



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ROHGI TOSHIO MENESES CHIKUSHI

**ABORDAGENS BASEADAS NA ANÁLISE DE SÉRIES TEMPORAIS NÃO
LINEARES PARA A DETECÇÃO DE MUDANÇAS DE CONCEITO NO
CONTEXTO DO APRENDIZADO ADAPTATIVO SUPERVISIONADO**

Recife

2021

ROHGI TOSHIO MENESES CHIKUSHI

ABORDAGENS BASEADAS NA ANÁLISE DE SÉRIES TEMPORAIS NÃO LINEARES PARA A DETECÇÃO DE MUDANÇAS DE CONCEITO NO CONTEXTO DO APRENDIZADO ADAPTATIVO SUPERVISIONADO

Tese apresentada ao Programa de Pós-Graduação em Ciências da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciências da Computação.

Área de Concentração: Inteligência Computacional.

Orientador (a): Prof. Dr. Roberto Souto Maior de Barros

Coorientador (a): Profa. Dra. Marilú Gomes Netto Monte da Silva

Recife

2021

Catálogo na fonte
Bibliotecária Nataly Soares Leite Moro, CRB4-1722

C534a Chikushi, Rohgi Toshio Meneses
Abordagens baseadas na análise de séries temporais não lineares para a detecção de mudanças de conceito no contexto do aprendizado adaptativo supervisionado / Rohgi Toshio Meneses Chikushi. – 2021.
219 f.: il., fig., tab.

Orientador: Roberto Souto Maior de Barros.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2021.
Inclui referências e apêndices.

1. Inteligência computacional. 2. Aprendizado de máquina. 3. Fluxos de dados. 4. Séries temporais não lineares. I. Barros, Roberto Souto Maior de (orientador). II. Título

006.31 CDD (23. ed.) UFPE - CCEN 2022 – 12

Rohgi Toshio Meneses Chikushi

**“Abordagens Baseadas na Análise de Séries Temporais não Lineares
para a Detecção de Mudanças de Conceito no Contexto do
Aprendizado Adaptativo Supervisionado”**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação. Área de Concentração: Inteligência Computacional.

Aprovado em: 25/08/2021.

Orientador: Prof. Dr. Roberto Souto Maior de Barros

BANCA EXAMINADORA

Prof. Dr. Germano Crispim Vasconcelos
Centro de Informática / UFPE

Prof. Dr. Paulo Mauricio Gonçalves Junior
Instituto Federal de Pernambuco

Prof. Dr. Wilson Rosa de Oliveira Junior
Departamento de Estatística e Informática /
UFRPE

Prof. Dr. Jean Paul Barddal
Escola Politécnica / PUC/PR

Prof. Dr. João Roberto Bertini Junior
Faculdade de Tecnologia / UNICAMP

Dedico esta tese aos meus pais, Akihiko e Margarida (in memoriam), a Catão Temístocles de Freitas Barbosa (in memoriam) e, em especial, aos meus amores, Alessandra e a pequena Hannah.

AGRADECIMENTOS

Agradeço primeiramente ao mistério da vida, a força que invocada ou não sempre está presente.

Ao meu orientador, Dr. Roberto Souto Maior de Barros, pelo profissionalismo, paciência e, principalmente, por ter contribuído diretamente na minha formação como pesquisador.

À minha coorientadora, Dra. Marilú Gomes Netto Monte da Silva, por seu suporte, troca de ideias e sugestões pertinentes para a melhoria do trabalho.

Aos amigos Bruno Iran Ferreira Maciel, Juan Isidro Gonzalez, Jose Luis Martinez Perez, Silas Garrido Teixeira de Carvalho Santos, Danilo Rafael de Lima Cabral, e aos demais membros do grupo de pesquisa pelo apoio no desenvolvimento desta tese.

À banca, pelas contribuições para a melhoria deste trabalho e ao Programa de Pós-graduação em Ciências de Computação do Centro de Informática (Cin-UFPE) pela estrutura oferecida ao desenvolvimento desta pesquisa. Agradeço também a todos os professores e ao pessoal da secretaria da pós pelo suporte durante o tempo que passei no Centro de Informática.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo suporte financeiro durante meu período no doutorado.

Agradeço também a todos os colegas e familiares pelo apoio e companheirismo.

Por fim, agradeço à minha querida e amada Alessandra que dividiu ao meu lado angústias e incertezas, mas sempre superadas com alegria.

RESUMO

Atualmente, algoritmos de Aprendizado de Máquina são aplicados em diversos domínios para a extração de informação em grandes volumes de dados. Apesar de modelos consolidados lidarem de forma efetiva com dados identicamente e independentemente distribuídos (i.i.d.), algoritmos aplicados a fluxos contínuos de dados devem lidar com distribuições não estacionárias. O desafio é manter um modelo de decisão atualizado, preciso e consistente, mesmo sujeito a mudanças nas distribuições de probabilidade dos dados, um fenômeno conhecido como mudança de conceito. Neste contexto, os algoritmos combinam um classificador e um detector para identificar mudanças na distribuição do erro das predições a fim de adaptar ou substituir rapidamente o modelo preditivo. Diversas propostas têm sido apresentadas na literatura para a detecção de mudanças de conceito com base na taxa de erro dos modelos preditivos. Em geral, a distribuição da taxa de erro fundamenta a maioria das abordagens baseadas em metodologias como a análise sequencial, o controle estatístico do processo, ou pelo monitoramento das distribuições por meio de janelas deslizantes, as quais assumem que os erros de predição são gerados de forma independente. Apesar da vasta aplicação desses detectores, estudos empíricos têm mostrado que a taxa de erro pode ser influenciada pela dependência temporal. Além disso, abordagens supervisionadas requerem dados rotulados, os quais podem ser difíceis de obter em muitas aplicações do mundo real. Nesta tese, ferramentas de Análise de Séries Temporais Não Lineares foram utilizadas com o objetivo de prover detectores não restritos ao pressuposto de observações i.i.d e mais apropriados para lidar com fluxos de dados sujeitos à dependência temporal. Neste sentido, foram propostos três detectores: Spectral Entropy Drift Detector (SEDD), Permutation Entropy Drift Detector (PEDD), e Recurrence Quantification Analysis Drift Detector (RQADD). Também foi proposto o Symbolic Labeling Adapter (SLA), uma abordagem de pseudo-rotulação simbólica com o intuito de expandir a aplicação de modelos adaptativos supervisionados a domínios onde fluxos de dados não são rotulados, visando a detecção de mudanças de conceito. Experimentos com os classificadores *Naive Bayes* e *Hoffding Tree* utilizando 15 detectores, 20 bases de dados reais e 360 artificiais, sugerem que o SEDD, embora não tenha superado o estado da arte em termos de acurácia na maioria dos cenários, não apresentou diferença estatística significativa em relação aos mesmos, sinalizando menos alarmes falsos. De modo semelhante, os detectores PEDD e RQADD foram competitivos acerca das detecções de *mudanças de conceito* na maioria das avaliações, principalmente utilizando o SLA com bases não rotuladas (30 artificiais e 26 de eletroence-

falograma). Desse modo, pode-se considerar os detectores propostos como uma alternativa competitiva, e a abordagem de pseudo-rotulação simbólica uma ferramenta promissora.

Palavras-chaves: aprendizado de máquina; fluxos de dados; mudança de conceito; séries temporais não lineares.

ABSTRACT

Currently, Machine Learning algorithms are applied in several domains to extract information from large datasets. Although most models effectively deal with identically and independently distributed (i.i.d.) data, algorithms applied to data streams must deal with non-stationary distributions. The challenge is to maintain an up-to-date, accurate and consistent decision model, even when changes in the probability distributions of the data occur, a phenomenon known as concept drift. In this context, algorithms usually combine a classifier and a detector to identify changes in the error distribution of the predictions in order to quickly adapt or replace the predictive model. Several proposals have been presented in the literature for detecting concept drift based on the error rate of predictive models. In general, the error rate distribution underlies most approaches based on methodologies such as sequential analysis, statistical process control, or by monitoring distributions through sliding windows, which assume that prediction errors are generated independently. Despite the wide application of these detectors, empirical studies have shown that the error rate can be influenced by temporal dependence. In addition, supervised approaches require labeled data, which can be difficult to obtain in many real-world applications. In this thesis, Nonlinear Time Series Analysis tools were used in order to propose detectors not restricted to the assumption of i.i.d. observations, and more appropriate to deal with data streams subjected to temporal dependence. In this sense, three detectors are proposed: Spectral Entropy Drift Detector (SEDD), Permutation Entropy Drift Detector (PEDD), and Recurrence Quantification Analysis Drift Detector (RQADD). It was also proposed the Symbolic Labeling Adapter (SLA), a symbolic pseudo-labeling approach in order to expand the application of supervised adaptive models to domains where data flows are not labelled, aiming to detect concept changes. Experiments with the classifiers *Naïve Bayes* and *Hoffding Tree* using 15 detectors, 20 real datasets and 360 artificial ones, suggest that SEDD, although it has not outperformed those state-of-art detectors in accuracy in most scenarios, did not show statistical difference in relation to them, signaling fewer false alarms. Similarly, the PEDD and RQADD detectors were competitive on the detections of *concept drift* in most evaluations, mainly using SLA with unlabeled datasets (30 artificial and 26 electroencephalogram). Therefore, the proposed detectors can be considered as a competitive alternative, and the symbolic pseudo-labeling approach a promising tool.

Keywords: machine learning; data streams; concept drift; nonlinear time series.

LISTA DE FIGURAS

Figura 1 – Série temporal discretizada (pontos marcados por círculos com segmentos verticais) obtida de uma série temporal contínua (linha contínua preta). . . .	53
Figura 2 – <i>Diagramas teia de aranha</i> (do Inglês <i>cobweb plot</i>) produzidos a partir dos estados iniciais $x_0^A = 0,4286$ e $x_0^C = 0,2764$, identificados pela linha tracejada.	60
Figura 3 – Série temporal obtida a partir da evolução da orbita periódica referente ao estado inicial x_0^A	60
Figura 4 – Série temporal obtida a partir da evolução da orbita aperiódica referente ao estado inicial x_0^C	60
Figura 5 – <i>Diagramas teia de aranha</i> da evolução do Mapa Logístico e suas respectivas séries temporais, considerando um <i>ponto-fixo</i> ($r = 2,9$), uma <i>órbita periódica</i> ($r = 3,2$), e uma <i>órbita caótica</i> ($r = 4,0$).	62
Figura 6 – RMSE obtido das séries temporais referentes à evolução dos estados iniciais \tilde{x}_0^A e \tilde{x}_0^B , considerando $r = 4,0$	63
Figura 7 – RMSE obtido das séries temporais referentes à evolução dos estados iniciais \tilde{x}_0^A e \tilde{x}_0^B , considerando $r = 3,2$	63
Figura 8 – Trajetória da evolução temporal dos estados em um <i>espaço de fase</i> no \mathbb{R}^3 com as séries temporais das variáveis x , y e z do Sistema de Lorenz.	65
Figura 9 – <i>Diagrama de bifurcação</i> do Mapa Logístico	66
Figura 10 – Série temporal gerada pelo Mapa Logístico Transiente considerando iterações do parâmetro $r(t)$ entre os valores 2,8 e 4,0, com incrementos de 10^{-5}	67
Figura 11 – Série temporal gerada pelo Sistema de Lorenz Transiente considerando iterações discretas do parâmetro $\rho(t)$ entre os valores 28,0 e 268,0, com incrementos de 0,002.	68
Figura 12 – Gráfico da Average Mutual Information Function (traduzido para o português, Função de Informação Mútua Média) (AMIF) para a estimação do parâmetro τ referente à variável x do Sistema de Lorenz, e as respectivas reconstruções do atrator, considerando as defasagens $\tau = 1, \tau = 8, \tau = 16$ e $\tau = 60$	71

Figura 13 – Ilustração do método False Nearest Neighbours (traduzido para o português, Falsos Vizinhos Próximos) (FNN) para a reconstrução do <i>espaço de fase</i> . Na cor preta, um segmento do atrator do Sistema de Lorenz no espaço $\mathbb{R}^{d=3}$. No plano XY , na cor cinza, o mesmo segmento no espaço $\mathbb{R}^{d=2}$	73
Figura 14 – Gráfico dos indicadores $E(1)$ e $E(2)$ obtidos do método de Cao aplicado sobre a série temporal da variável x do Sistema de Lorenz.	75
Figura 15 – Reconstrução do atrator por meio da série temporal da variável x do Sistema de Lorenz, considerando a dimensão estimada $m_r = 3$	76
Figura 16 – Ilustração do Diagrama de Recorrência (RP) sobre o atrator reconstruído a partir da série temporal da variável x do Sistema de Lorenz. A Figura (a) representa o processo de construção do RP apresentado na Figura (b).	78
Figura 17 – Exemplos de diagramas de recorrência para um sistema com dinâmica periódica (a), caótica (b) e estocástica (c).	79
Figura 18 – Padrões de todas as ordenações possíveis para $m = 4$ ($4! = 24$ símbolos). O eixo das abscissas corresponde aos índices de permutação das componentes $(x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$ para qualquer estado \mathbf{v}_i no espaço $\mathbb{R}^{d=4}$	81
Figura 19 – Magnitudes (A_ξ) relacionadas às frequências (ξ) obtidas como componentes espectrais da série temporal apresentada na Figura 1.	82
Figura 20 – Ângulos de fase (ϕ_ξ) em radianos, relacionados às respectivas magnitudes (A_ξ) apresentadas na Figura 19.	82
Figura 21 – (A) Um fluxo de dados sintético com duas janelas de observações consecutivas, destacadas com linhas tracejadas na cor preta e cinza, respectivamente; (B) seus respectivos histogramas (barras pretas e cinzas) em que não é possível identificar qualquer diferença; (C) espectros de potência normalizados, os quais apresentam diferenças sutis entre as janelas; Finalmente, (D) os respectivos espaços de fase, os quais são significativamente diferentes.	85
Figura 22 – (A) Amostra do fluxo de erro com o ponto de mudança na instância 500 (linha cinza tracejada); (B) A série temporal obtida após a aplicação do Mapa de Bernoulli.	87
Figura 23 – (A) Série temporal do fluxo de erros obtida por meio do Mapa de Bernoulli; (B) Entropia Espectral; (C) Entropia de Shannon; (D) Determinismo; e (E) Entropia de Permutação. Linhas tracejadas marcam as <i>mudanças de conceito</i> do fluxo de dados.	94

Figura 24 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.	122
Figura 25 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.	123
Figura 26 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base Agrawal desbalanceada, considerando um intervalo de confiança de 95%.	129
Figura 27 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal desbalanceada, considerando um intervalo de confiança de 95%.	130
Figura 28 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.	135
Figura 29 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.	136
Figura 30 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.	140
Figura 31 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.	141
Figura 32 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com duas classes, considerando um intervalo de confiança de 95%.	147
Figura 33 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com duas classes, considerando um intervalo de confiança de 95%.	148
Figura 34 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.	151

Figura 35 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.	152
Figura 36 – Comparação dos resultados para o consumo de memória (<i>MEM</i>) e tempo de processamento (<i>TIME</i>) usando o teste post-hoc de Nemenyi sobre as 6 bases artificiais, considerando um intervalo de confiança de 95%.	157
Figura 37 – Comparação dos classificadores e detectores com HT e NB usando o teste <i>post-hoc</i> de Nemenyi sobre as bases reais rotuladas, considerando um intervalo de confiança de 95%, e o nível de significância $\alpha = 0,05$	161
Figura 38 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 3$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.	167
Figura 39 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 3$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.	168
Figura 40 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 4$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.	171
Figura 41 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 4$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.	172
Figura 42 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 5$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.	174
Figura 43 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 5$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.	175

Figura 44 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi sobre todas as bases de dados com rotulações simbólicas geradas a partir dos sinais de EEG, considerando um intervalo de confiança de 95%.	180
Figura 45 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi sobre todas as bases de dados com rotulações simbólicas geradas a partir dos sinais de EEG, considerando um intervalo de confiança de 95%.	181
Figura 46 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.	200
Figura 47 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.	201
Figura 48 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base Agrawal desbalanceada, considerando um intervalo de confiança de 95%.	202
Figura 49 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal desbalanceada, considerando um intervalo de confiança de 95%.	203
Figura 50 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.	205
Figura 51 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.	206
Figura 52 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.	208
Figura 53 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.	209

Figura 54 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com duas classes, considerando um intervalo de confiança de 95%.	211
Figura 55 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com duas classes, considerando um intervalo de confiança de 95%.	212
Figura 56 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.	214
Figura 57 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.	215

LISTA DE TABELAS

Tabela 1 – Algoritmo para estimar a Entropia Espectral.	88
Tabela 2 – Algoritmo para estimar o Determinismo com base na Análise de Quantificação de Recorrência (RQA).	90
Tabela 3 – Algoritmo para estimar a Entropia de Permutação.	92
Tabela 4 – Algoritmo para identificar a mudança de conceito.	95
Tabela 5 – Exemplo do mapeamento de uma série temporal em instâncias Independente e Identicamente Distribuídas (i.i.d.) (coluna $v^{2,1}$) obtidas após a reconstrução do <i>espaço de fase</i> , considerando os parâmetros $m = 2$, e $\tau = 1$	97
Tabela 6 – Exemplo da generalização do mapeamento, em que o conjunto de dados de entrada (\mathcal{X}) e retorno (\mathcal{Y}) estão definidos em uma dimensão m ($\mathcal{X}, \mathcal{Y} \in \mathbb{R}^m$), e as instâncias resultantes, na dimensão $2m$ ($\mathcal{X} \wedge \mathcal{Y} \in \mathbb{R}^{2m}$).	98
Tabela 7 – Exemplo da pseudo-rotulação simbólica, em que o conjunto de símbolos de retorno (\mathcal{W}) estão definidos na dimensão 1 ($\mathcal{W} \in \mathbb{N}$), e são obtidos a partir do conjunto \mathcal{Y} . Nesse caso, as instâncias resultantes estão definidas na dimensão $m + 1$ ($\mathcal{X} \wedge \mathcal{W} \in \mathbb{R}^{m+1}$).	99
Tabela 8 – Algoritmo para gerar a pseudo-rotulação simbólica.	99
Tabela 9 – Características das bases de dados reais utilizadas nos experimentos com seus respectivos repositórios.	106
Tabela 10 – Bases de dados de EEG selecionadas com os respectivos pontos de mudanças.	109
Tabela 11 – Detectores utilizados nos experimentos com, respectivamente, o número de parâmetros, e o tamanho das janelas de observações.	110
Tabela 12 – Percentual das frequências em que os detectores ficaram na barra ligada ao topo do ranqueamento dos gráficos de Critical Differences (traduzido para o português, Diferença Crítica) (CD).	118
Tabela 13 – Resultados obtidos usando HT, considerando a base artificial Agrawal balanceada, com um intervalo de confiança de 95%.	121
Tabela 14 – Resultados obtidos usando HT, considerando a base artificial Agrawal desbalanceada, com um intervalo de confiança de 95%.	128

Tabela 15 – Resultados obtidos usando HT, considerando a base artificial LED tendo 5% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.	134
Tabela 16 – Resultados obtidos usando HT, considerando a base artificial LED tendo 10% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.	139
Tabela 17 – Resultados obtidos usando HT, considerando a base artificial RBF com duas classes, com um intervalo de confiança de 95%.	146
Tabela 18 – Resultados obtidos usando HT, considerando a base artificial RBF com cinco classes, com um intervalo de confiança de 95%.	150
Tabela 19 – Resultados do uso de memória em bytes e tempo médio de execução em segundos obtidos usando os classificadores Hoeffding Tree (HT) e Naïve Bayes (NB), considerando as 6 bases artificiais obtidas dos geradores, com um intervalo de confiança de 95%.	156
Tabela 20 – Resultados obtidos usando HT e NB, considerando as 20 bases reais, com um intervalo de confiança de 95%.	160
Tabela 21 – Resultados obtidos usando HT e NB a partir das rotulações simbólicas da base artificial do Sistema Lorenz Transiente, considerando a dimensão de imersão $m = 3$ e a defasagens $\tau = 3$, com um intervalo de confiança de 95%.	166
Tabela 22 – Resultados obtidos usando HT e NB a partir das rotulações simbólicas da base artificial do Sistema Lorenz Transiente, considerando a dimensão de imersão $m = 4$ e a defasagens $\tau = 3$, com um intervalo de confiança de 95%.	170
Tabela 23 – Resultados obtidos usando HT e NB a partir das rotulações simbólicas da base artificial do Sistema Lorenz Transiente, considerando a dimensão de imersão $m = 5$ e a defasagens $\tau = 3$, com um intervalo de confiança de 95%.	173
Tabela 24 – Resultados obtidos usando HT e NB, considerando as 26 bases de EEG, com um intervalo de confiança de 95%.	179
Tabela 25 – Questões de pesquisa e as respectivas respostas obtidas por meio da análise dos resultados experimentais.	186

Tabela 26 – Resultados obtidos usando NB, considerando a base artificial Agrawal balanceada, com um intervalo de confiança de 95%.	198
Tabela 27 – Resultados obtidos usando NB, considerando a base artificial Agrawal desbalanceada, com um intervalo de confiança de 95%.	199
Tabela 28 – Resultados obtidos usando NB, considerando a base artificial LED tendo 5% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.	204
Tabela 29 – Resultados obtidos usando NB, considerando a base artificial LED tendo 10% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.	207
Tabela 30 – Resultados obtidos usando NB, considerando a base artificial RBF com duas classes, com um intervalo de confiança de 95%.	210
Tabela 31 – Resultados obtidos usando NB, considerando a base artificial RBF com cinco classes, com um intervalo de confiança de 95%.	213
Tabela 32 – Resultados com as acurácias (%) obtidas usando HT, considerando as 20 bases reais.	216
Tabela 33 – Resultados com as acurácias (%) obtidas usando NB, considerando as 20 bases reais.	217
Tabela 34 – Resultados com as medidas ARL obtidas usando HT, considerando as 20 bases reais.	218
Tabela 35 – Resultados com as medidas ARL obtidas usando NB, considerando as 20 bases reais.	219

LISTA DE ABREVIATURAS E SIGLAS

ADWIN	Adaptive Windowing
AM	<i>Aprendizado de Maquina</i> (do inglês, <i>Machine Learning</i>)
AMIF	Average Mutual Information Function (traduzido para o português, Função de Informação Mútua Média)
ARIMA	Autoregressive Integrated Moving Average (traduzido para o português, Modelo Auto-regressivo Integrado de Médias Móveis)
ARL	Average Running Length (traduzido para o português, Comprimento Médio de Sequência)
CD	Critical Differences (traduzido para o português, Diferença Crítica)
CUSUM	Cumulative Sum
DDM	Drift Detection Method
ECDD	EWMA for Concept Drift Detection
EDDM	Early Drift Detection Method
EEG	Eletroencefalograma
EWMA	Exponentially Weighted Moving Average
FAC	Autocorrelation Function (traduzido para o português, Função de Autocorrelação)
FFT	Fast Fourier Transform (traduzido para o português, Transformada Rápida de Fourier)
FHDDM	Fast Hoeffding Drift Detection Method
FNN	False Nearest Neighbours (traduzido para o português, Falsos Vizinhos Próximos)
FPDD	Fisher Proportions Drift Detector
FSDD	Fisher Squared Drift Detector
FTDD	Fisher Test Drift Detector
HDDM	Drift Detection Method based on the Hoeffding's inequality

HT	Hoeffding Tree
i.i.d.	Independente e Identicamente Distribuídas
IoT	Internet of Things
k-NN	k-Nearest Neighbors (traduzido para o português, k-Vizinhos Mais Próximos)
LFR	Linear Four Rates
MDDM	McDiarmid Drift Detection Method
MDR	Missed Detection Rate (traduzido para o português, Taxa de Detecções Perdidas)
MOA	Massive Online Analysis Framework
MTD	Mean Time for Detection (traduzido para o português, Tempo Médio de Detecção)
MTFA	Mean Time Between False Alarms (traduzido para o português, Tempo Médio Entre Alarmes Falsos)
MTR	Mean Time Ratio (traduzido para o português, Razão de Tempo Médio)
NB	Naïve Bayes
NLTA	Nonlinear Time Series Analysis (traduzido para o português, Análise de Séries Temporais Não Lineares)
PAC	Probably Approximately Correct
PE	Entropia de Permutação
PEDD	Permutation Entropy Drift Detector
PHT	Page-Hinckley Test
PS	Espectro de Potência
RDDM	Reactive Drift Detection Method
RMSE	Root Mean Square Error
RP	Diagrama de Recorrência
RQA	Análise de Quantificação de Recorrência
RQADD	Recurrence Quantification Analysis Drift Detector

SE	Entropia Espectral
SEDD	Spectral Entropy Drift Detector
SEED	SEED Drift Detector
SEQD2	Sequential Hypothesis Testing Drift Detector 2
SEQDRIFT	Sequential Hypothesis Testing Drift Detector 1
SLA	Symbolic Labeling Adapter
SLT	Statistical Learning Theory (traduzido para o português, Teoria do Aprendizado Estatístico)
SPC	Statistical Process Control (traduzido para o português, Controle Estatístico de Processos)
SPRT	Sequential Probability Ratio Test
SSC	Surrogate Stability Concept
STEPD	Statistical Test of Equal Proportions
SVD	Singular Value Decomposition (traduzido para o português, Decomposição em Valores Singulares)
TSFTM	Theiler's Surrogate Fourier Transformed Method
VFDT	Very Fast Decision Tree
WSTD	Wilcoxon Rank Sum Test Drift Detector

LISTA DE SÍMBOLOS

Γ	espaço de fase
m	dimensão de reconstrução do espaço de fase
τ	retardo temporal
$\mathbf{v}_i^{m,\tau}$	estado do espaço de fase reconstruído na dimensão m com defasagem τ
π_*^m	símbolo/padrão ordinal obtido de um estado na dimensão m
\mathcal{L}	classificador base
\mathcal{D}	detector de mudança de conceito
\mathcal{X}	conjunto dos atributos de entrada
\mathcal{Y}	conjunto dos atributos alvos/rótulos
y_i	atributos alvo/rótulo
\mathbf{x}_i	atributo de entrada
x_{ij}	componente j do atributo \mathbf{x}_i
ξ	componente de frequência na dimensão espectral
A_ξ	amplitude do componente de frequência ξ
ϕ_ξ	fase do componente de frequência ξ
$\{x_t\}$	série temporal \mathbf{s}
x_t	observação t da série temporal \mathbf{s}
\mathbf{R}	diagrama de recorrência
ε	raio de vizinhança do diagrama de recorrência

SUMÁRIO

1	INTRODUÇÃO	26
1.1	CONTEXTUALIZAÇÃO E MOTIVAÇÃO	26
1.2	OBJETIVO, HIPÓTESES, E QUESTÕES DE PESQUISA	29
1.3	ORGANIZAÇÃO DO TRABALHO	31
2	FLUXO DE DADOS E MUDANÇA DE CONCEITO	33
2.1	FLUXO DE DADOS	33
2.1.1	Aprendizado adaptativo	34
2.1.2	Classificadores aplicados a fluxos contínuos de dados	37
2.1.2.1	<i>Very Fast Decision Tree</i>	37
2.1.2.2	<i>Naïve Bayes</i>	38
2.1.2.3	<i>Classificadores ingênuos</i>	39
2.2	MUDANÇA DE CONCEITO	40
2.2.1	Detectores de mudança de conceito	41
2.2.1.1	<i>Análise Sequencial</i>	42
2.2.1.2	<i>Controle Estatístico de Processos</i>	44
2.2.1.3	<i>Monitoramento em Janelas de Distribuições</i>	48
2.3	DEPENDÊNCIA TEMPORAL	51
2.4	RESUMO DO CAPÍTULO	52
3	ANÁLISE DE SÉRIES TEMPORAIS NÃO LINEARES	53
3.1	SÉRIES TEMPORAIS	53
3.2	DINÂMICA NÃO LINEAR E CAOS	55
3.2.1	Órbitas, atratores e caos	56
3.2.1.1	<i>Mapa de Bernoulli</i>	58
3.2.1.2	<i>Mapa Logístico</i>	61
3.2.1.3	<i>Sistema de Lorenz</i>	64
3.2.2	Bifurcação e modelos transientes	65
3.3	RECONSTRUÇÃO DO ESPAÇO DE FASE	68
3.3.1	Retardo Temporal	69
3.3.2	Dimensão de imersão	72

3.4	DETECTANDO ALTERAÇÕES NA DINÂMICA DE SÉRIES TEMPORAIS NÃO LINEARES	76
3.4.1	Análise de Quantificação de Recorrência	77
3.4.2	Entropia de permutação	80
3.4.3	Entropia espectral	81
3.5	RESUMO DO CAPÍTULO	83
4	PROPOSTA DE DETECTORES E PSEUDO-ROTULAÇÃO SIM- BÓLICA	84
4.1	AVALIAÇÃO ENTRE JANELAS DE OBSERVAÇÕES	84
4.2	MAPEAMENTO DO FLUXO DE ERROS EM UMA SÉRIE TEMPORAL ESCALAR	86
4.2.1	Aplicação do Mapa de Bernoulli	86
<i>4.2.1.1</i>	<i>Estimador da Entropia Espectral</i>	<i>88</i>
<i>4.2.1.2</i>	<i>Estimador do Determinismo da RQA</i>	<i>89</i>
<i>4.2.1.3</i>	<i>Estimador da Entropia de Permutação</i>	<i>91</i>
4.2.2	Heurística de detecção	92
4.3	PSEUDO-ROTULAÇÃO SIMBÓLICA	96
4.4	RESUMO DO CAPÍTULO	100
5	ANÁLISE EMPÍRICA E ESTATÍSTICA	102
5.1	BASES DE DADOS	102
5.1.1	Bases Artificiais	102
<i>5.1.1.1</i>	<i>Geradores de bases artificiais</i>	<i>103</i>
<i>5.1.1.2</i>	<i>Sistema de Lorenz Transiente</i>	<i>104</i>
5.1.2	Bases do mundo real	105
<i>5.1.2.1</i>	<i>Bases reais rotuladas</i>	<i>105</i>
<i>5.1.2.2</i>	<i>Bases de Eletroencefalograma (EEG)</i>	<i>108</i>
5.2	CONFIGURAÇÕES DOS EXPERIMENTOS	109
5.2.1	Métricas de avaliação das detecções de mudanças de conceito . . .	111
5.2.2	Testes estatísticos	113
5.2.3	Análise de sensibilidade dos parâmetros	114
5.3	RESULTADOS E ANÁLISES DO EXPERIMENTOS COM BASES ARTIFI- CIAIS	115
5.3.1	Análise dos resultados com as bases artificiais	117

5.3.1.1	<i>Resultados e análises considerando classes balanceadas e desbalanceadas . . .</i>	120
5.3.1.2	<i>Resultados e análises considerando 5% e 10% de ruído</i>	132
5.3.1.3	<i>Resultados e análises considerando as configurações com 2 e 5 classes . . .</i>	143
5.3.2	Avaliação do uso de memória e tempo de processamento	155
5.4	RESULTADOS E ANÁLISE DO EXPERIMENTO COM BASES REAIS ROTULADAS	158
5.5	RESULTADOS E ANÁLISE DO EXPERIMENTO COM O SISTEMA DE LORENZ	163
5.6	RESULTADOS E ANÁLISE DO EXPERIMENTO COM BASES DE EEG . . .	176
5.7	RESUMO DO CAPÍTULO	182
6	CONSIDERAÇÕES FINAIS	183
6.1	PUBLICAÇÃO RELACIONADA À TESE	187
	REFERÊNCIAS	188
	APÊNDICE A – RESULTADOS EXPERIMENTAIS DAS BASES ARTIFICIAIS COM O NB	198
	APÊNDICE B – RESULTADOS EXPERIMENTAIS COM BASES REAIS	216

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO E MOTIVAÇÃO

Atualmente, modelos de *Aprendizado de Máquina* (do inglês, *Machine Learning*) (AM) são aplicados em diversos domínios para a extração de conhecimento em grandes volumes de dados (HAN; PEI; KAMBER, 2011). Usualmente, a indução do modelo ocorre de modo *offline*, onde o algoritmo de aprendizado tem acesso a todo o conjunto de dados disponível para o treinamento. Apesar de algoritmos consolidados lidarem de forma efetiva com conjuntos de dados estáticos, em algumas aplicações do mundo real, esses algoritmos devem atuar em ambientes dinâmicos onde as instâncias são coletadas na forma de *fluxos contínuos de dados* (do Inglês, *data stream*) (SCHLIMMER; GRANGER, 1986; GAMA, 2010; GAMA et al., 2014; KRAWCZYK et al., 2017).

Em cenários como o tráfego em redes de computadores, localização por GPS, chamadas de telefones móveis, comércio eletrônico, Internet of Things (IoT), dentre muitos outros, fluxos contínuos de dados são transferidos em grande volume e velocidade, inviabilizando a aplicação de abordagens convencionais de AM devido a restrições computacionais de armazenamento e tempo de processamento (GAMA, 2010; GAMA et al., 2014). Além disso, os dados podem ser desestruturados, ruidosos, não-estacionários, e apresentarem dependência temporal. Nesse contexto, o desafio é manter um modelo de decisão atualizado, preciso e consistente, mesmo sujeito a alterações nas distribuições dos dados, um fenômeno conhecido como *mudança de conceito* (do Inglês *concept drift*) (SCHLIMMER; GRANGER, 1986; TSYMBAL, 2004; GAMA et al., 2014; KRAWCZYK et al., 2017).

Nestes cenários, algoritmos de aprendizado *online* são aplicados considerando uma *estratégia adaptativa* (GAMA et al., 2014). Diferentemente do aprendizado *offline*, o qual pressupõe que as instâncias são geradas por uma distribuição estacionária e coletadas antes da fase de treinamento, na *estratégia adaptativa* o aprendizado ocorre de modo *incremental*. Nessa abordagem, os modelos são continuamente atualizados a cada nova instância, e geralmente o treinamento ocorre numa única passagem. Além disso, algoritmos adaptativos são projetados para lidar dinamicamente com a não estacionariedade da distribuição dos dados e consumir menos recursos computacionais.

Várias abordagens foram propostas para o aprendizado em cenários de *fluxo de dados* (TSYMBAL et al., 2008; BIFET, 2009; ELWELL; POLIKAR, 2011; GAMA et al., 2014; KRAWCZYK

et al., 2017). Do mesmo modo, diversos algoritmos têm sido apresentados na literatura para a detecção de *mudança de conceito* (COSTA et al., 2017; SETHI; KANTARDZIC, 2017; KRAWCZYK et al., 2017; BARROS; SANTOS, 2018; BARROS; HIDALGO; CABRAL, 2018). Nessas propostas, busca-se tratar a *mudança de conceito* de forma explícita, por meio de medidas de performance de classificadores, ou não-supervisionada, por meio da avaliação dos atributos dos dados.

Em geral, os algoritmos combinam um classificador e um detector com o objetivo de identificar mudanças na distribuição dos dados para atualizar ou substituir rapidamente o modelo preditivo (GAMA et al., 2014). Assumindo que a única informação recebida pelo detector é o resultado da predição, em geral 0 se a predição estiver correta ou 1 se estiver incorreta, pode-se considerar o fluxo de erros como uma sequência de ensaios de Bernoulli (ROSS et al., 2012). Essas estratégias reduzem o problema de detecção da *mudança de conceito* ao problema de identificar diferenças significativas na distribuição da taxa de erro, considerando a hipótese de observações independentes (ROSS et al., 2012; GAMA et al., 2014; ŽLIOBAITĚ et al., 2015). Neste sentido, uma configuração comum é usar o detector para analisar os resultados das predições de um classificador base e aplicar alguma regra de decisão sobre a distribuição estimada do erro para informar se o modelo está desatualizado devido a mudanças na distribuição dos dados.

Diferentes métodos têm sido propostos na literatura para a detecção de *mudança de conceito* com base na não estacionariedade da distribuição da taxa de erro de modelos preditivos (BARROS; SANTOS, 2018). Apesar da eficiência de muitos detectores, estudos empíricos têm mostrado que a taxa de erro pode ser influenciada pela dependência temporal (ŽLIOBAITĚ et al., 2015; BIFET, 2017). Esse efeito está relacionado à propagação da acurácia devido a sucessivas atualizações incrementais do modelo (BIFET, 2017). Desse modo, nenhuma garantia teórica pode ser fornecida por esses métodos quando as observações não são Independente e Identicamente Distribuídas (i.i.d.) (ŽLIOBAITĚ et al., 2015). Conseqüentemente, informações relevantes da dependência temporal entre os erros das predições são desconsideradas quando apenas a não estacionariedade da distribuição da taxa de erro é avaliada (COSTA; RIOS; MELLO, 2016).

Outra dificuldade é que abordagens supervisionadas requerem dados rotulados, os quais podem ser difíceis de obter em muitas aplicações do mundo real onde a rotulagem é demorada, cara e em alguns casos inviável (DYER; CAPO; POLIKAR, 2014). Além do mais, a rotulagem de dados para retreinamento e atualização de modelos com a única finalidade de detecção de *mudança de conceito* é uma tarefa dispendiosa e supérflua (SETHI; KANTARDZIC, 2015; SETHI;

KANTARDZIC, 2017).

Algoritmos aplicados a *fluxos contínuos de dados* pressupõe independência temporal entre os dados e priorizam o paradigma de aprendizado supervisionado (TSYMBAL et al., 2008; BIFET et al., 2009; ELWELL; POLIKAR, 2011; GAMA et al., 2014; KRAWCZYK et al., 2017). Segundo a Statistical Learning Theory (traduzido para o português, Teoria do Aprendizado Estatístico) (SLT) (LUXBURG; SCHÖLKOPF, 2011), a abordagem supervisionada tem como pressuposto a existência de uma distribuição de probabilidade conjunta desconhecida $P(\mathcal{X} \times \mathcal{Y})$ sobre o conjunto de combinações entre exemplos $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ e classes (ou rótulo) $\mathcal{Y} = \{y_1, \dots, y_m\}$. O processo de aprendizado se dá pela busca de um modelo $f: \mathcal{X} \rightarrow \mathcal{Y}$, o qual deve mapear cada exemplo \mathbf{x}_i (*i.i.d.*) à sua respectiva classe y_i , sobre o conjunto \mathcal{F} de possíveis modelos que melhor represente $P(\mathcal{X} \times \mathcal{Y})$, onde $f \in \mathcal{F}$. No entanto, para fluxos contínuos de diversos domínios do mundo real, não há garantias da disponibilidade de dados estruturados, previamente rotulados e de uma distribuição de probabilidade conjunta fixa. Além disso, esses fluxos de dados estão sujeitos a ruídos, a não-estacionaridade, a apresentarem não-linearidade e dependência temporal (SCHREIBER, 1999; MARWAN et al., 2007; KANTZ; SCHREIBER, 2004; STAM, 2005). Entretanto, a maioria dos estudos atuais assume que um conceito é composto por um conjunto de observações *i.i.d.*, geralmente modelados por meio de processos lineares e estacionários (COSTA; RIOS; MELLO, 2016).

Recentemente, abordagens não supervisionadas fundamentadas na Nonlinear Time Series Analysis (traduzido para o português, Análise de Séries Temporais Não Lineares) (NLTA) (KANTZ; SCHREIBER, 2004; BRADLEY; KANTZ, 2015) foram propostas para a detecção de *mudança de conceito* em fluxos de dados contendo dependência temporal (VALLIM; MELLO, 2014; COSTA; RIOS; MELLO, 2016; COSTA et al., 2017). Esses estudos sugerem que a *mudança de conceito* pode ser caracterizada por relações temporais entre os dados, e a sua detecção pode ser realizada ao mapear as dependências dos dados em um *espaço de fase* (TAKENS, 1981), bem como, por meio da avaliação de divergências entre gráficos de espectro de potência (VALLIM; MELLO, 2014). Essas abordagens levam em consideração a dependência temporal, assumindo que as observações não são necessariamente *i.i.d.* (VALLIM; MELLO, 2014; COSTA; RIOS; MELLO, 2016).

Além da finalidade de detecção de *mudança de conceito*, modelos de classificação também podem ser inferidos a partir de dados com dependências temporais por meio da aplicação do teorema de imersão de Takens (TAKENS, 1981), de modo a desdobrar esses dados em instâncias *i.i.d.* (PAGLIOSA; MELLO, 2017). Esse teorema é a base da NLTA (KANTZ; SCHREIBER, 2004;

BRADLEY; KANTZ, 2015), o qual viabiliza a reconstrução do *espaço de fase*, definido por um conjunto de estados $\mathbf{v}_i^{m,\tau} = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$ obtidos a partir de uma série temporal $\mathbf{s} = \{x_0, x_1, x_2, \dots, x_{N-1}\}$, onde o *retardo temporal* τ e a dimensão de reconstrução m são parâmetros livres.

A partir da estimação dos parâmetros, o *espaço de fase* pode ser utilizado como uma ferramenta importante para a avaliação das relações temporais entre os dados, permitindo a identificação de *mudanças de conceito* em termos da dinâmica do processo responsável pela geração do *fluxo de dados* (COSTA; RIOS; MELLO, 2016). Além disso, o *espaço de fase* possibilita uma avaliação simbólica da dinâmica do processo subjacente aos dados (BANDT; POMPE, 2002; CAO et al., 2004; AMIGÓ, 2010), a qual, em conjunto com a transformação de dados proposta por Pagliosa e Mello (2017), foram utilizadas nesta pesquisa para viabilizar a rotulação automática de *fluxos de dados unidimensionais* com a finalidade de detecção de *mudanças de conceito* relacionadas à dinâmica do *fluxo de dados*.

Inspirado nessas novas abordagens não supervisionadas e fundamentadas na NLTA, esta pesquisa considera a dependência temporal dos erros produzidos por previsões consecutivas ao longo do tempo. Assim, propõe-se utilizar ferramentas da NLTA no contexto do aprendizado adaptativo supervisionado para lidar com a dependência temporal do fluxo de erros, bem como prover rótulos artificiais, com o objetivo de detectar *mudanças de conceito*.

1.2 OBJETIVO, HIPÓTESES, E QUESTÕES DE PESQUISA

Nesta tese, técnicas da Nonlinear Time Series Analysis (traduzido para o português, Análise de Séries Temporais Não Lineares) (NLTA) são utilizadas com o objetivo de demonstrar a sua aplicabilidade para tratar das seguintes limitações relacionadas à detecção de *mudanças de conceito* no contexto do aprendizado supervisionado: i) a lacuna em relação à disponibilidade de métodos apropriados para lidar com a dependência temporal; e ii) o problema da rotulagem em fluxo de dados.

No caso da dependência temporal, técnicas da NLTA foram utilizadas para avaliar o fluxo de erros de uma maneira alternativa, ou seja, a partir da sua dinâmica, ao invés da não estacionaridade da distribuição da taxa de erro. Especificamente, por meio do Mapa de Bernoulli (DEVANEY, 1992; OTT, 2002), o fluxo de erros é mapeado em uma série temporal a fim de caracterizar a dinâmica dos erros entre as previsões. Consequentemente, considera-se a dependência temporal do fluxo de erros produzidos por previsões consecutivas ao longo do

tempo. Esse mapeamento possibilita descrever a informação sobre as relações temporais entre as predições em uma escala de mensuração apropriada para a aplicação de ferramentas derivadas da NLTA. Neste sentido, foram desenvolvidos três algoritmos fundamentados na NLTA, especificamente na Análise de Quantificação de Recorrência (RQA) (TRULLA et al., 1996), Entropia de Permutação (PE) (BANDT; POMPE, 2002; CAO et al., 2004), e Entropia Espectral (SE) (POWELL; PERCIVAL, 1979; INOUYE et al., 1991), com o objetivo de detectar *mudanças de conceito*. Vale ressaltar que, diferentemente das abordagens não supervisionado utilizadas em Vallim e Mello (2014), Costa, Rios e Mello (2016), Costa et al. (2017), nesta tese, técnicas da NLTA são adaptadas ao contexto supervisionado para avaliar o fluxo de erros das predições e identificar *mudanças de conceito*.

Para o caso da rotulagem, a aplicação de técnicas da NLTA têm por objetivo contornar as dificuldades envolvidas na disponibilidade de rótulos para a indução de um classificador base em domínios com *fluxos de dados* unidimensionais não rotulados. Seu fundamento está na dinâmica simbólica (BANDT; POMPE, 2002; AMIGÓ, 2010; LIND; MARCUS, 2021), a qual foi utilizada para viabilizar predições na evolução dos padrões ordinais de estados em um *espaço de fase* a fim de identificar mudanças na dinâmica do *fluxo de dados*. Com base numa *pseudo rotulação simbólica*, busca-se viabilizar tarefas de classificação usando padrões ordinais como rótulos, os quais são obtidos automaticamente a partir dos estados. Desse modo, levando em consideração a relação entre a dinâmica real e a simbólica, a indução de um classificador base considera o estado atual \mathbf{v}_i como atributos de entrada, e o padrão simbólico do estado consecutivo \mathbf{v}_{i+1} como o respectivo rótulo π_{i+1} . Presume-se que cada símbolo (ou rótulo) π_{i+1} possui alguma relação significativa com o seu respectivo estado \mathbf{v}_{i+1} . Conseqüentemente, um método foi desenvolvido com o objetivo de fornecer *pseudo rótulos* relacionados à dinâmica subjacente do *fluxo de dados*, com a finalidade de detectar *mudanças de conceito*.

Para investigar os pontos mencionados acima, vários conceitos da NLTA foram utilizados de modo a extrapolar o estado da arte delimitado por detectores fundamentados no pressuposto de observações i.i.d.. Isto posto, a seguinte hipótese foi levantada:

Hipótese. *Por lidar com a dinâmica do processo gerador do fluxo de erros das predições, de modo a considerar o viés da dependência temporal sobre a indução de modelos adaptativos supervisionados, ferramentas da NLTA apresentam desempenho superior em comparação com abordagens que avaliam a não estacionaridade da distribuição da taxa de erro e pressupõem observações i.i.d..*

Com o intuito de avaliar essa hipótese de pesquisa, diversos cenários com *mudanças de conceito* abruptas e graduais foram considerados com o objetivo de verificar a resiliência dos detectores em diferentes intervalos de tempo de transição entre conceitos, bem como a influência da dependência temporal no desempenho dos modelos adaptativos. Os métodos derivados da NLTA foram confrontados com métodos do estado da arte, fundamentados na *análise sequencial*, no *controle estatístico de processos* e na *avaliação de distribuições entre janelas* (GAMA et al., 2014), com base na acurácia e em métricas de qualidade da detecção de *mudança de conceito* (BASSEVILLE; NIKIFOROV, 1993; BIFET, 2017). Como parte da metodologia de investigação, experimentos foram projetados para responder as seguintes questões de pesquisa:

QP 1. *Os detectores fundamentados na NLTA apresentam desempenho superior ao estado da arte em termos de acurácia e detecção?*

QP 2. *Os classificadores acoplados a detectores de mudança de conceito aplicados a fluxos de dados com dependência temporal apresentam diferenças significativas em termos de acurácia em comparação a um classificador persistente (um classificador ingênuo que realiza a predição da instância atual com o rótulo da instância anterior (ŽLIOBAITĚ et al., 2015))?*

QP 3. *A rotulação de um fluxo de dados unidimensional por meio de padrões ordinais, definidos sobre os estados de um espaço de fase obtido a partir da aplicação do teorema de imersão de Takens (TAKENS, 1981), viabiliza a aplicação de modelos adaptativos supervisionados para fins de detecção da mudança de conceito de modo eficaz?*

QP 4. *Ao considerar a rotulação simbólica usando padrões ordinais dos estados, detectores acoplados a classificadores apresentam melhor desempenho em termos da detecção de mudanças na dinâmica do fluxo de dados em comparação ao estado da arte usados na avaliação direta dos dados brutos?*

1.3 ORGANIZAÇÃO DO TRABALHO

Os demais capítulos desta tese encontram-se estruturadas da seguinte forma:

Capítulo 2 - Fluxo de Dados e Mudança de Conceito: apresenta os fundamentos do aprendizado adaptativo relacionados à extração de informação em *fluxos de dados*, os métodos

do estado da arte usados para a detecção de *mudança de conceito* e discute o problema da dependência temporal nos modelos adaptativos supervisionados.

Capítulo 3 - Análise de Séries Temporais Não Lineares: apresenta os conceitos da análise de séries temporais não lineares, cujo objetivo é compreender a dinâmica por trás dos registros temporais e extrair informações implícitas quando o modelo de dados apresenta uma dinâmica determinística em algum espaço de estados. Também são discutidas as ferramentas utilizadas para avaliar mudanças na dinâmica de séries temporais não lineares.

Capítulo 4 - Proposta de Detectores e Pseudo-rotulação Simbólica: descreve as abordagens propostas para a detecção de *mudança de conceito*, bem como a estratégia para rotulação automática por meio da dinâmica simbólica. As abordagens de detecção abrangem algoritmos baseados na RQA, PE, e SE apresentadas no Capítulo 3. A estratégia de rotulação utiliza a relação entre os estados em um *espaço de fase* reconstruído, viabilizando a aplicação de classificadores a fluxo de dados unidimensionais com a finalidade de identificar *mudanças de conceito*.

Capítulo 5 - Análise Empírica e Estatística: apresenta os resultados e as respectivas análises referentes aos experimentos usados na avaliação dos métodos propostos no Capítulo 4. As seções seguintes descrevem as bases de dados, os experimentos e suas configurações, os testes estatísticos, e a metodologia utilizada para avaliar os resultados.

Capítulo 6 - Considerações Finais: discute as considerações finais sobre os principais pontos abordados nesta tese, incluindo as contribuições alcançadas e os trabalhos futuros.

2 FLUXO DE DADOS E MUDANÇA DE CONCEITO

Este capítulo apresenta os fundamentos do aprendizado adaptativo relacionados a fluxos contínuos de dados e à mudança de conceito (*concept drift*). As seções finais abordam métodos do estado da arte para a detecção de mudanças de conceito e o problema da dependência temporal sobre os modelos adaptativos.

2.1 FLUXO DE DADOS

Em cenários, tais como o tráfego TCP/IP, dados de sensores e dispositivos IoT, geolocalização, análise de atividade do uso cartões de crédito, interações em redes sociais, transações de comércio eletrônico (e-commerce), previsão de demanda de eletricidade, dentre muitos outros, uma quantidade massiva de dados trafega em alta velocidade, impondo restrições ao processamento dessa grande massa de dados (AGGARWAL, 2007; GAMA, 2010; BIFET et al., 2018). Entretanto, avanços tecnológicos recentes têm facilitado a coleta e análise desses dados na forma de fluxos contínuos (GAMA et al., 2014; BIFET et al., 2018).

Fluxos de dados são constituídos por uma sequência contínua de itens, denominados instâncias ou observações, os quais podem ser avaliados em blocos ou um a um (GAMA, 2010). Nesse contexto, o seu gerenciamento difere do gerenciamento de conjuntos de dados estáticos tradicionais em diversos aspectos (AGGARWAL, 2007; GAMA, 2010; BIFET et al., 2018). O grande volume e velocidade desses fluxos superam a capacidade de armazenamento e do tempo de processamento, inviabilizando avaliações em múltiplas passagens, comumente aplicadas a conjuntos de dados tradicionais (AGGARWAL, 2007). Normalmente, os conjuntos de dados tradicionais são previamente rotulados por especialistas, o que pode ser um obstáculo em muitos ambiente de fluxos contínuos de dados devido a rotulagem ser um processo demorado, caro, e em alguns casos inviável (DYER; CAPO; POLIKAR, 2014; COSTA; RIOS; MELLO, 2016). Outra dificuldade está na impossibilidade de controle em relação a ordem do processamento dos dados devido à sua natureza ubíqua. Além disso, um dos principais desafios refere-se à não garantia de dois pressupostos fundamentais aos modelos de aprendizagem: i) os dados devem ser identicamente e independentemente distribuídos (i.i.d); e ii) a estacionaridade da distribuição dos dados ao longo do processo de aprendizado (LUXBURG; SCHÖLKOPF, 2011; ŽLIOBAITĚ et al., 2015; GAMA et al., 2014).

Devido a essas particularidades, as seguintes características devem ser consideradas no aprendizado com fluxos contínuos de dados (GAMA, 2010; GAMA et al., 2014; KRAWCZYK et al., 2017; BIFET et al., 2018): i) as instâncias surgem sequencialmente, e não há controle sobre a ordem em que serão recebidas; ii) a taxa de transferência é relativamente alta em relação ao poder de processamento do sistema, de modo que apenas uma varredura é possível, ou seja, cada instância (ou bloco de instâncias) deve ser inspecionada em um única passagem; iii) após o processamento, as instâncias são descartada ou arquivadas temporariamente. Entretanto, é inviável armazenar todos os dados em memória, não sendo possível recuperá-los facilmente; iv) normalmente os dados não são estruturados, e a rotulagem pode ser impossível em alguns casos; e v) fluxos de dados são suscetíveis a mudanças, não sendo possível garantir a estacionaridade da distribuição das instâncias, ou que sejam identicamente e independentemente distribuídas (i.i.d);

Algoritmos de aprendizado *supervisionados*, *não supervisionados*, e *semisupervisionados* têm sido propostos para a mineração de dados na forma fluxos contínuos (AGGARWAL, 2007; GAMA, 2010; GAMA et al., 2014; DYER; CAPO; POLIKAR, 2014; ŽLIOBAITĖ et al., 2015; COSTA; RIOS; MELLO, 2016; GOMES et al., 2017; SETHI; KANTARDZIC, 2017; KRAWCZYK et al., 2017). Apesar das dificuldades relacionadas à rotulagem de dados produzidos em grande volume e velocidade, a maioria dos algoritmos são dedicados ao paradigma *supervisionado* (GAMA, 2010; GAMA et al., 2014; ŽLIOBAITĖ et al., 2015; GOMES et al., 2017; SETHI; KANTARDZIC, 2017; KRAWCZYK et al., 2017). Por outro lado, algoritmos *não supervisionados* são usados na identificação de alterações da distribuição dos dados, pois não requerem qualquer conhecimento prévio sobre os mesmos (AGGARWAL, 2007; GAMA, 2010; COSTA; RIOS; MELLO, 2016; SETHI; KANTARDZIC, 2017). Devido às dificuldades relacionadas ao aprendizado *supervisionados*, abordagens *semisupervisionadas* buscam lidar com o problema da escassez de rótulos em fluxo de dados (DYER; CAPO; POLIKAR, 2014). Independente do paradigma adotado, algoritmos aplicados ao contexto de fluxos contínuos de dados devem lidar com o processo de aprendizado de modo *online*, adequando o modelo a mudanças na distribuição dos dados, um fenômeno denominado *mudança de conceito* (TSYMBAL, 2004; GAMA et al., 2014).

2.1.1 Aprendizado adaptativo

Nos últimos anos, pesquisas no campo da aprendizagem de máquina têm se dedicado ao estudo e desenvolvimento de algoritmos de *aprendizado online* para a extração de informação

em fluxo contínuos de dados (AGGARWAL, 2007; GAMA, 2010; GAMA et al., 2014; ŽLIOBAITĖ et al., 2015; KRAWCZYK et al., 2017; GOMES et al., 2017; BARROS; SANTOS, 2019). Diferentemente do *aprendizado offline*, em que todos os dados de treinamento estão disponíveis, são i.i.d, e com uma distribuição estacionária, esses algoritmos devem manter modelos consistentes mesmo diante de alterações do processo gerador dos dados (FACELI et al., 2011). Algoritmos aplicados a fluxo de dados atualizam modelos de modo *adaptativo*, geralmente enfatizando os dados mais recentes, alterando o modelo a cada instância (ou lote de instâncias). Não existe predefinição entre dados de treinamento e de teste, as instâncias são apresentadas sequencialmente e rapidamente descartadas, e o modelo precisa ser atualizado imediatamente de acordo com os dados de entrada (BIFET, 2009; GAMA, 2010; GAMA et al., 2014).

No *aprendizado offline*, um modelo define uma hipótese indutiva a partir da inferência de uma distribuição de probabilidade conjunta desconhecida $P(\mathcal{X} \times \mathcal{Y})$ sobre o conjunto de combinações entre exemplos $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ e classes (ou rótulos) $\mathcal{Y} = \{y_1, \dots, y_m\}$. Para problemas de classificação, onde $P(\mathcal{X} \times \mathcal{Y})$ é estacionária, o aprendizado pode ser formalizado a partir da Statistical Learning Theory (traduzido para o português, Teoria do Aprendizado Estatístico) (SLT) (VAPNIK, 1999; LUXBURG; SCHÖLKOPF, 2011; VAPNIK, 2013). Nessa formulação, a aprendizagem é definida por meio da busca de uma relação funcional $f: \mathcal{X} \rightarrow \mathcal{Y}$, onde f define um classificador com o melhor desempenho possível. O classificador precisa ter acesso a alguns exemplos de treinamento $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n) \in \mathcal{X} \times \mathcal{Y}$, mas sem qualquer pressuposto sobre os espaços \mathcal{X} ou \mathcal{Y} . Entretanto, assume-se uma distribuição oculta de probabilidade conjunta $P(\mathcal{X} \times \mathcal{Y})$, e que os exemplos de treinamento (\mathbf{x}_i, y_i) são amostrados dessa distribuição de forma i.i.d..

Para a garantia do aprendizado, os seguintes pressupostos devem ser atendidos (LUXBURG; SCHÖLKOPF, 2011): i) nenhuma suposição deve ser feita sobre a distribuição de probabilidade conjunta P , ou seja, ela pode ser qualquer distribuição relacionada a $\mathcal{X} \times \mathcal{Y}$; ii) a distribuição de probabilidade conjunta subjacente não muda com o tempo, e nenhuma ordenação dos exemplos de treinamento deve ser assumida; iii) apesar da distribuição ser desconhecida no momento do treinamento do modelo, as suas propriedades podem ser estimadas indiretamente caso uma quantidade suficiente de exemplos para esta finalidade esteja disponível; iv) os exemplos de treinamento devem ser independentes e amostrados de maneira idêntica (i.i.d.); e v) devido a ruídos e sobreposição de classes, a relação entre rótulos e exemplos não segue uma regra determinística.

O *aprendizado offline* exige que todos os dados de treinamento estejam disponíveis, e

somente após a convergência da hipótese indutiva o modelo deve ser utilizado (FACELI et al., 2011; GAMA et al., 2014). Essa dificuldade tem motivado o desenvolvimento de abordagens de *aprendizado online incremental* (GAMA et al., 2014). Nesse modo de *aprendizado online*, os dados podem ser processados sequencialmente, e o modelo atualizado continuamente, em operação, à medida que mais exemplos são disponibilizados. Apesar dessa vantagem em relação ao *aprendizado offline*, a sua aplicação ainda presume cenários onde a distribuição conjunta $P(\mathcal{X} \times \mathcal{Y})$ deve ser estacionária.

Com a evolução da abordagem incremental surgem os algoritmos de *aprendizado online adaptativos* (GAMA et al., 2014). Nessa configuração de *aprendizado online*, presume-se que os padrões das distribuições dos dados podem sofrer mudanças ao longo do tempo devido a alterações no processo gerador do fluxo, exigindo algoritmos de aprendizado projetados para detectar e incorporar essas mudanças dinamicamente (HULTEN; SPENCER; DOMINGOS, 2001; SEBASTIÃO et al., 2010; GAMA, 2010; GAMA et al., 2014). Algoritmos adaptativos podem utilizar algum mecanismo para identificar mudanças inesperadas na evolução do processo gerador dos dados, e caso necessários, atualizar o modelo diante da *mudança de conceito*. De acordo com Gama et al. (2014), os métodos de adaptação podem ser classificados como *implícitos (cegos)* ou *explícitos (reativos)*. Métodos *implícitos* são abordagens passivas que atualizam o modelo periodicamente, independente da ocorrência de *mudança de conceito*. Desse modo, essas abordagens apenas adaptam o modelo a possíveis alterações na distribuição dos dados, em geral indiscriminadamente, sem necessariamente sinalizá-las previamente. Por outro lado, métodos *explícitos* são abordagens ativas que atualizam o classificador sempre que uma *mudança de conceito* for identificada. Esses métodos implementam algum mecanismo de detecção para sinalizar mudanças na distribuição dos dados a fim de atualizar rapidamente o modelo preditivo.

Abordagens promissoras relacionadas à adaptação implícita são os *comitês dinâmicos de classificadores*, os quais têm sido usado para lidar com fluxos contínuos de dados com relativo sucesso (KRAWCZYK et al., 2017; GOMES et al., 2017; BARROS; SANTOS, 2019). Nessa abordagem, um conjunto de modelos de decisão são combinados para produzir uma previsão global. Esses modelos serão adicionados e removidos dinamicamente de acordo com o respectivo desempenho individual. Métodos adaptativos explícitos são usados principalmente com um único classificador incremental, embora alguns comitês também os utilizem (KRAWCZYK et al., 2017; BARROS; SANTOS, 2018). Como esses métodos implementam mecanismos explícitos de detecção, eles informam o instante em que ocorre a *mudança de conceito*, possibilitando

avaliações do contexto da aplicação para alguma tomada de decisão.

2.1.2 Classificadores aplicados a fluxos contínuos de dados

O aprendizado preditivo a partir de fluxos contínuos de dados requer novos algoritmos com capacidade para atender as seguintes propriedades (DOMINGOS; HULTEN, 2001; GAMA et al., 2014): i) aprendizado incremental em tempo real; ii) processamento em tempo constante; iii) restrições de armazenamento e acesso de exemplos já processados; iv) capacidade de adaptar o modelo preditivo a *mudanças de conceito*. Embora esses requisitos não sejam restritos a modelos preditivos, apenas métodos adaptativos explícitos são avaliados no escopo desta tese.

Os classificadores incrementais Naïve Bayes (NB) (JOHN; LANGLEY, 1995) e Very Fast Decision Tree (VFDT) (HULTEN; SPENCER; DOMINGOS, 2001) são os mais frequentemente usados em experimentos e estão disponíveis no Framework MOA (BIFET et al., 2010; BIFET et al., 2018). Além desses, o *Classificador de Classe Majoritária* e o *Classificador Persistente* são sugeridos em Žliobaitė et al. (2015) como referências, principalmente em cenários com classes desbalanceadas, e onde os rótulos apresentem dependência temporal.

2.1.2.1 Very Fast Decision Tree

VFDT é um algoritmo de aprendizado incremental que ajusta dinamicamente o seu viés a cada exemplo avaliado (HULTEN; SPENCER; DOMINGOS, 2001; FACELI et al., 2011). A árvore de decisão define uma estrutura hierárquica composta por nós, ramos, e folhas (nós terminais). Cada nó interno contém um teste de um atributo, em que ramos representam partições dos seus respectivos valores. As folhas armazenam as estatísticas suficientes sobre os atributos dos exemplos e representam segmentações, ou seja, um ponto de classificação. A partir do nó raiz, localizado no topo da árvore, ramificações são construídas recursivamente, substituindo folhas por nós. O algoritmo utiliza o limite de Hoeffding (HOEFFDING, 1963) para definir uma quantidade mínima n de exemplos, e determinar a expansão de uma folha assim que houver evidência estatística suficiente.

O limite de Hoeffding garante que, com probabilidade $1 - \delta$, a média verdadeira \bar{r} de uma variável aleatória r , pertencente a um intervalo R , seja pelo menos $\bar{r} - \epsilon$, onde

$$\epsilon = \sqrt{\frac{R^2 \ln(\frac{1}{\delta})}{2n}}. \quad (2.1)$$

Seja $G(\cdot)$ a função de ganho de informação, e $\Delta G = G(\mathbf{x}_1) - G(\mathbf{x}_2)$ a diferença do ganho entre dois atributos \mathbf{x}_1 e \mathbf{x}_2 , respectivamente. Sendo o maior ganho $G(\mathbf{x}_1)$ e o segundo maior ganho $G(\mathbf{x}_2)$, se $\Delta G > \epsilon$ com n exemplos avaliados, o limite de Hoeffding assegura que, com probabilidade $1 - \delta$, a folha correspondente ao atributo \mathbf{x}_1 deve ser transformada em um nó de decisão. Caso, após considerar um conjunto de exemplos, a diferença do mérito entre os dois melhores testes de divisão não satisfaça o limite de Hoeffding, o algoritmo procede examinando mais exemplos (FACELI et al., 2011).

2.1.2.2 Naïve Bayes

No NB, a indução do modelo consiste apenas em computar as ocorrências de atributos e classes, conforme uma perspectiva probabilística (MITCHELL, 1997; DUDA; HART; STORK, 2012). Trata-se de um modelo estatístico generativo que tem por base uma hipótese de independência estrita, a qual assume que os atributos são condicionalmente independentes (JOHN; LANGLEY, 1995). Apesar dessa condição, o NB apresenta um bom desempenho em diversos domínios, mesmo com evidências da dependência entre os atributos dos exemplos (DOMINGOS; PAZZANI, 1997; MITCHELL, 1997).

Assim como diversos outros métodos de aprendizado, o NB está fundamentado no *teorema de Bayes* (MITCHELL, 1997; DUDA; HART; STORK, 2012). Métodos *bayesianos* assumem que a probabilidade da classe y_i , dado um conjunto de atributos \mathbf{x} , depende da relação entre a probabilidade *a priori* da classe, $P(y_i)$, a probabilidade dos exemplos pertencerem à classe, $P(\mathbf{x}|y_i)$, e a probabilidade de ocorrência desses exemplos $P(\mathbf{x}_j)$. Essa relação deriva do *teorema de Bayes*:

$$P(y_j|\mathbf{x}) = \frac{P(\mathbf{x}|y_j)P(y_j)}{P(\mathbf{x})}. \quad (2.2)$$

Para problemas de classificação, $P(y_j|\mathbf{x})$ define uma *função discriminante*, a qual separa o espaço de exemplos em classes distintas. Por meio da estimativa máxima de probabilidade *a posteriori* (do inglês, *Maximum A Posteriori* - MAP), o custo de associar \mathbf{x} à classe incorreta é minimizado quando \mathbf{x} é associado à classe y_j para a qual $P(y_j|\mathbf{x})$ é máxima (DUDA; HART; STORK, 2012). Formalmente, $P(\mathbf{x}|y_j)$ é designado por *verossimilhança*, e a hipótese de máxima verossimilhança pode ser expressa por $h_{MV} = \arg \max P(\mathbf{x}|y_i)$ (FACELI et al., 2011). Devido a dificuldades práticas em computar os fatores da Equação 2.2, assume-se a

premissa da independência condicional, considerando que cada um dos atributos contribui de forma independente para a estimação da classe, de modo que $P(\mathbf{x}|y_i)$ pode ser descrito como $P(x_1|y_i) \times P(x_2|y_i) \times \dots \times P(x_n|y_i)$ (FACELI et al., 2011). Assim, o classificador *naïve* Bayes pode ser expresso por meio da *função discriminante* e da Equação 2.2, onde $P(y_j|\mathbf{x})$, a probabilidade de um exemplo pertencer à classe y_i , pode ser descrita pela proporção:

$$P(y_j|\mathbf{x}) = \frac{P(y_j) \prod_{i=1}^n P(\mathbf{x}_i|y_j)}{P(\mathbf{x})}. \quad (2.3)$$

Como o denominador $P(\mathbf{x})$ é o mesmo para todas as classes, a Equação 2.3 pode ser simplificada considerando $P(y_j|\mathbf{x}) \propto P(y_j) \prod_{i=1}^n P(\mathbf{x}_i|y_j)$. O processo de indução ocorre de modo incremental, ou seja, não é necessário reprocessar todos os exemplos de treinamento a medida que novos exemplos são avaliados.

2.1.2.3 Classificadores ingênuos

De acordo com Žliobaitė et al. (2015), um *Classificador de Classe Majoritária* e um *Classificador Persistente* podem ser usados como referenciais para avaliar o desempenho da classificação em *fluxos contínuos de dados* usando diferentes informações sobre as estimativas de probabilidades dos rótulos. Esses *classificadores ingênuos* utilizam apenas os rótulos y_j para a classificação e são usados para estabelecer um desempenho mínimo significativo na avaliação de projeto de classificadores (ŽLIOBAITĖ et al., 2015).

O *Classificador de Classe Majoritária* sempre prevê $\hat{y} = \arg \max_i P(y = y_i)$ para qualquer observação \mathbf{x}_i , ou seja, a classe que tem a máxima probabilidade *a priori*. Sendo y_M a classe majoritária, a acurácia será igual à probabilidade *a priori* dessa classe. O *Classificador Persistente* prevê o mesmo rótulo observado previamente, ou seja, $\hat{y} = y_{t-1}$, para qualquer observação \mathbf{x}_i . Esse classificador leva em consideração a dependência temporal. Nesse caso, a estratégia é prever $\hat{y} = \arg \max_i P(y_t = i|y_{t-1})$ Žliobaitė et al. (2015).

Considerando que os exemplos são i.i.d., a acurácia do *Classificador de Classe Majoritária* será maior ou igual a do *Classificador Persistente*. Caso os rótulos contenham alguma dependência temporal, tal que $\sum_{y=1}^k P(y^i, y^{i-1}) > P(y_M)$, onde k é o número de classes e y_M é a classe majoritária, então a acurácia do *Classificador Persistente* será maior que a do *Classificador de Classe Majoritária*. Portanto, em cenários com classes desbalanceadas ou dependência temporal, ao menos esses dois classificadores devem ser levados em consideração (ŽLIOBAITĖ

et al., 2015).

Nesta tese, uma extensão do *Classificador Persistente* denominada *Classificador Temporalmente Aumentado* (do Inglês, *Temporally Augmented Classifier*) foi utilizada como referencial para indicar dependências temporais nos rótulos das instâncias de um *fluxo de dados*. Trata-se de uma heurística alternativa que pode incorporar ordens superiores de dependência temporal a um modelo preditivo h_t , onde $\hat{y}_t = h_t(\mathbf{x}_i, y_{t-1}, \dots, y_{t-\ell})$ (ŽLIOBAITĚ et al., 2015). Além dos atributos \mathbf{x}_i , essa heurística considera a ordem ℓ de dependência temporal dos rótulos $y_{t-1}, \dots, y_{t-\ell}$. Assim, quanto maior ℓ , mais longa é a dependência temporal considerada. Embora qualquer classificador base h_t possa ser utilizado, para o propósito de identificação da dependência temporal dos rótulos, foi mantido o classificador padrão VFDT.

2.2 MUDANÇA DE CONCEITO

A viabilidade de classificadores em generalizar hipóteses a partir de dados tem favorecido a priorização do paradigma de aprendizado supervisionado no cerne de vários problemas relacionados a extração de informação em fluxo de dados (BAENA-GARCÍA et al., 2006; TSYMBAL et al., 2008; BIFET, 2009; GAMA, 2010; GAMA et al., 2014; BARROS; SANTOS, 2018; BARROS; SANTOS, 2019; KRAWCZYK et al., 2017). No entanto, a capacidade de generalização dos classificadores depende da suposição de *estacionaridade*, ou seja, que os exemplos são gerados por uma mesma distribuição, de modo i.i.d. (VALIANT, 1984; ŽLIOBAITĚ, 2011). Essa suposição é frequentemente violada devido a mudanças na dinâmica de processos do mundo real, provocando a degradação da hipótese indutiva do modelo ao longo do tempo (GAMA, 2010).

Formalmente, *mudança de conceito* (do Inglês *concept drift*) pode ser definida como $\exists \mathbf{x} : p_{t_0}(\mathbf{x}, y) \neq p_{t_1}(\mathbf{x}, y)$, onde p_{t_0} e p_{t_1} representam distribuições de probabilidades conjuntas distintas, respectivamente nos instantes t_0 e t_1 (GAMA et al., 2014). Nessa formalização, a mudança pode ocorrer nas seguintes perspectivas: i) na distribuição de probabilidade *a priori* das classes $p(y)$; ii) na distribuição de probabilidade condicional da classe $p(\mathbf{x}|y)$; e iii) na distribuição de probabilidade *a posteriori* das classes $p(y|\mathbf{x})$. Como alterações na distribuição de probabilidade *a posteriori* afetam fronteiras de decisão de modelos preditores, essas alterações são consideradas *mudança de conceito reais* (*real concept drift*). Quando a fronteira de decisão permanece inalterada, mudanças em $p(\mathbf{x})$ ainda afetam $p(\mathbf{x}|y)$, sendo denominadas *mudanças de conceito virtuais* (*virtual concept drift*) (GAMA et al., 2014; ŽLIOBAITĚ, 2011).

Neste contexto, fluxo de dados são caracterizados pela evolução de uma sequência S_i de

exemplos $e_t = (\mathbf{x}_t, y_t)$, onde cada e_t é gerado por uma distribuição estacionária p_{θ_i} durante algum instante de tempo t . Sendo $S_1, S_2, \dots, S_k, \dots$, as sequências distintas de instâncias geradas pelas respectivas distribuições $p_{\theta_1}, p_{\theta_2}, \dots, p_{\theta_k}, \dots$, para $p_{\theta_i} \neq p_{\theta_{i+1}}$, a transição entre sequências consecutivas S_i e S_{i+1} define uma mudança de conceito. Caso a distribuição p_{θ_i} permaneça estacionária, pode-se aproximar uma hipótese indutiva com significativa precisão avaliando-se um número significativo de exemplos da sequência S_i geradas por p_{θ_i} . Na ocorrência de *mudanças de conceito*, instâncias de sequências consecutivas S_i e S_{i+1} surgem misturadas. Durante essa fase de transição, instâncias geradas pela distribuição $p_{\theta_{i+1}}$ serão identificadas como ruídos pelo modelo induzido na sequência S_i (GAMA, 2010).

O desafio é determinar o ponto exato de mudança diante do tipo *de mudança de conceito*. Mudanças abrupta indicam alterações súbitas na distribuição de probabilidade dos dados. Em geral, são mais fáceis de detectar e requerem poucas instâncias para a sua identificação. Por outro lado, mudanças graduais exigem um número maior de instâncias por serem mais difíceis de identificar em sua fase inicial. Em casos extremos, mudanças repentinas e frequentes prejudicam a inferência de uma hipótese indutiva, e conseqüentemente, a adaptação do modelo. Por outro lado, mudanças sutilmente graduais podem ser confundidas com estacionaridade (GAMA, 2010).

Existe um custo entre a eficácia da capacidade adaptativa para o rastreamento de mudanças e a estabilidade e robustez contra ruídos (GAMA, 2010). Detectores projetados para reagir rapidamente aos primeiros sinais de mudança são propensos a confundir ruídos nos dados com *mudanças de conceito* efetivas, apresentando comportamento instável e baixa acurácia preditiva em ambientes ruidosos. Por outro lado, modelos robustos a ruídos podem apresentar baixa sensibilidade, sinalizando apenas ocorrências de mudanças abruptas ou realizando adaptações tardias à nova distribuição do fluxo de dados (WIDMER, 1994). Modelos adaptativos devem diferenciar *ruídos* de *mudanças de conceito*, conseqüentemente, algoritmos para detecção dessas mudanças devem combinar robustez contra *ruídos* com sensibilidade à *mudanças de conceito* (GAMA, 2010).

2.2.1 Detectores de mudança de conceito

Em geral, quando os rótulos estão disponíveis, os métodos de detecção avaliam os resultados das predições do classificador e aplicam alguma regra de decisão sobre a distribuição estimada da predição de erro para tentar identificar se o modelo está desatualizado devido a

mudanças na distribuição dos dados. A partir de um fluxo de dados na forma de pares (x_i, y_i) , a *mudança de conceito* refere-se a alterações na distribuição de probabilidade conjunta $P(\mathbf{x}, y) = P(y|\mathbf{x}) \times P(\mathbf{x})$ sempre que $P(y|\mathbf{x})$ é afetada ao longo do tempo (GAMA et al., 2014). Essas mudanças, inerentes à dinâmica do fluxo de dados, podem ser rastreadas por meio dos erros das previsões, em que o classificador realiza uma previsão (\hat{y}_i) , que em seguida é comparada ao rótulo real (y_i) para decidir se a previsão estava correta $(\hat{y}_i = y_i)$ ou não $(\hat{y}_i \neq y_i)$. Nesse contexto, uma configuração comum é usar um detector para analisar o fluxo de erros de um classificador de forma que uma nova instância substitua o classificador anterior sempre que um alarme de mudança na distribuição de dados for emitido.

Diferentes abordagens explícitas têm sido propostas para o aprendizado a partir de fluxos de dados contendo *mudança de conceito* (SETHI; KANTARDZIC, 2017; BARROS; SANTOS, 2018; BARROS; SANTOS, 2019). Da mesma forma, os detectores utilizam estratégias distintas para avaliar o desempenho do classificador com o objetivo de atualizá-lo sempre que houver alteração na distribuição da taxa de erro. A não estacionaridade fundamenta a maioria das propostas apresentadas na literatura para a detecção de mudanças de conceito (SETHI; KANTARDZIC, 2017; KRAWCZYK et al., 2017; BARROS; SANTOS, 2018). De acordo com Gama et al. (2014), esses detectores sinalizam indícios de variações na distribuição dos dados avaliando a acurácia de classificadores por meio de técnicas derivadas da *análise sequencial*, do *controle estatístico de processos* e pelo monitoramento de distribuições usando janelas deslizantes.

2.2.1.1 Análise Sequencial

A Sequential Probability Ratio Test (SPRT) (WALD, 1973) é a base de vários detectores. Seja S_1^n uma sequência de exemplos, onde a subsequência S_1^w é gerada por uma distribuição desconhecida p_{θ_0} e a subsequência S_w^n é gerada por outra distribuição p_{θ_1} , para $1 < w < n$. A mudança da distribuição de probabilidade de p_{θ_0} para p_{θ_1} no exemplo w aumenta a expectativa de observar exemplos advindos da distribuição p_{θ_1} . Considerando as observações independentes, a hipótese nula H_0 é que S_w^n é gerado a partir da distribuição p_{θ_0} e a hipótese alternativa H_1 é que S_w^n é gerado por outra distribuição conhecida p_{θ_1} . A taxa de verossimilhança entre as duas distribuições é calculado por (BASSEVILLE; NIKIFOROV, 1993; GAMA et al., 2014):

$$\Lambda_w^n = \log \frac{P(S_w^n | p_{\theta_1})}{P(S_w^n | p_{\theta_0})} = \sum_{i=w}^n \log \frac{p_{\theta_1}[x_i]}{p_{\theta_0}[x_i]} = \Lambda_w^{n-1} + \log \frac{p_{\theta_1}[x_n]}{p_{\theta_0}[x_n]}. \quad (2.4)$$

Os limiares α e β são definidos a partir da taxa de erro alvo. Aceita-se H_0 , caso $\Lambda_w^{n-1} < \alpha$. Se $\Lambda_w^{n-1} > \beta$, H_1 será aceita. Quando $\alpha \leq \Lambda_w^{n-1} \leq \beta$, a decisão é adiada e o próximo exemplo x_n do fluxo é adicionado ao conjunto, conseqüentemente, Λ_w^n é calculado para uma nova comparação com os limiares. Uma mudança é detectada sempre que a hipótese alternativa H_1 é aceita.

No contexto do *aprendizado adaptativo*, técnicas baseadas em abordagens univariadas monitoram alguma métrica de performance para sinalizar mudanças nas distribuições (SETHI; KANTARDZIC, 2017). O Cumulative Sum (CUSUM) (PAGE, 1954) e o Page-Hinckley Test (PHT) (PAGE, 1954) são frequentemente usadas para esse propósito. CUSUM tem como entrada os resíduos de qualquer preditor, possui baixo custo de armazenamento e sua acurácia depende dos parâmetros δ e λ . A sinalização da mudança é disparada quando a média da sequência desvia significativamente de zero. O PHT foi proposto para detecção de mudanças no processamento de sinais, trata-se de uma variante do CUSUM para identificar desvios abruptos da média em processos Gaussianos (MOUSS et al., 2004).

A avaliação por meio do CUSUM pode ser definida pela Equação 2.5, considerando um deslocamento positivo em relação à média, onde $g_0 = 0$ define a estatística inicial da soma cumulativa superior.

$$g_t = \max(0, g_{t-1} + (\epsilon_t - \delta)). \quad (2.5)$$

O teste avalia se $g_t > \lambda$ para sinalizar alarmes e redefine $g_t = 0$ a cada detecção. Nesse caso, ϵ_t corresponde a observação da medida de performance no tempo corrente t , λ é o limiar de decisão e δ representa a magnitude tolerável de variação. Os parâmetros δ e λ controlam o *tradeoff* entre detecções imediatas contra o incremento de alarmes falsos, valores pequenos do parâmetro δ favorecem o aumento das sinalizações. Caso detecção ocorra no sentido negativo em relação a média, a Equação 2.5 deve ser trocada pela Equação 2.6 e o teste será avaliado para $g_t < -\lambda$, de modo que $g_0 = 0$ representa a estatística inicial da soma cumulativa inferior. Em geral, pode-se avaliar alterações em ambos os sentidos (BASSEVILLE; NIKIFOROV, 1993).

$$g_t = \min(0, g_{t-1} - (\epsilon_t - \delta)). \quad (2.6)$$

O Page-Hinckley Test (PHT) é definido pela Equação 2.7, onde o método atualiza o valor mínimo para m_T , denotado como M_T e obtido pela expressão: $M_T = \min(m_t, t = 1 \dots T)$. A variável m_T representa a diferença acumulada entre os valores observados e a respectiva

média até o instante T .

$$m_T = \sum_{t=1}^T (\epsilon_t - \bar{\epsilon}_T - \delta), \bar{\epsilon}_T = \frac{1}{T} \sum_{t=1}^T \epsilon_t. \quad (2.7)$$

A detecção é determinada por meio dos parâmetros δ e λ de forma semelhante ao CUSUM, de modo que diferenças significativas dada por $PH_T = m_T - M_T$, sinalizam mudanças abruptas quando $PH_T > \lambda$. Valores elevados para o parâmetro λ reduzem a sinalização de alarmes falsos nas identificações, mas favorecem a taxa de falsos negativos. Ambos os testes têm baixo custo de armazenamento, sendo também usados de modo incremental no rastreamento de medidas de performance em algoritmos de aprendizado *online*.

Outra proposta é o Linear Four Rates (LFR) (WANG; ABRAHAM, 2015). A abordagem fundamenta a detecção de mudança de conceito por meio da avaliação de medidas de performance numa matriz de contingência de modo a lidar com fluxos de dados desbalanceados (SETHI; KANTARDZIC, 2017).

2.2.1.2 Controle Estatístico de Processos

Com base nos pressupostos do *framework* Probably Approximately Correct (PAC) (VALIANT, 1984), assume-se que a taxa de erro do modelo decresce com o aumento do número de instâncias, caso a distribuição dos dados mantenha-se estacionária (HAUSSLER, 1990). No contexto da detecção de mudança de conceito, técnicas estatísticas para monitoramento de controle de qualidade (Statistical Process Control (traduzido para o português, Controle Estatístico de Processos) (SPC)) (MACGREGOR; KOURTI, 1995; WOODALL; MONTGOMERY, 1999) são usadas no rastreamento da taxa de erro de modelos durante o aprendizado *online* (GAMA et al., 2014). O fundamento está na interpretação da evolução do aprendizado como um processo de controle de qualidade. Desse modo, indícios da mudança de conceito são identificados quando incrementos significativos da taxa de erro violam os pressupostos do PAC (SETHI; KANTARDZIC, 2017).

Formalmente, os erros de predição para cada instância (\mathbf{x}_t, y_t) de uma sequência de exemplos S_i podem ser interpretados como um processo de Bernoulli. Assim, assume-se a taxa de erro em S_i como uma variável aleatória descrita por uma distribuição Binomial. Para cada instância, a taxa de erro será a probabilidade p_t de obter um erro na predição ($\hat{y}_t \neq y_t$) com desvio padrão $\sigma_t = \sqrt{\frac{p_t(1-p_t)}{t}}$. Detectores de mudança de conceito gerenciam a cada instân-

cia, p_{min} , a proporção mínima de erros, e σ_{min} , a variação mínima da proporção de erros. Sempre que $p_t + \sigma_t < p_{min} + \sigma_{min}$, as proporções mínimas são atualizadas, então $p_{min} = p_t$ e $\sigma_{min} = \sigma_t$ (GAMA et al., 2004; GAMA, 2010).

Considerando um contexto estacionário, fundamenta-se a convergência da distribuição Binomial em uma distribuição Normal com o aumento significativo do tamanho da amostra, e conseqüentemente, de acordo com o PAC, o decaimento da taxa de erro do preditor. Assumindo que a distribuição permanecerá a mesma até a ocorrência da mudança de conceito, o intervalo de confiança $1 - \alpha/2$ para p_t quando $n > 30$ será aproximadamente $p_t \pm z \times \sigma_t$ (GAMA et al., 2014). A partir do parâmetro z , o nível de confiança para sinalizar alerta define os seguintes estados:

- *In-Control*: enquanto $p_t + \sigma_t < p_{min} + 2\sigma_{min}$. Indica estabilidade na taxa de erro, conseqüentemente a distribuição permanece estacionária;
- *warning*: ocorre para $p_{min} + 2\sigma_{min} \leq p_t + \sigma_t < p_{min} + 3\sigma_{min}$. Com nível de confiança de 95%, a mudança da distribuição pode ser *persistente*; e
- *drift*: definido quando $p_t + \sigma_t > p_{min} + 3\sigma_{min}$. Com nível de confiança de 99%, admite-se que as observações atuais são geradas por uma nova distribuição.

Essa avaliação permite monitorar a velocidade da mudança por meio do intervalo temporal entre a sinalização de *warning* e o *drift*, e são a base de detectores como o Drift Detection Method (DDM) (GAMA et al., 2014).

Uma abordagem similar, a Exponentially Weighted Moving Average (EWMA) (ROBERTS, 1959), foi proposta para avaliar a evolução da média de uma sequência de variáveis aleatórias unidimensionais x_1, x_2, \dots, x_n considerando o decaimento exponencial de observações anteriores. A EWMA também sinaliza variações da média em termos do número de desvios padrões.

O algoritmo computa a estimativa atual da média μ_t por meio da recorrência definida na Equação 2.8. A recorrência representa o decaimento exponencial das observações, tendo x_t como medida corrente, para $t > 0$. Assumindo $\mu_t = \mu_0$ e $\mu_t = \mu_c$, respectivamente, antes e após a ocorrência da *mudança de conceito*, a flutuação de Z_t identifica uma alteração de μ_0 até uma nova estimativa da média μ_c (ROSS et al., 2012).

$$Z_t = (1 - \lambda)Z_{t-1} + \lambda x_t, Z_0 = \mu_0. \quad (2.8)$$

Independentemente da distribuição, o valor esperado do estimador Z_t é definido por $\mu_{Z_t} = \mu_t$ e seu desvio padrão pela Equação 2.9. O sinal de detecção ocorre quando $Z_t > \mu_0 + L\sigma_{Z_t}$. O parâmetro de controle L determina o quanto Z_t diverge de μ_0 para uma mudança ser declarada. A suavização do decaimento exponencial para a estimativa da média e do desvio padrão é definida pelo parâmetro λ .

$$\sigma_{Z_t} = \sqrt{\frac{\lambda}{2-\lambda}(1 - (1-\lambda)^{2t})}\sigma_x. \quad (2.9)$$

Como no DDM, a distribuição da taxa de erro é interpretada como um processo de Bernoulli (YEH et al., 2008). Assim, as probabilidades $p_0 = \mu_0$ e $p_c = \mu_c$ definem, respectivamente, as taxas de erro antes e após a mudança. Entretanto, no cenário de fluxo de dados, p_0 não é conhecida *a priori* e deve ser estimada juntamente com σ_0 . Para contornar essa restrição, o EWMA for Concept Drift Detection (ECDD) (ROSS et al., 2012) introduz o estimador definido na Equação 2.10. Diferentemente de Z_t , o estimador $\hat{p}_{0,t}$ não pondera as observações mais recentes com maior significância. Assim, esse estimador é menos sensível a mudanças em p_0 e portanto converge mais lentamente para o novo valor p_c (ROSS et al., 2012).

$$\hat{p}_{0,t} = \frac{1}{t} \sum_{i=1}^t x_i = \frac{t-1}{t} \hat{p}_{0,t-1} + \frac{1}{t} x_t. \quad (2.10)$$

A partir da definição do estimador $\hat{p}_{0,t}$, define-se com a propriedade elementar da distribuição de Bernoulli, a variância do estimador na Equação 2.11 (YEH et al., 2008). Desse modo, a sinalização da mudança passa a ser definida quando $Z_t > \hat{p}_{0,t} + L\sigma_{Z_t}$, onde parâmetro de controle L está relacionado com a taxa média de detecções de falsos positivos (Average Running Length (traduzido para o português, Comprimento Médio de Sequência) (ARL)).

$$\sigma_{Z_t}^2 = \sqrt{\frac{\lambda}{2-\lambda}(1 - (1-\lambda)^{2t})} \hat{p}_{0,t}(1 - \hat{p}_{0,t}). \quad (2.11)$$

Um valor apropriado para L pode ser obtido via simulação de Monte Carlo (VERDIER; HILGERT; VILA, 2008; ROSS et al., 2012). Entretanto, devido ao custo computacional, uma adaptação da solução direcionada ao CUSUM (SPARKS, 2000) foi proposta em Ross et al. (2012). Trata-se de uma aproximação polinomial dada em função de p_0 e ARL por meio de técnicas convencionais de regressão para estimar os coeficientes de uma lista de polinômios. Assim, fornecendo um valor ARL desejado, o respectivo valor para L pode ser encontrado com custo $\mathcal{O}(1)$.

Vários detectores estão fundamentados no *controle estatístico de processos*. Uma extensão do DDM para contextos de mudanças graduais, o Early Drift Detection Method (EDDM) (BAENA-GARCÍA et al., 2006), monitora a distância entre erros de classificação consecutivos em vez da taxa de erro para sinalizar a detecção. Assim, enquanto a distribuição dos dados permanecer estacionária, as distâncias entre erros consecutivos tendem a aumentar e detecções são sinalizadas caso as distâncias diminuam significativamente. Outro método relacionado é o Statistical Test of Equal Proportions (STEPD) (NISHIDA; YAMAUCHI, 2007). O fundamento do método está na comparação da acurácia das predições de uma janela de observações recentes contra a acurácia geral, obtida de observações desde a última sinalização de detecção, tendo por base um teste de hipóteses de iguais proporções com correção de continuidade. Assume-se que as precisões das predições devam ser as mesmas caso a distribuição da acurácia permaneça estacionária, do contrário, ocorrerá uma diminuição significativa na precisão, indicando uma alteração do conceito.

Melhorias recentes em relação aos métodos de detecção DDM e STEPD são apresentadas nas propostas do Reactive Drift Detection Method (RDDM) (BARROS et al., 2017), Wilcoxon Rank Sum Test Drift Detector (WSTD) (BARROS; HIDALGO; CABRAL, 2018) e no Fisher Test Drift Detector (FTDD) (CABRAL; BARROS, 2018).

Com base no DDM, o RDDM (BARROS et al., 2017) busca reduzir periodicamente o número de instâncias em longos períodos de estacionaridade para lidar com um problema de perda de desempenho, característico do DDM. Supõe-se que essa queda seja devido à diminuição da sensibilidade como resultado da indução de um grande número de instâncias pertencentes a um mesmo conceito. Com a finalidade de reduzir esse *overfitting*, utiliza-se um mecanismo explícito para descartar periodicamente instâncias mais antigas de conceitos muito longos e o recálculo das estatísticas de um conjunto mínimo das instâncias mais recentes presentes no conceito.

O WSTD (BARROS; HIDALGO; CABRAL, 2018) foi inspirado no STEPD. Trata-se de uma proposta fundamentada em uma eficiente implementação do teste da soma de postos de *Wilcoxon* (WILCOXON, 1945). Esse teste não paramétrico é aplicável quando as amostras são independentes e as medições estão ao menos em escala ordinal. As principais diferenças dos dois métodos são o tipo de teste estatístico usado para comparar as amostras e o limite de observações obtidas desde a última sinalização de detecção. Como propostas de métodos estatisticamente precisos, independentemente do tamanho de janelas de observações, o FTDD (CABRAL; BARROS, 2018) e seus derivados, o Fisher Proportions Drift Detector (FPDD) e o

Fisher Squared Drift Detector (FSDD), também estão fundamentado no STEPD. O FPDD utiliza o teste exato de *Fisher* (FISHER, 1922) quando as janelas de observações (antigas e recentes) contém um número pequeno de predições. Caso contrário, também será usado o teste de iguais proporções de acordo com o STEPD. De modo semelhante, o FSDD também aplica o teste exato de *Fisher* quando as janelas de observações contém um número pequeno de predições, mas utiliza o teste *chi-quadrado* no lugar do teste de iguais proporções. Por fim, o FTDD utiliza o teste exato de *Fisher* em todos os cenários.

Como esses detectores provêm do DDM, os seus parâmetros de configurações são semelhantes. Além disso, todos permitem monitorar a velocidade da mudança por meio do intervalo temporal entre a sinalização de *warning* e o *drift*.

2.2.1.3 Monitoramento em Janelas de Distribuições

No contexto de fluxo contínuos de dados, a detecção da mudanças de conceito por meio da comparação de distribuições foi introduzida por (KIFER; BEN-DAVID; GEHRKE, 2004). Trata-se de um abordagem robusta a ruídos e com identificação precisa do ponto de mudança. A desvantagem em relação à *análise sequencial* e ao *Controle Estatístico de Processos*, onde o modelo avalia uma única instância a cada incremento temporal, está na necessidade de armazenar e avaliar predições em duas janelas de instâncias (*chunks*) para disparar a detecção. Em geral, confronta-se uma janela de referência fixa para resumo de características de instâncias passadas contra uma janela deslizante para resumo de informações de instâncias recentes. As distribuições das janelas são comparadas por meio de testes de hipóteses, considera-se a hipótese alternativa para indicar mudança de conceito (GAMA et al., 2014).

Uma proposta representativa é o Adaptive Windowing (ADWIN) (BIFET; GAVALDÀ, 2007), uma metodologia de estimação e detecção para o problema de monitoramento por meio do valor esperado μ em fluxos de *bits* ou valores reais. O ADWIN mantém um histograma das observações considerando uma janela W de tamanho ajustável, estatisticamente consistente com a hipótese de estacionaridade da distribuição dentro da janela, por meio da avaliação de μ . A manutenção da estacionaridade a cada observação aumenta o tamanho da janela, o contrário, induz o descarte de uma subjanela de dados mais antigos. Assim, a variação no tamanho da janela W evidencia mudança na distribuição de probabilidade dos dados.

Pela formalização da metodologia (BIFET; GAVALDÀ, 2007), dado uma sequência de valores reais $x_1, x_2, x_3, \dots, x_t$, com a restrição $x_t \in [0, 1]$. Assumindo o valor esperado μ_t para os

valores x_t gerados independentemente por uma distribuição p_t , o ADWIN mantém a janela ajustável W como a concatenação de duas partições $W_0 \cdot W_1$. As partições W_0 e W_1 representam subjanelas das observações antigas e atuais, respectivamente. Sendo $\tilde{\mu}_{W_0}$ e $\tilde{\mu}_{W_1}$ as estimativas do valor esperado obtidas em W_0 e W_1 , a diferença absoluta das estimativas em relação a um limiar identifica a mudança de conceito. Assim, $|\tilde{\mu}_{W_0} - \tilde{\mu}_{W_1}| \geq \epsilon_{cut}$ determina o descarte de W_0 de acordo com o teorema da metodologia. Uma rigorosa garantia de performance é obtida por meio do limite Hoeffding (SCHMIDT; SIEGEL; SRINIVASAN, 1995) para a estimação do limiar, desse modo:

$$\epsilon_{cut} = \sqrt{\frac{1}{2m} \times \ln \frac{4|W|}{\delta'}}, m = \frac{1}{\frac{1}{|W_0|} + \frac{1}{|W_1|}}. \quad (2.12)$$

Onde m é a media harmônica, $|W|$, $|W_0|$ e $|W_1|$, são os tamanhos das janelas. O ADWIN adapta-se automaticamente à taxa de mudança, não assume qualquer distribuição particular dos dados e possui como único parâmetro o nível de confiança $\delta \in (0, 1)$.

O Sequential Hypothesis Testing Drift Detector 1 (SEQDRIFT), e a sua versão mais recente Sequential Hypothesis Testing Drift Detector 2 (SEQD2) (SRIPRAKAS; PEARS; KOH, 2013; PEARS; SAKTHITHASAN; KOH, 2014), baseiam-se na *desigualdade de Bernstein* (BERNSTEIN, 1946), e propõe melhorar alguns aspectos do ADWIN, reduzindo a complexidade computacional, a taxa de falsos positivos e aumentando a velocidade das detecções. O SEQD2 utiliza uma eficiente estratégia de amostragem a partir dos dados de entrada em dois repositórios de instâncias. Os repositórios são denominados *left* e *right*. O primeiro, combina instâncias novas e antigas por meio da estratégia de amostragem (VITTER, 1985), o segundo, armazena apenas as novas instâncias. A detecção é sinalizada comparando-se as médias dos dois repositórios com base em um limite superior, usando a *desigualdade de Bernstein*, para identificar se as amostras são de uma mesma população. Sua limitação está no pressuposto que as observações seguem uma distribuição normal.

Outra proposta com base no ADWIN, o SEED Drift Detector (SEED) (HUANG et al., 2014), também compara duas partições dentro de uma janela W e utiliza um teste semelhante, com correção de Bonferroni, para detectar a mudança de conceito. No SEED, os pontos de mudança candidatos são os pontos iniciais e finais em cada lote de instâncias. Lotes adjacentes são examinados e agrupados se forem considerados suficientemente semelhantes. Essa compactação de lotes remove os pontos de mudança candidatos que têm menor probabilidade de serem verdadeiros pontos de mudança de conceito. A justificativa é que lotes maiores garantem

estimativas mais estáveis das probabilidades em comparação com as estimativas dos blocos originais.

Com base na *desigualdade de Hoeffding* (HOEFFDING, 1963), Frías-Blanco et al. (2015) propuseram o Drift Detection Method based on the Hoeffding's inequality (HDDM), mais especificamente os detectores $HDDM_A$ e o $HDDM_W$. Ambos definem intervalos de confiança para comparar a média da acurácia do classificador por meio dos testes estatísticos A-Test e W-Test, respectivamente. O $HDDM_A$ emprega uma média móvel, enquanto o $HDDM_W$ usa uma média móvel ponderada de modo a enfatizar as instâncias mais recentes. Os autores afirmam que o $HDDM_A$ é ideal para detectar mudanças abruptas, e o $HDDM_W$ é mais adequado para identificar mudanças graduais.

Outro detector fundamentado na *desigualdade de Hoeffding* é o Fast Hoeffding Drift Detection Method (FHDDM) (PESARANGHADER; VIKTOR, 2016). A ideia é sinalizar uma *mudança de conceito* quando uma diferença significativa $\Delta\mu = \mu^m - \mu^t \geq \varepsilon_d$ entre a maior média μ^m das avaliações das predições observadas até o momento e a média μ^t das avaliações atuais, ambas calculadas em uma janela deslizante de tamanho n , ultrapassar um limiar ε_d . A partir de uma probabilidade de erro previamente definida δ_d , a desigualdade de Hoeffding indica que uma *mudança de conceito* ocorre se $\Delta\mu \geq \varepsilon_d$, onde

$$\varepsilon_d = \sqrt{\frac{1}{2n} \ln\left(\frac{1}{\delta_d}\right)}. \quad (2.13)$$

A família de detectores McDiarmid Drift Detection Method (MDDM) (PESARANGHADER; VIKTOR; PAQUET, 2018) foi proposta como uma extensão do FHDDM, a qual utiliza a *desigualdade de McDiarmid* (MCDIARMID, 1989) no lugar da *desigualdade de Hoeffding*. A ideia desses métodos é que ponderando os resultados das predições da janela deslizante, e colocando mais ênfase nos elementos mais recentes, a *mudança de conceito* pode ser detectada mais rapidamente e com mais eficiência. De modo semelhante ao FHDDM, μ_w^t define a média ponderada da avaliação de n predições, no tempo t , e μ_w^m é a maior média ponderada observada atualmente. Dado o intervalo de confiança δ_w , a *desigualdade de McDiarmid* sugere uma *mudança de conceito* se $\Delta\mu_w = \mu_w^m - \mu_w^t$, onde

$$\varepsilon_w = \sqrt{\frac{\sum_{i=1}^n v_i^2}{2} \ln\left(\frac{1}{\delta_w}\right)}, \text{ e } v_i = \frac{w_i}{\sum_{i=1}^n w_i}. \quad (2.14)$$

Três esquemas de ponderação são considerados, um esquema aritmético, definido por $w_i = 1 + (i - 1) \times d$, onde $d \geq 0$ é a diferença entre dois pesos consecutivos; um esquema

geométrico, definido por $w_i = r^{(i-1)}$, onde $r \geq 1$ é a razão entre dois pesos consecutivos; e o esquema de Euler que é definido por $r = e^\lambda$, onde $\lambda \geq 0$. Esses esquemas definem, respectivamente, os detectores $MDDM_A$, $MDDM_G$ e o $MDDM_E$. Os autores afirmam que esses detectores reduzem atrasos na detecção, com poucos falsos negativos, enquanto mantém a acurácia elevada.

2.3 DEPENDÊNCIA TEMPORAL

As causas da *mudança de conceito* podem ser diversas, pode-se distinguir fatores de alterações no contexto da aprendizagem a partir de mudanças nos atributos observados, ou devido a influências de variáveis ocultas (GAMA, 2010). Em problemas do mundo real, propriedades importantes do domínio da aplicação podem não ser capturadas pelos atributos (HARRIES; SAMMUT; HORN, 1998), conseqüentemente, a evolução dessas propriedades ocultas ao longo do tempo pode influenciar a dinâmica da aprendizagem. Um dos fatores de impacto na identificação da mudança de conceito é a dependência temporal (ŽLIOBAITĚ et al., 2015). Conhecida como autocorrelação, trata-se de uma característica investigada tradicionalmente na *análise de séries temporais* (BOX et al., 2015). A dependência temporal restringe a aplicabilidade de muitos testes estatísticos fundamentados na suposição de observações i.i.d.. É importante enfatizar que quando há dependência temporal nos erros das predições, as garantias teóricas dos detectores levantados na Seção 2.2.1 não são asseguradas (ŽLIOBAITĚ et al., 2015).

Apesar da eficiência de muitos detectores com base na avaliação da distribuição da taxa de erro, estudos empíricos mostraram que a taxa de erro pode ser influenciada pela dependência temporal (ŽLIOBAITĚ et al., 2015; BIFET, 2017). Este efeito está relacionado à propagação da acurácia por causa das sucessivas atualizações incrementais do modelo (BIFET, 2017). Considerando que uma predição precisa ser realizada para cada observação ao longo do tempo, uma sequência de modelos h_1, \dots, h_i, \dots é construída de forma incremental, levando em consideração todos ou um subconjunto do modelo anterior h_{i-1} , observações anteriores $\mathbf{x}_1, \dots, \mathbf{x}_{i-1}$, e rótulos verdadeiros y_1, \dots, y_{i-1} , definidos como $h_i = f(h_{i-1}, \mathbf{x}_1, \dots, \mathbf{x}_{i-1}, y_1, \dots, y_{i-1})$, onde f é o algoritmo para atualização do modelo (ŽLIOBAITĚ et al., 2015). Desse modo, informações relevantes da dependência temporal entre os erros das predições são desconsideradas quando apenas a não estacionariedade da distribuição da taxa de erro é avaliada. Caso os rótulos das instâncias apresentem dependência temporal, o fluxo de erros resultante das predições dos rótulos também será afetado (as observações não serão i.i.d.). Logo, as garantias teóricas dos

detectores de *mudanças de conceito* serão violadas (ŽLIOBAITĖ et al., 2015).

Recentemente, novas abordagens foram propostas para detecção de *mudança de conceito virtuais* em *fluxos de dados* contendo dependência temporal, considerando ferramentas de sistemas dinâmicos (VALLIM; MELLO, 2014; COSTA; RIOS; MELLO, 2016; COSTA et al., 2017). Esses métodos preservam relações não lineares e são aplicados a *fluxo de dados* sujeitos a dependência temporal e a ruídos para análise de mudanças em sua dinâmica. Entretanto, métodos para detecção de mudança de conceito aplicados a *fluxo de dados* no contexto do *aprendizado adaptativo supervisionado* com o uso de técnicas da NLTA ainda são raros. Esses estudos sugerem que a *mudança de conceito* pode ser caracterizada por relações temporais entre os dados, e a identificação dessas mudanças pode ser alcançada ao avaliar as dependências do dados em um espaço de fase (TAKENS, 1981), bem como, por meio da divergência entre gráficos de espectro de potência (VALLIM; MELLO, 2014). Essas abordagens levam em consideração possíveis dependências temporais, assumindo que as observações não são necessariamente i.i.d..

Técnicas da NLTA são largamente usadas para a identificação de mudanças em *fluxos de dados* subjacentes a diversos fenômenos biológicos e naturais sujeitos a ruídos, a não estacionaridade, a apresentarem não linearidade e dependência temporal (SCHREIBER, 1999; MARWAN et al., 2007; KANTZ; SCHREIBER, 2004; STAM, 2005; DONNER et al., 2010; DONNER et al., 2011). Essas técnicas fundamentam os métodos propostos nesta tese e seus conceitos serão discutidos no próximo capítulo.

2.4 RESUMO DO CAPÍTULO

Este capítulo apresentou os fundamentos do aprendizado adaptativo relacionados a fluxos contínuos de dados e à *mudança de conceito*. As seções finais abordaram métodos tradicionais consolidados na literatura e baseados na *análise sequencial*, no *controle estatístico de processos* e pela *avaliação de distribuições entre janelas*. Apesar das evidências a respeito da dependência temporal, abordagens aplicadas ao aprendizado em fluxos contínuos pressupõe dados i.i.d.. Estudos recentes sugerem que a avaliação do efeito da dependência temporal sobre os modelos adaptativos podem vir a contribuir no ganho de informação para o desenvolvimento de métodos mais apropriados para a detecção de *mudanças de conceito*, contornando as restrições teóricas dos métodos do estado da arte.

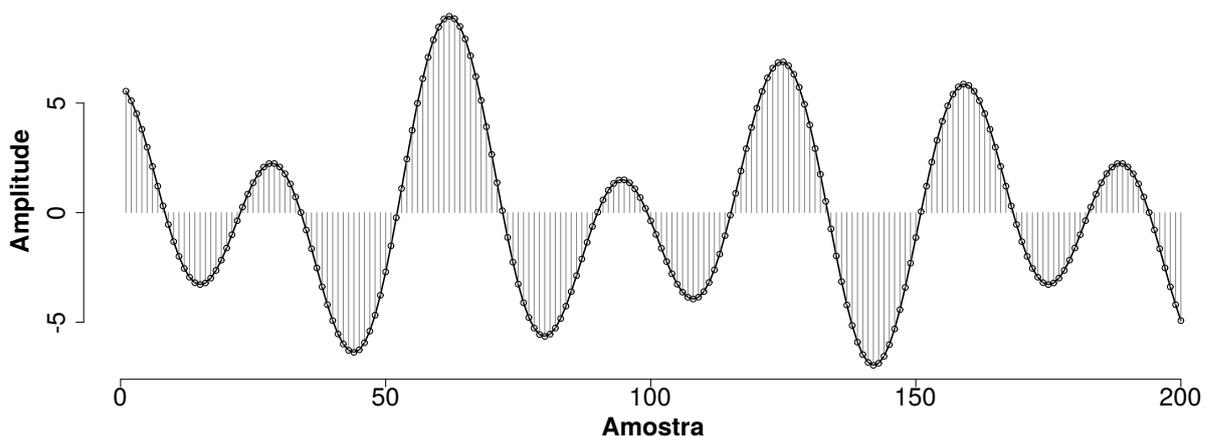
3 ANÁLISE DE SÉRIES TEMPORAIS NÃO LINEARES

Este capítulo apresenta os conceitos da análise de séries temporais não lineares fundamentada na teoria de Sistemas Dinâmicos, cujo objetivo é compreender a dinâmica por trás dos registros temporais e extrair informações a partir de um espaço de fase. As seções seguintes descrevem a coleção de ideias e técnicas para a análise de séries temporais e os métodos para identificar e quantificar as mudanças na dinâmica do processo subjacente aos dados.

3.1 SÉRIES TEMPORAIS

Uma série temporal é uma sequência de observações $x_1, x_2, \dots, x_t, \dots$ que descreve a evolução de alguma característica relacionada ao fenômeno de interesse (MORETTI; TOLOI, 2006). Usualmente, as observações $x_t \in \mathbb{R}$ são obtidas considerando intervalos regulares¹ (ilustrado na Figura 1), ordenadas de acordo com um índice cronológico t . Nesta tese, considera-se apenas séries temporais discretas ($t \in \mathbb{N}$) com um número finito de N observações, definidas como $\mathbf{s} = \{x_t\}$, onde $t = 0, 1, 2, 3, \dots, N - 1$.

Figura 1 – Série temporal discretizada (pontos marcados por círculos com segmentos verticais) obtida de uma série temporal contínua (linha contínua preta).



Fonte: Elaborada pelo autor (2021).

Um dos objetivos da análise de séries temporais é compreender a dinâmica do fenômeno oculto por trás de observações obtidas ao longo do tempo (KANTZ; SCHREIBER, 2004; BRADLEY; KANTZ, 2015). As abordagens mais utilizadas empregam modelos lineares para inter-

¹ discretização da série contínua $x(t) = \cos(2\pi 20t + \pi/2) + 3\cos(2\pi 30t - \pi/3) + 5\cos(2\pi 50t + \pi/5)$

pretar estruturas regulares considerando todo o comportamento irregular da dinâmica como uma interferência aleatória (KANTZ; SCHREIBER, 2004; BOX et al., 2015).

Tradicionalmente, as irregularidades e periodicidades de fenômenos temporais são analisadas no domínio temporal ou espectral (KANTZ; SCHREIBER, 2004; SHUMWAY; STOFFER, 2017). Nesse sentido, busca-se avaliar a dinâmica do processo subjacente para fins de classificação, descrição, modelagem e previsão, assumindo um viés predominantemente estocástico ou determinístico (SCHREIBER, 1999; KANTZ; SCHREIBER, 2004; PALIT; POPOVIC, 2006; BOX et al., 2015).

Independente do viés adotado, dados coletados de fenômenos do mundo real normalmente tendem a apresentar uma mistura de componentes aleatórios e determinísticos associados a alguma ordem de correlação serial (SCHREIBER, 1999; RIOS; MELLO, 2016). Considerando essa suposição, abordagens no domínio temporal tentam descrever a dinâmica de séries temporais por meio de ferramentas estatísticas, como os modelos Autoregressive Integrated Moving Average (traduzido para o português, Modelo Auto-regressivo Integrado de Médias Móveis) (ARIMA) (BOX et al., 2015), ou utilizando ferramentas físicas e matemáticas fundamentadas na NLTA (SCHREIBER, 1999; KANTZ; SCHREIBER, 2004).

As abordagens estatísticas e espectrais constituem um paradigma linear, as quais assumem que pequenas causas devem provocar pequenos efeitos na dinâmica intrínseca do sistema (KANTZ; SCHREIBER, 2004). Por outro lado, abordagens físicas e matemáticas correspondem a um paradigma não linear, em que tanto a estrutura regular quanto o comportamento irregular do fenômeno são determinados por regras não lineares (SCHREIBER, 1999). Ferramentas estatísticas são aplicadas a séries temporais para descrever componentes que incluem comportamento predominantemente estocástico (SCHREIBER, 1999; BOX et al., 2015). Essa abordagem consiste na decomposição da série entre componentes com alguma dependência temporal e outros com variações aleatórias ou resíduos. O pressuposto é que influências aleatórias externas atuando no processo gerador dos dados explicam a irregularidade enquanto a estrutura encontrada na sequência de observações pode ser descrita pela dinâmica da dependência temporal, a qual pode revelar tendências e padrões sazonais (SCHREIBER, 1999; SHUMWAY; STOFFER, 2017; BOX et al., 2015).

Diferentemente da análise linear, em que a estrutura regular da dinâmica intrínseca do processo é descrita por meio de correlações lineares e a irregularidade atribuída a algum fator aleatório externo, a abordagem não linear realiza uma avaliação menos reducionista da complexidade do fenômeno subjacente (KANTZ; SCHREIBER, 2004; SMALL, 2005; BRADLEY;

KANTZ, 2015). Usualmente, métodos advindos da *Teoria do Caos e Sistemas Dinâmicos* buscam quantificar e modelar, de modo mais natural, a descrição da dinâmica oculta por trás dos dados (BANDT; POMPE, 2002; OTT, 2002; KANTZ; SCHREIBER, 2004; SMALL, 2005; MARWAN et al., 2007; BRADLEY; KANTZ, 2015).

No contexto da dinâmica não linear, a ligação mais direta entre essa teoria e o mundo real é a NLTA (SCHREIBER, 1999; KANTZ; SCHREIBER, 2004). O fundamento dessa análise encontra-se no teorema de imersão de Takens (TAKENS, 1981), o qual demonstra, sob pressupostos necessários, que o espaço de estados de um sistema dinâmico pode ser reconstruído por meio de uma série temporal. A partir dessa reconstrução, diversas medidas referentes à dinâmica do processo gerador da série podem ser obtidas, permitindo a diferenciação de regimes dinâmicos distintos (SCHREIBER, 1999; MARWAN et al., 2007; DONNER et al., 2010; DONNER et al., 2011).

Apesar da vasta literatura sobre análise de séries temporais, poucas abordagens levam em consideração que séries temporais são um tipo especial de fluxo de dados (VALLIM; MELLO, 2014; COSTA; RIOS; MELLO, 2016; CAVALCANTE; MINKU; OLIVEIRA, 2016). Além disso, como apresentado na Seção 2.2.1, a maioria dos algoritmos aplicados na detecção de *mudança de conceito* consideram o uso de abordagens estatísticas relacionados à análise de séries temporais para identificar alterações na distribuição do fluxo de dados tendo como fundamento a hipótese de observações i.i.d. (WALD, 1973; BASSEVILLE; NIKIFOROV, 1993; ROSS et al., 2012; GAMA et al., 2014; SETHI; KANTARDZIC, 2017). Entretanto, séries temporais obtidas de fenômenos do mundo real tendem a apresentar alguma dependência temporal (SCHREIBER, 1999; KANTZ; SCHREIBER, 2004; MORETTI; TOLOI, 2006). Logo, ferramentas da NLTA (SCHREIBER, 1999; BANDT; POMPE, 2002; CAO et al., 2004; MARWAN et al., 2007) são alternativas para lidar com o problema da dependência temporal e a não-estacionaridade das observações.

3.2 DINÂMICA NÃO LINEAR E CAOS

A coleção de ideias e técnicas conhecidas como *análise de séries temporais não lineares* pode ser extremamente eficaz quando o objetivo é inferir propriedades de algum sistema dinâmico subjacente aos dados (BRADLEY; KANTZ, 2015). Nesse contexto, considera-se que os dados das séries são obtidos a partir da evolução de algum sistema dinâmico não-linear (SCHREIBER, 1999; KANTZ; SCHREIBER, 2004).

Esses sistemas podem ser definidos por meio das regras que governam a evolução de uma sequência de estados $\mathbf{v}_t \in \mathbb{R}^d$ (OTT, 2002; SAVI, 2006). Presume-se que cada estado \mathbf{v}_t

contenha informação suficiente da configuração atual do sistema, de modo que uma função determinística f possa definir o estado seguinte sem qualquer ambiguidade. Essas regras são definidas por meio de funções não-lineares em que duas categorias principais podem ser descritas a partir do tipo da evolução temporal: sistemas contínuos ou fluxos, e sistemas discretos ou mapas (SCHREIBER, 1999; OTT, 2002).

Fluxos são aqueles em que a evolução temporal decorre de um campo vetorial $\mathbf{v}(t)$ continuamente transformado por um sistema de equações diferenciais

$$\dot{\mathbf{v}}(t) = \mathbf{F}(\mathbf{v}(t)). \quad (3.1)$$

Diferentemente dos fluxos, em que a evolução temporal decorre de modo contínuo, nos mapas, a evolução decorre de forma discreta no tempo. Nesse caso, cada estado \mathbf{v}_t é definido a partir de um estado anterior \mathbf{v}_{t-1} , onde $t \in \mathbb{N}$, e

$$\mathbf{v}_{t+1} = \mathbf{f}(\mathbf{v}_t). \quad (3.2)$$

Uma das principais aplicações dos mapas é auxiliar a análise de sistemas contínuos (SAVI, 2006). Além do mais, a discretização possibilita avaliar a dinâmica de um sistema considerando uma perspectiva simbólica. Nessa abordagem, os estados podem ser mapeados em sequências de zeros e uns, de modo que o principal interesse é evitar erros de arredondamento inerentes a operações com números reais (SAVI, 2006).

3.2.1 Órbitas, atratores e caos

A partir de um estado inicial \mathbf{v}_0 , um sistema dinâmico f evolui para o próximo estado $\mathbf{v}_1 = f(\mathbf{v}_0)$ após a primeira iteração. Na próxima iteração, $f(\mathbf{v}_1)$ leva ao estado \mathbf{v}_2 , de modo que a evolução até o estado atual pode ser descrita como $f(f(\mathbf{v}_0)) = f^2(\mathbf{v}_0)$. A generalização em k iterações resulta no estado \mathbf{v}_k definido como $\mathbf{v}_k = f^k(\mathbf{v}_0)$, o qual será um elemento do conjunto de estados $\Gamma = \{\mathbf{v}_0, f^1(\mathbf{v}_0), f^2(\mathbf{v}_0), \dots, f^k(\mathbf{v}_0), f^{k+1}(\mathbf{v}_0), \dots\}$, denominado *espaço de fase*. Formalmente, a sequência de estados $\{\mathbf{v}_0, f^1(\mathbf{v}_0), f^2(\mathbf{v}_0), \dots, f^k(\mathbf{v}_0)\} \subset \Gamma$ descreve uma órbita, ou trajetória, a partir do estado inicial \mathbf{v}_0 até o estado \mathbf{v}_k (ALLIGOOD; SAUER; YORKE, 1997), cujo $f : \Gamma \rightarrow \Gamma$.

Diferentes trajetórias podem ser obtidas a partir de estados iniciais distintos $\tilde{\mathbf{v}}_0^i = \mathbf{v}_0 + \delta_i$, onde $\delta_i \in \mathbb{R}^d$ representa uma perturbação sutil do estado inicial \mathbf{v}_0 referente a cada trajetória

alternativa $\tilde{\Gamma}^i$. Essas trajetórias fornecem informações relevantes sobre a evolução da dinâmica do sistema, onde todos os estados possíveis são representados em um espaço denominado *espaço de fase* (ALLIGOOD; SAUER; YORKE, 1997). Nesse espaço, a evolução dos estados iniciais pode revelar padrões como *órbitas periódicas*, *pontos-fixos*, e *órbitas caóticas* (ALLIGOOD; SAUER; YORKE, 1997; SAVI, 2006).

Uma *órbita periódica* é definida quando a repetição do estado \mathbf{v}_p apresenta um período k , ou k -ciclo, de modo que $\mathbf{v}_p = \mathbf{f}^k(\mathbf{v}_p)$ após k iterações. Um *ponto-fixo* é um caso particular em que $k = 1$. Nesse caso, o sistema permanece no estado \mathbf{v}_p após cada iteração, em que $\mathbf{v}_p = \mathbf{f}(\mathbf{v}_p)$. Esse comportamento define um *órbita de período 1* (um), ou 1-ciclo, devido o mapa \mathbf{f} retornar ao estado \mathbf{v}_p numa única iteração. Um *ponto-fixo* pode ser caracterizado como estável ou instável caso os estados vizinhos desse ponto sejam atraídos ou repelidos, respectivamente, durante a evolução de suas *órbitas* (ALLIGOOD; SAUER; YORKE, 1997; DEVANEY, 1992).

A partir de um *ponto-fixo* \mathbf{v}_p e um raio $\varepsilon \in \mathbb{R}$, uma região de vizinhança pode ser definida como

$$N(\mathbf{v}_p, \varepsilon) = \{\mathbf{v}_j : \|\mathbf{v}_j - \mathbf{v}_p\| \leq \varepsilon\}, \quad (3.3)$$

onde $\|\cdot\|$ é a distância Euclidiana. Caso todos os pontos \mathbf{v}_j convirjam para \mathbf{v}_p , de modo que $\lim_{k \rightarrow \infty} \mathbf{v}_p = \mathbf{f}^k(\mathbf{v}_j)$, o estado \mathbf{v}_p definirá um *atrator* e a região de vizinhança $N(\mathbf{v}_p, \varepsilon)$ será uma *bacia de atração* (ALLIGOOD; SAUER; YORKE, 1997). Essencialmente, um *atrator* pode ser um único *ponto-fixo*, um conjunto de pontos regularmente visitados, ou uma *órbita caótica* (ALLIGOOD; SAUER; YORKE, 1997). De qualquer modo, o *atrator* captura todas as *órbitas* iniciadas na sua *bacia de atração*, de modo que o estado do sistema tende para essa região do *espaço de fase* na medida em que as *órbitas* evoluem (SAVI, 2006).

Uma *órbita caótica* é aperiódica e apresenta sensibilidade a variações mínimas das condições iniciais (ALLIGOOD; SAUER; YORKE, 1997). Nesse caso, a divergência entre os estados $\tilde{\mathbf{v}}_k^A = \mathbf{f}^k(\tilde{\mathbf{v}}_0^A)$ e $\tilde{\mathbf{v}}_k^B = \mathbf{f}^k(\tilde{\mathbf{v}}_0^B)$ obtidos, respectivamente, durante a evolução de duas *órbitas caóticas* $\tilde{\Gamma}^A$ e $\tilde{\Gamma}^B$, deve aumentar exponencialmente com o incremento de k iterações. As condições iniciais de vizinhança entre $\tilde{\Gamma}^A$ e $\tilde{\Gamma}^B$ são definidas a partir dos seus respectivos estados iniciais $\tilde{\mathbf{v}}_0^A$ e $\tilde{\mathbf{v}}_0^B$, considerando uma diferença infinitesimal $\Delta\delta = \|\delta_A - \delta_B\|$. Esse comportamento está associado a atratores com uma estrutura complexa, denominados *atratores estranhos* ou *atratores caóticos*, os quais estão relacionados ao conceito de *caos* (LI; YORKE, 1975; DEVANEY, 1992; ALLIGOOD; SAUER; YORKE, 1997).

O *caos* tem sido definido como o comportamento estocástico de sistemas determinísticos

(SAVI, 2006). Desse modo, muitos sistemas dinâmicos são passíveis de apresentar comportamento caótico, desde que algumas condições sejam satisfeitas (SAVI, 2006). Por exemplo, sistemas autônomos, os quais não dependem explicitamente do tempo, definidos por equações diferenciais ordinárias no $\mathbb{R}^{d \geq 3}$, podem apresentar comportamento caótico. Da mesma maneira, mapas inversíveis, os quais possuem seu inverso f^{-1} definido, de modo a possibilitar retroceder numa trajetória, podem apresentar *caos* caso estejam contidos no $\mathbb{R}^{d \geq 2}$. Entretanto, mapas não-inversíveis podem apresentar comportamento caótico mesmo no $\mathbb{R}^{d=1}$ (DEVANEY, 1992; ALLIGOOD; SAUER; YORKE, 1997; OTT, 2002; SAVI, 2006).

Além da sensibilidade às condições iniciais, um outro requisito importante relacionado ao *caos* é a *não-linearidade* (ALLIGOOD; SAUER; YORKE, 1997; SCHREIBER, 1999). Assim, caso o sistema seja governado por regras lineares, essa condição será suficiente para descartar um comportamento caótico (ALLIGOOD; SAUER; YORKE, 1997). Para detalhar esses conceitos, os sistemas dinâmicos caóticos apresentados a seguir ilustram a relação entre a *teoria da dinâmica não-linear* e a análise de séries temporais.

3.2.1.1 Mapa de Bernoulli

No estudo de sistemas dinâmicos caóticos, mapas unidimensionais são os modelos mais simples para descrever as propriedades subjacentes a tais sistemas (OTT, 2002). Dentre esses modelos, o Mapa de Bernoulli, definido na Equação 3.4, é não-inversível e seu comportamento pode ser facilmente descrito ao considerar a *dinâmica simbólica*, uma ferramenta importante para entender o comportamento caótico de sistemas dinâmicos (DEVANEY, 1992; LIND; MARCUS, 2021). Por definição, o operador *mod* não exige números inteiros como entrada: $a \bmod b = c \iff \exists k \in \mathbb{Z} \mid a - bk = c$, com $c < b$. Logo $a \bmod 1$ implica que, se $a \in \mathbb{R}$, resta na definição apenas a parte fracionária de a . Assim, na Equação 3.4, a operação $2x_t \bmod 1$ trunca a parte inteira a cada iteração, retornando apenas a parte numérica depois do ponto decimal.

$$x_{t+1} = B(x_t) = \begin{cases} 2x_t & \text{if } 0 \leq x_t \leq 0,5, \\ 2x_t - 1 & \text{if } 0,5 < x_t \leq 1, \end{cases} \equiv 2x_t \bmod 1, \quad 0 \leq x_0 \leq 1. \quad (3.4)$$

Por meio de iterações sucessivas a partir de um estado inicial $x_0 \in \mathbb{R}$, pode-se gerar uma sequência de valores, uma órbita do mapa, em que $x_1 = B(x_0)$, $x_2 = B(x_1) = B^2(x_0)$, $x_3 =$

$B(x_2) = B^3(x_0), \dots$, onde $x_{t+m} = B^m(x_t)$. Uma maneira mais adequada de descrever este processo é representar a condição inicial x_0 como uma sequência binária infinita

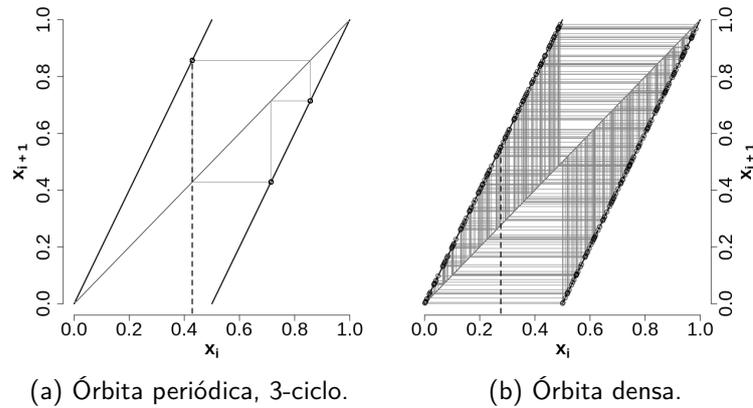
$$x_0 = 0, b_1 b_2 b_3 \dots \equiv \sum_{j=1}^{\infty} 2^{-j} b_j = \frac{b_1}{2^1} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \dots, b_j \in \{0, 1\}. \quad (3.5)$$

Desse modo, cada iteração produz uma nova sequência de dígitos binários, onde o primeiro dígito é descartado e os restantes são deslocados para a esquerda. Devido a esse procedimento, o Mapa de Bernoulli também é conhecido como *deslocamento de Bernoulli*, no qual os estados são definidos a cada deslocamento. Assim, os estados da órbita do sistema, $x_1 = 0, b_2 b_3 b_4 \dots$, $x_2 = 0, b_3 b_4 b_5 \dots$, $x_3 = 0, b_4 b_5 b_6 \dots$, etc., mapeados em sequências infinitas de 0's e 1's, definem a sua *dinâmica simbólica* (DEVANEY, 1992).

O mapeamento dos estados em expansões binárias permite avaliar o comportamento das orbitas por meio de padrões simbólicos (DEVANEY, 1992; SAVI, 2006). Por exemplo, no Mapa de Bernoulli, os *pontos-fixos* são definidos por $\{0.0000 \dots\} \equiv \{0.\bar{0}\}$ ou $\{0.1111 \dots\} \equiv \{0.\bar{1}\}$. Da mesma maneira, *orbitas periódicas* com k -ciclos podem ser descritas por blocos de 0's e 1's repetidos indefinitivamente. Nesse caso, a expansão binária $x_0^A = \{0.011011011 \dots\} \equiv \{0.\overline{011}\} \equiv \{0.\overline{110}\} \equiv \{0.\overline{101}\}$ representa uma órbita de período 3, ou 3-ciclos, e a expansão binária $x_0^B = \{0.101010 \dots\} \equiv \{0.\overline{10}\} \equiv \{0.\overline{01}\}$ descreve uma órbita de período 2, ou 2-ciclos. Além desses comportamentos, uma órbita aperiódica e densa, relacionado ao comportamento caótico, pode ser definida pela expansão binária $x_0^C = \{0.0100011011000001010011100101110111 \dots\}$ (DEVANEY, 1992; SAVI, 2006). Nesse caso, a sequência binária é definida por meio da união de blocos com um bit $\{0, 1\}$, dois bits $\{00, 01, 10, 11\}$, três bits $\{000, 001, 010, 011, 100, 101, 110, 111\}$, etc, em que os elementos estão ordenados de modo ascendente.

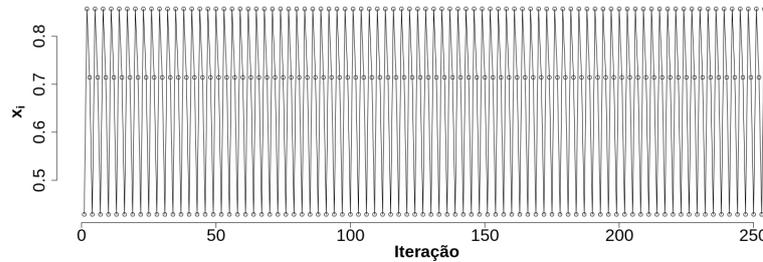
Considerando os estados iniciais x_0^A e x_0^C , os *diagramas teia de aranha* (do Inglês *cobweb plot*) das Figura 2 (a) e (b), ilustram os respectivos comportamentos de suas órbitas durante 256 iterações (ou deslocamentos), onde esses estados iniciais estão destacados por uma linha tracejada. Nessas figuras, $x_0^A \cong 0,4286$ e $x_0^C \cong 0,2764$, no domínio \mathbb{R} . Essa aproximação é devido à restrição da representação dessas expansões binárias em um computador digital. Como exemplo, $x_0^C \cong 0,2764 \equiv 0.0100011011000010001001101\dots$, o qual, de acordo com a Equação 3.5, é expresso como $x_0^C \equiv \frac{0}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{0}{2^4} + \frac{0}{2^5} + \frac{1}{2^6} + \frac{1}{2^7} + \frac{0}{2^8} + \frac{1}{2^9} + \frac{1}{2^{10}} \dots$. Nos diagramas das Figura 2 (a) e (b), uma bissetriz definida pelo gráfico da função identidade $f(x) = x$, determina a localização dos *pontos-fixos* sempre que $f(x) = B(x)$ (ALLIGOOD; SAUER; YORKE, 1997). Nesse caso, esses pontos são obtidos quando $x = 0$ ou $x = 1$.

Figura 2 – Diagramas teia de aranha (do Inglês *cobweb plot*) produzidos a partir dos estados iniciais $x_0^A = 0,4286$ e $x_0^C = 0,2764$, identificados pela linha tracejada.



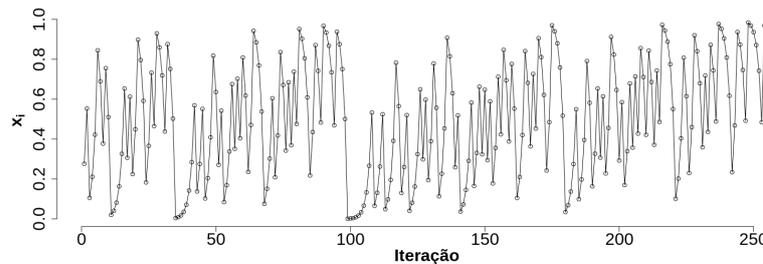
Fonte: Elaborada pelo autor (2021).

Figura 3 – Série temporal obtida a partir da evolução da órbita periódica referente ao estado inicial x_0^A .



Fonte: Elaborada pelo autor (2021).

Figura 4 – Série temporal obtida a partir da evolução da órbita aperiódica referente ao estado inicial x_0^C .



Fonte: Elaborada pelo autor (2021).

Numa outra perspectiva, a evolução dessas órbitas são descritas por meio das séries temporais ilustradas nas Figuras 3 e 4, respectivamente. Na Figura 3, a série apresenta um comportamento regular, característico de um sistema determinístico. Entretanto, na Figura 4, a série assemelha-se a um processo estocástico.

3.2.1.2 Mapa Logístico

O Mapa Logístico (MAY, 1976), definido na Equação 3.6, foi proposto como um modelo de crescimento populacional, no qual a sua dinâmica varia de acordo com a taxa de reprodução $r \in \mathbb{R}$.

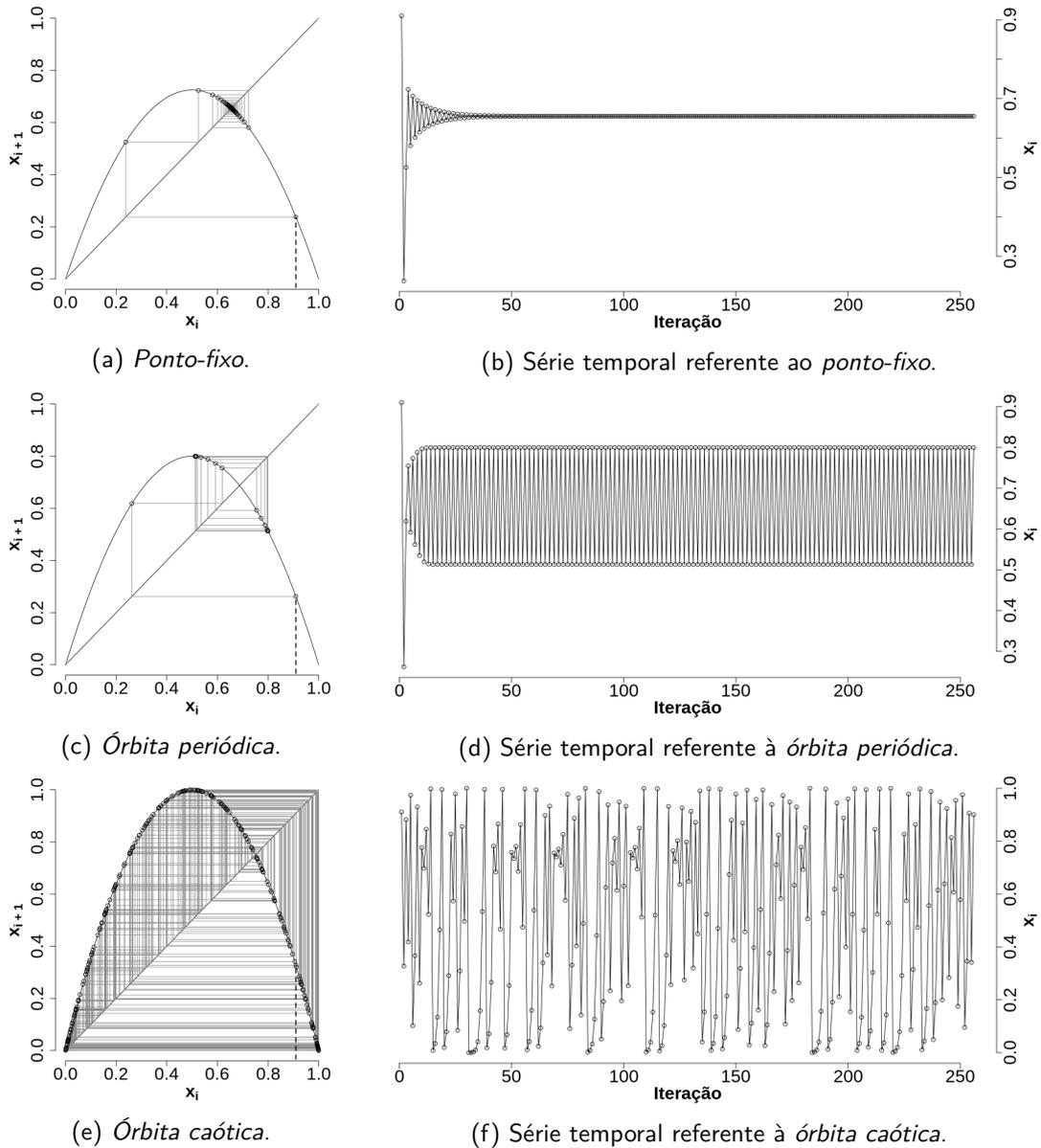
$$x_{t+1} = L(x_t) = r \times x_t(1 - x_t). \quad (3.6)$$

Na verdade, trata-se de um sistema dinâmico unidimensional, em que o parâmetro r é responsável por diferentes regimes (DEVANEY, 1992). Dependendo do valor desse parâmetro, o mapa pode evoluir para um *ponto-fixo*, uma *órbita periódica*, ou uma *órbita caótica*, a partir de um mesmo estado inicial x_0 . Como exemplo, a Figura 5 ilustra essas três dinâmicas considerando a evolução das órbitas em *diagramas teia de aranha*, bem como, as respectivas séries temporais, ambas com 256 iterações.

A Figura 5a ilustra uma trajetória definida com $r = 2,9$, e o estado inicial $x_0 = 0,91$, destacado por uma linha tracejada. Nesse *diagrama teia de aranha*, a trajetória converge para o *ponto-fixo* $x_k = 0,655$, a qual também pode ser observada por meio da série temporal apresentada na Figura 5b. De modo semelhante, as Figuras 5c e 5d apresentam, respectivamente, o *diagrama teia de aranha* e a séries temporal correspondente, definidos com $r = 3,2$, a partir do mesmo estado inicial, também destacado por uma linha tracejada. Nesse caso, as iterações evoluem para uma *órbita periódica*, a qual oscila entre os estados $x_k = 0,799$ e $x_{k+1} = 0,513$. Nessas configurações, as trajetórias apresentam iterações transitórias antes de convergir para valores assintóticos como um *ponto-fixo* ou uma *órbita periódica*. Finalmente, na Figura 5e, ao considerar $r = 4,0$, e o mesmo estado inicial $x_0 = 0,91$, a evolução da *órbita caótica* tende a abranger significativamente os estados do intervalo $[0, 1]$. Além disso, a série temporal ilustrada na Figura 5f parece apresentar um comportamento estocástico.

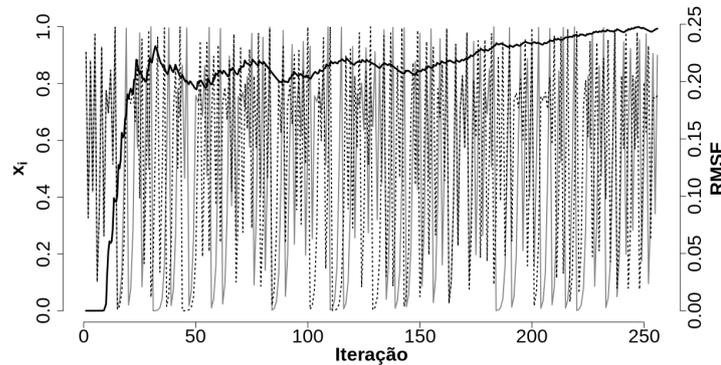
Com a configuração $L(x_t) = 4,0 \times x_t(1 - x_t)$, ou seja, $r = 4,0$, o Mapa Logístico apresenta sensibilidade às condições iniciais. Como exemplo, duas trajetórias bastante distintas podem ser obtidas a partir dos estados iniciais $\tilde{x}_0^A = x_0 + \delta_A$ e $\tilde{x}_0^B = x_0 + \delta_B$, onde $\delta_A = 0$ não altera a órbita $\tilde{\Gamma}^A$ em relação ao estado inicial $x_0 = 0,91$, e $\delta_B = 10^{-6}$ representa uma perturbação sutil desse estado inicial referente à órbita $\tilde{\Gamma}^B$, de modo que $\Delta\delta = |\delta_A - \delta_B| = 10^{-6}$. Nesse caso, a diferença entre a evolução das duas órbitas caóticas $\tilde{\Gamma}^A$ e $\tilde{\Gamma}^B$ deve aumentar significativamente ao longo das iterações. As Figuras 6 e 7 ilustram,

Figura 5 – *Diagramas teia de aranha* da evolução do Mapa Logístico e suas respectivas séries temporais, considerando um *ponto-fixo* ($r = 2,9$), uma *órbita periódica* ($r = 3,2$), e uma *órbita caótica* ($r = 4,0$).



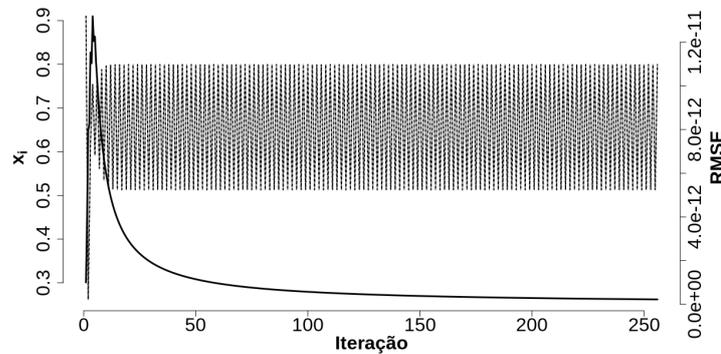
Fonte: Elaborada pelo autor (2021).

Figura 6 – RMSE obtido das séries temporais referentes à evolução dos estados iniciais \tilde{x}_0^A e \tilde{x}_0^B , considerando $r = 4, 0$.



Fonte: Elaborada pelo autor (2021).

Figura 7 – RMSE obtido das séries temporais referentes à evolução dos estados iniciais \tilde{x}_0^A e \tilde{x}_0^B , considerando $r = 3, 2$.



Fonte: Elaborada pelo autor (2021).

respectivamente, o efeito da sensibilidade às condições iniciais considerando a configuração caótica do mapa com $r = 4, 0$, bem como a configuração periódica $L(x_t) = 3, 2 \times x_t(1 - x_t)$, ou seja, $r = 3, 2$.

Nessas figuras, as séries temporais referentes às órbitas $\tilde{\Gamma}^A$ e $\tilde{\Gamma}^B$ foram plotadas, respectivamente, com a cor cinza e pontilhada com a cor preta. A linha contínua, destacada ao longo das iterações representa a Root Mean Square Error (RMSE), definido na Equação 3.7, a qual foi utilizada para computar a diferença entre as séries temporais das órbitas $\tilde{\Gamma}^A$ e $\tilde{\Gamma}^B$ a cada iteração.

$$RMSE = \sqrt{\frac{\sum_{i=0}^{i=k} (L^i(\tilde{x}_0^A) - L^i(\tilde{x}_0^B))^2}{k}}. \quad (3.7)$$

Na Figura 6, o RMSE aumenta significativamente ao longo da evolução das *órbitas caóticas*. Por outro lado, na Figura 7, a qual ilustra a evolução das *órbitas periódicas*, o RMSE apresenta um decaimento exponencial ao longo das iterações. Nessa configuração, as séries temporais

das órbitas $\tilde{\Gamma}^A$ e $\tilde{\Gamma}^B$ ficam sobrepostas após as iterações transitórias. Apesar da diferença insignificante dos estados iniciais ($\Delta\delta = 10^{-6}$), na evolução das *órbitas caóticas* $\tilde{\Gamma}^A$ e $\tilde{\Gamma}^B$ essa diferença entre as trajetórias foi bastante significativa.

3.2.1.3 Sistema de Lorenz

Um exemplo de sistema dinâmico contínuo não-linear é o Sistema de Lorenz (LORENZ, 1963), definido pelo sistema de equações

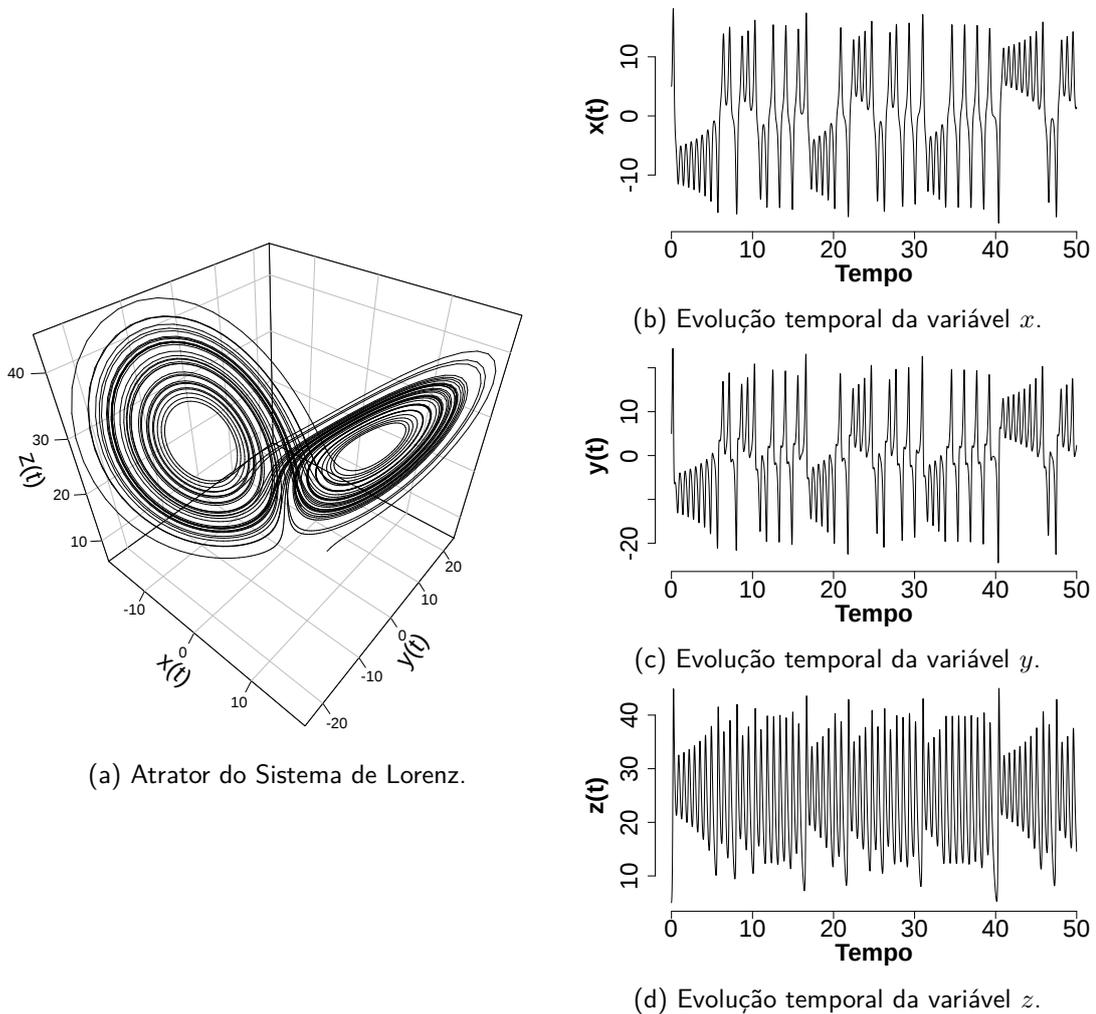
$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z.\end{aligned}\tag{3.8}$$

Esse sistema é um modelo simplificado para descrever os movimentos de convecção atmosférica. O parâmetro σ , proporcional ao número de Prandtl, e o parâmetro ρ , proporcional ao número de Rayleigh, estão relacionados, respectivamente, à difusividade térmica e à forma de transferência de calor, a qual ocorre principalmente por condução ou convecção. Por outro lado, o parâmetro β depende das propriedades geométricas da região atmosférica em consideração (SPARROW, 1982).

Uma configuração usual para esses parâmetros é $\sigma = 10$, $\rho = 28$ e $\beta = 8/3$, os quais devem ser números reais e positivos. A partir da definição desses parâmetros, as variáveis x , y e z definem o estado atual, enquanto \dot{x} , \dot{y} e \dot{z} correspondem ao deslocamento dos estados ao longo do tempo. Fisicamente, a variável x mede a taxa de escoamento convectivo, enquanto as variáveis y e z descrevem, respectivamente, variações de temperatura nas direções horizontal e vertical (SPARROW, 1982).

A Figura 8 ilustra a trajetória da evolução temporal dos estados em um *espaço de fase* no \mathbb{R}^3 , bem como as suas componentes x , y e z . Foram realizadas 10^5 iterações com um intervalo de tempo definido como 10^{-2} , considerando a parametrização $\sigma = 10$, $\rho = 28$ e $\beta = 8/3$. O estado inicial foi definido como $\mathbf{v}_0 = (x_0, y_0, z_0) = (5, 5, 5)$. A Figura 8a descreve a órbita do sistema, ou seja, o conjunto de estados percorrido de modo contínuo ao longo do tempo. Nesse caso, trata-se do atrator do Sistema de Lorenz. Nas Figuras 8b, 8c e 8d, as séries temporais $x(t)$, $y(t)$, e $z(t)$ descrevem, respectivamente, os deslocamentos \dot{x} , \dot{y} e \dot{z} ao longo do tempo.

Figura 8 – Trajetória da evolução temporal dos estados em um *espaço de fase* no \mathbb{R}^3 com as séries temporais das variáveis x , y e z do Sistema de Lorenz.



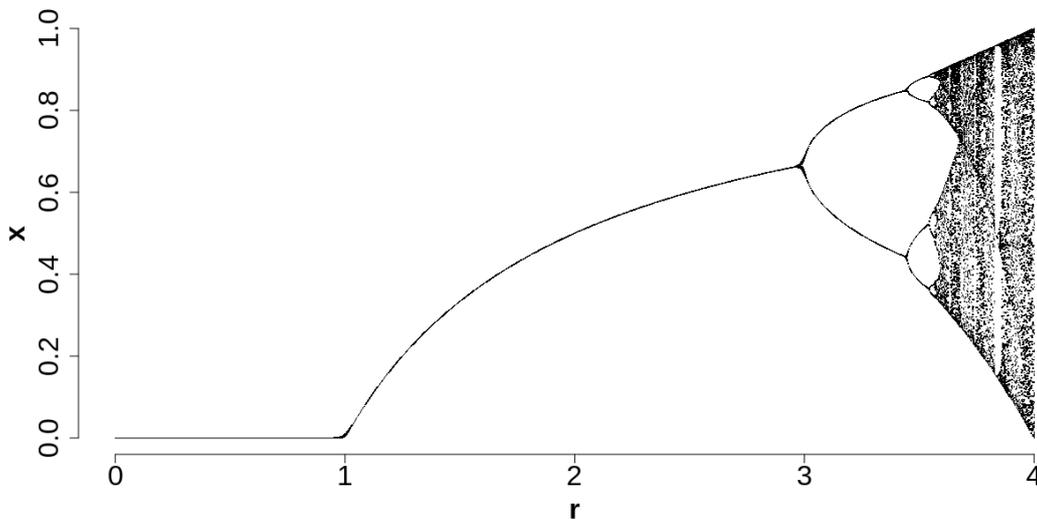
Fonte: Ilustração adaptada de COSTA (2017).

3.2.2 Bifurcação e modelos transientes

Nos modelos apresentados na seção anterior, diferentes comportamentos assintóticos foram classificados como *ponto-fixa*, *órbita periódica*, e *órbita caótica*. De modo a encontrar mecanismos pelos quais o sistema alterna de um comportamento simples para um altamente complexo, busca-se identificar como os valores assintóticos da variável de estado podem mudar conforme o sistema é configurado (ALLIGOOD; SAUER; YORKE, 1997).

Tomando como exemplo o Mapa Logístico, definido na Equação 3.6, a Figura 9 ilustra o seu comportamento num *diagrama de bifurcação*. Nesse diagrama, os valores assintóticos da variável de estado $x \in [0; 1]$ são plotados ao longo do eixo das ordenadas versus os valores do parâmetro de controle $r \in [0; 4]$, descrito no eixo das abscissas (ALLIGOOD; SAUER; YORKE,

Figura 9 – Diagrama de bifurcação do Mapa Logístico



Fonte: Elaborada pelo autor (2021).

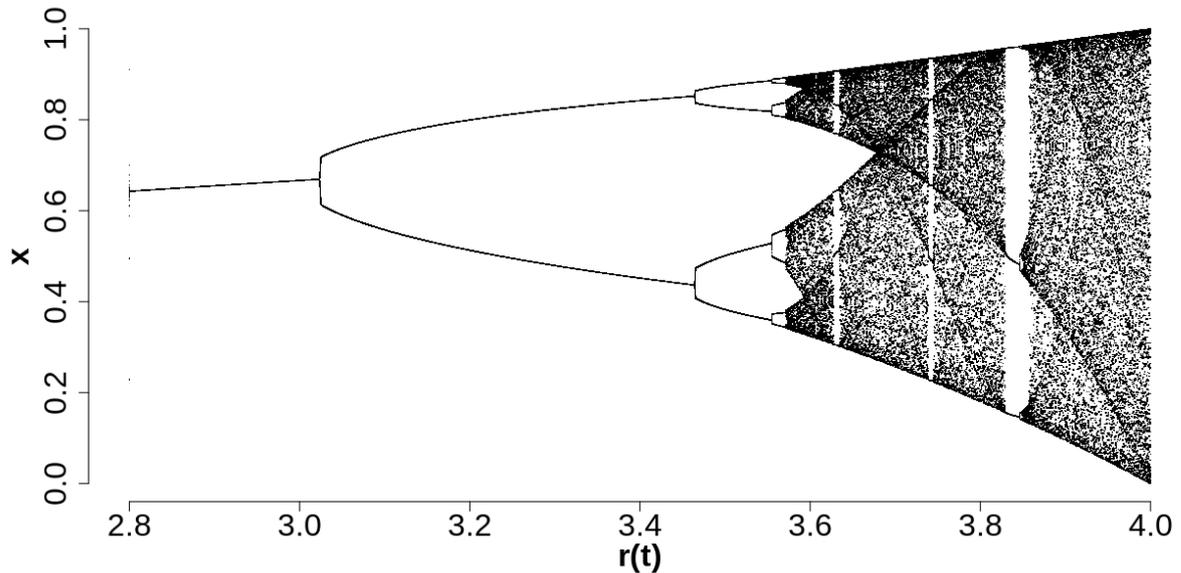
1997). Normalmente, pode ocorrer mais de um ponto ao longo do eixo das ordenadas para cada valor desse parâmetro. Conforme ilustrado na Figura 9, diferentes comportamentos podem ser identificados por meio do *diagrama de bifurcação*. Nesse gráfico, o caos aparece após uma sequência de ramificações em forma de bifurcações ao longo dos incrementos do parâmetro r . Essa sequência de bifurcações inicia a partir de $r = 3,0$, como uma órbita de período 2, ou 2-ciclo, condição já ilustrada na Figura 5e para $r = 3,2$. Com o aumento do parâmetro r , o sistema evolui para 4-ciclo ($r = 3,45$), 8-ciclo ($r = 3,54$), 16-ciclo ($r = 3,56$), etc., sempre dobrando o período, até alcançar $r = 4,0$, onde finalmente termina como um sistema caótico.

Semelhante ao *diagrama de bifurcação*, o Mapa Logístico Transiente (TRULLA et al., 1996) e o Sistema de Lorenz Transiente (CAO et al., 2004) são modelos usados como geradores de séries temporais com características transientes. A Figura 10 apresenta uma série temporal consistindo de 120.000 pontos, gerada pelo Mapa Logístico Transiente. Para essa série, o estado inicial foi definido como $x_0 = 0,65$ e o parâmetro $r(t)$ incrementado de 2,8 até 4,0, com uma variação $\Delta t = 10^{-5}$ a cada iteração.

Apesar da semelhança com o *diagrama de bifurcação*, a série temporal resultante apresenta apenas um único ponto ao longo do eixo das ordenadas para cada valor do parâmetro $r(t)$. Nesse diagrama, regimes dinâmicos distintos são observados com 1-ciclo, 2-ciclo, 4-ciclo, 8-ciclo, caótico, 6-ciclo, caótico novamente, 5-ciclo, caótico novamente, 3-ciclo, e por fim, caótico. Além disso, devido ao comportamento transiente, as posições dos valores de $r(t)$ onde as bifurcações ocorrem são ligeiramente diferentes dos pontos de bifurcação usuais do

diagrama de bifurcação (TRULLA et al., 1996; CAO et al., 2004).

Figura 10 – Série temporal gerada pelo Mapa Logístico Transiente considerando iterações do parâmetro $r(t)$ entre os valores 2,8 e 4,0, com incrementos de 10^{-5} .

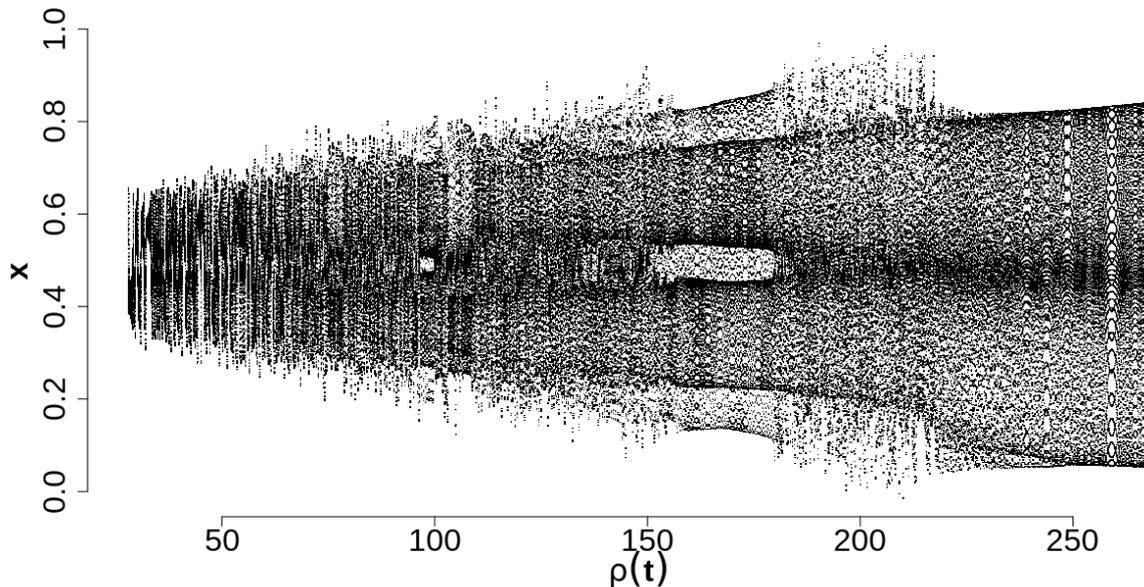


Fonte: Elaborada pelo autor (2021).

Seguindo a mesma ideia, a Figura 11 apresenta o comportamento de uma série temporal, também com 120.000 pontos, gerada pelo Sistema de Lorenz Transiente a partir do estado inicial $\mathbf{v}_0 = (5, 5, 5)$. Para resolver o sistema, o método de Runge-Kutta de Quarta Ordem foi utilizado considerando o passo de integração $\Delta t = 0,01$. O parâmetro $\rho(t)$ varia de 28,0 a 268,0 com incrementos de 0,002, de acordo com Cao et al. (2004). A série temporal resultante é mais complexa, e teoricamente apresenta janelas periódicas quando $99,524 < \rho(t) < 100,795$, $145 < \rho(t) < 166$ e $\rho(t) > 214,4$. Entretanto, devido ao comportamento transiente, a característica periódica da primeira janela ($99,524 < \rho(t) < 100,795$) não está mais presente (CAO et al., 2004).

É importante destacar que esses dois modelos transientes têm sido aplicados ao contexto da detecção de mudanças dinâmicas em séries temporais (TRULLA et al., 1996; CAO et al., 2004), bem como na detecção de *mudança de conceito* em cenários de *fluxo de dados* (COSTA et al., 2017; COSTA, 2017).

Figura 11 – Série temporal gerada pelo Sistema de Lorenz Transiente considerando iterações discretas do parâmetro $\rho(t)$ entre os valores 28,0 e 268,0, com incrementos de 0,002.



Fonte: Elaborada pelo autor (2021).

3.3 RECONSTRUÇÃO DO ESPAÇO DE FASE

Empiricamente, para séries temporais extraídas de problemas reais, não existe uma maneira de identificar, *a priori*, a forma das equações de evolução e nem as variáveis de estado. Normalmente, observa-se uma única variável por meio de uma série temporal, na qual são aplicadas técnicas para identificar a dimensão e a reconstrução da trajetória dos estados em um *espaço de fase* topologicamente equivalente ao original (KANTZ; SCHREIBER, 2004; SMALL, 2005). Essa coleção de ideias e técnicas, denominada *análise de séries temporais não lineares*, pode ser extremamente eficaz quando o modelo de dados apresenta uma dinâmica determinística em algum *espaço de estados* (BRADLEY; KANTZ, 2015). Embora a hipótese do determinismo não seja apropriada para caracterizar aspectos estocásticos, essa abordagem permite inferir, a partir dos dados de uma série temporal, propriedades de algum sistema dinâmico oculto (KANTZ; SCHREIBER, 2004; BRADLEY; KANTZ, 2015).

A reconstrução do *espaço de fase* é o núcleo da *análise de séries temporais não lineares* (SCHREIBER, 1999; BRADLEY; KANTZ, 2015). Basicamente, os principais métodos utilizados nessa reconstrução são: o método das coordenadas derivativas (PACKARD et al., 1980), o método baseado em Singular Value Decomposition (traduzido para o português, Decomposição em Valores Singulares) (SVD) (BROOMHEAD; KING, 1986), e o método das coordenadas de-

fasadas (PACKARD et al., 1980; TAKENS, 1981). Devido a vasta aplicação do método das coordenadas defasadas (ALLIGOOD; SAUER; YORKE, 1997; KANTZ; SCHREIBER, 2004), apenas abordagens relacionadas a esse método foram utilizadas nesta tese.

O método das coordenadas defasadas está fundamentado no *teorema de imersão de Takens*, proposto por Packard et al. (1980) e formalizado por Takens (1981), o qual permite resolver um problema de complexidade considerável, ou seja, reconstruir a dinâmica de um sistema não linear a partir de uma única série temporal (BRADLEY; KANTZ, 2015). A reconstrução não resulta no mesmo atrator original, mas ainda mantém a equivalência topológica, sendo extremamente útil para inferir conclusões acerca da verdadeira dinâmica do sistema (SCHREIBER, 1999; BRADLEY; KANTZ, 2015; COSTA; RIOS; MELLO, 2016).

De acordo com o teorema, a dinâmica de um sistema pode ser desdobrada no domínio de estados $\Gamma = \{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ a partir de uma sequência de observações no domínio temporal $\mathbf{s} = \{x_0, x_1, x_2, \dots, x_{N-1}\}$. A abordagem padrão para a reconstrução do *espaço de fase* fundamenta-se na estimação de dois parâmetros livres: o *retardo temporal* τ e a *dimensão de imersão* m . Assumindo que os parâmetros foram estimados corretamente, a detecção de mudanças dinâmicas pode ser realizada usando vários métodos (TRULLA et al., 1996; BANDT; POMPE, 2002; CAO et al., 2004; MARWAN et al., 2007; DONNER et al., 2010; DONNER et al., 2011). Além disso, a partir de uma reconstrução bem sucedida, o *espaço de fase* pode ser usado como uma estrutura de análise importante para entender as relações temporais entre as observações, permitindo identificar *mudanças de conceito* em termos da dinâmica do processo subjacente ao *fluxo de dados* (COSTA; RIOS; MELLO, 2016; COSTA et al., 2017).

3.3.1 Retardo Temporal

A qualidade da reconstrução do *espaço de fase* Γ está implicitamente relacionada com a escolha do *retardo temporal* τ , definido como retardos temporais adicionados aos elementos da série temporal \mathbf{s} para construção de um conjunto de estados

$$\mathbf{v}_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, \dots, x_{t+(m-1)\tau}\} \in \mathbb{R}^m, \mathbf{v}_t \in \Gamma. \quad (3.9)$$

Em teoria, qualquer valor diferente de zero pode ser utilizado para esse parâmetro (SMALL, 2005; BRADLEY; KANTZ, 2015). Entretanto, para dados reais, limitados e sujeitos a ruídos, deve-se levar em consideração um valor apropriado. Tipicamente, um valor pequeno para o parâmetro τ define uma forte correlação entre as componentes dos estados \mathbf{v}_t , comprimindo

a reconstrução numa linha diagonal. Por outro lado, um valor muito grande provocará distorções na trajetória dos estados devido à independência significativa entre as suas componentes (KANTZ; SCHREIBER, 2004).

Diversos métodos foram propostos para estimar o *retardo temporal* τ , visando representar o sistema original de forma topologicamente apropriada (FRASER; SWINNEY, 1986; ALBANO; PASSAMANTE; FARRELL, 1991; ABARBANEL et al., 1993; KEMBER; FOWLER, 1993). Dentre essas propostas, os métodos mais usados são a Autocorrelation Function (traduzido para o português, Função de Autocorrelação) (FAC) (ABARBANEL et al., 1993) e a Average Mutual Information Function (traduzido para o português, Função de Informação Mútua Média) (AMIF) (FRASER; SWINNEY, 1986).

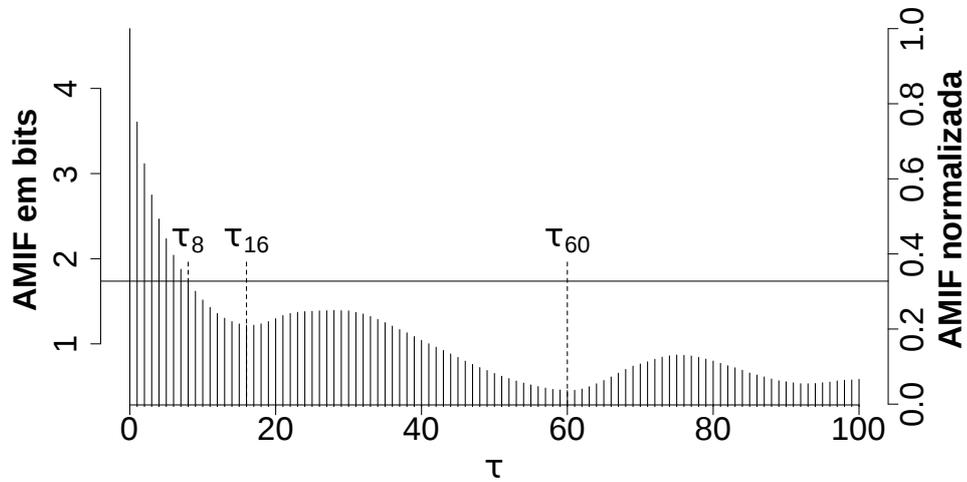
A FAC define uma medida da dependência linear entre séries temporais distintas, ou em um mesma série defasada, podendo ser utilizada como uma primeira estimativa. Em geral, a estimativa é dada pelo primeiro valor abaixo de zero da FAC, ou o primeiro valor abaixo do limiar $1/e$, onde e é o número de Euler (KANTZ; SCHREIBER, 2004; SMALL, 2005). Para alguns sistemas simples, a aplicação da FAC resulta em uma pobre reconstrução, mas suficiente na maioria dos casos (SMALL, 2005; SAVI, 2006). Por outro lado, a AMIF, definida na Equação 3.10, é mais efetiva, principalmente ao lidar com dados de natureza não linear. Na verdade, trata-se de uma função de correlação mais geral, podendo ser aplicada à séries temporais com estrutura linear ou não linear (FRASER; SWINNEY, 1986; KANTZ; SCHREIBER, 2004). Devido à maior aceitação da AMIF para a estimação do *retardo temporal* τ , somente esse método foi utilizado ao longo desta tese.

$$I(\tau) = \sum P(x_t, x_{t+\tau}) \log_2 \frac{P(x_t, x_{t+\tau})}{P(x_t)P(x_{t+\tau})}. \quad (3.10)$$

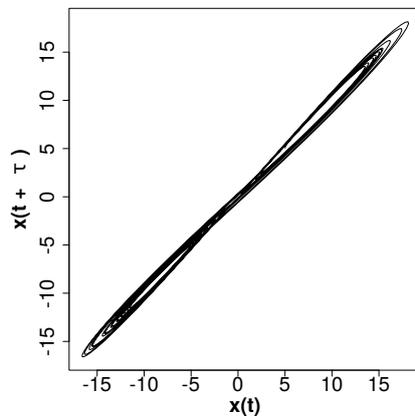
A estimação do parâmetro τ é sugerida pelo primeiro mínimo local da AMIF, indicando uma boa aproximação do valor ideal (FRASER; SWINNEY, 1986; KANTZ; SCHREIBER, 2004). Igualmente, pode-se considerar o primeiro decaimento abaixo do limiar $1/e$, após a normalização da AMIF (SEMMLOW; GRIFFEL, 2014). A Figura 12 exhibe o gráfico da AMIF considerando a escala em número de bits, bem como na forma normalizada. Nesse gráfico, a linha horizontal contínua identifica o limiar $1/e$, no caso normalizado. A Figura também ilustra a reconstrução de *espaços de fases* com diferentes defasagens a partir da série temporal da variável x do Sistema de Lorenz (Figura 8b).

A Figura 12a exhibe a AMIF aplicada à série temporal da variável x do Sistema de Lorenz

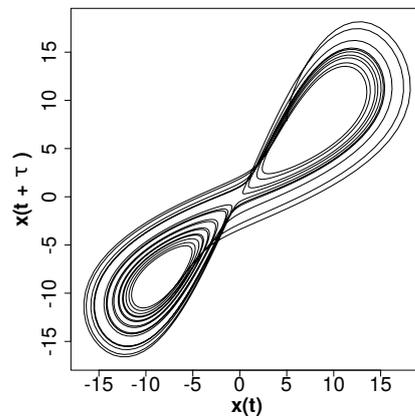
Figura 12 – Gráfico da AMIF para a estimação do parâmetro τ referente à variável x do Sistema de Lorenz, e as respectivas reconstruções do atrator, considerando as defasagens $\tau = 1, \tau = 8, \tau = 16$ e $\tau = 60$.



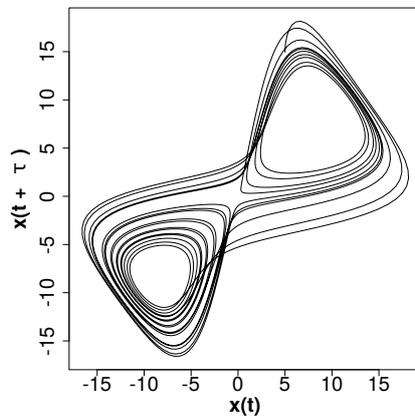
(a) Gráfico da AMIF para a estimação do parâmetro τ com 100 defasagens da série temporal referente à variável x do Sistema de Lorenz. As linhas verticais pontilhadas destacam o primeiro *retardo temporal* $\tau = 8$ abaixo do limiar $1/e$ (linha horizontal contínua), e dois mínimos locais, respectivamente $\tau = 16$, e $\tau = 60$.



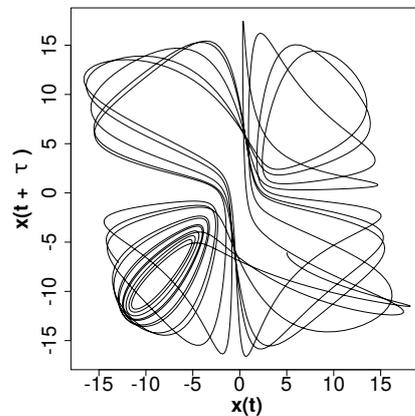
(b) Reconstrução com $\tau = 1$.



(d) Reconstrução com $\tau = 8$.



(c) Reconstrução com $\tau = 16$.



(e) Reconstrução com $\tau = 60$.

Fonte: Ilustração adaptada de COSTA (2017).

(Figura 8b), considerando 100 defasagens. A linha horizontal contínua refere-se ao limiar $1/e$ na versão normalizada da AMIF, em que o primeiro *retardo temporal* $\tau = 8$ abaixo desse limiar está destacada por uma linha vertical pontilhada. As linhas verticais pontilhadas seguintes destacam dois mínimos locais referentes às defasagens $\tau = 16$ e $\tau = 60$, respectivamente. Nesse caso, a primeira linha determina o primeiro mínimo local, sugerido como o valor ótimo para a reconstrução do *espaço de fase*.

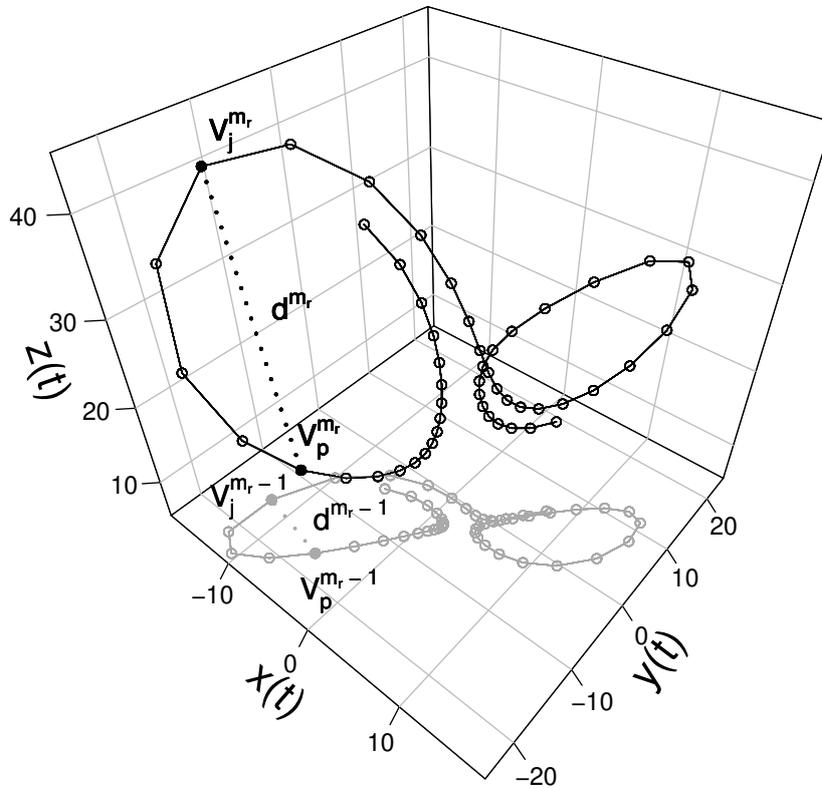
As Figuras 12b, 12c, 12d e 12e exibem, respectivamente, as reconstruções do atrator do Sistema de Lorenz no espaço $\mathbb{R}^{d=2}$, considerando as defasagens $\tau = 1, \tau = 8, \tau = 16$ e $\tau = 60$. Visualmente, as reconstruções mais semelhantes ao atrator original foram obtidas com o parâmetro $\tau = 8$, definido pelo primeiro decaimento abaixo do limiar $1/e$, seguido do primeiro mínimo local, identificado por $\tau = 16$. Para o *retardo temporal* $\tau = 1$, ocorre uma forte correlação entre as componentes dos estados, comprimindo a reconstrução do atrator ao longo de uma linha diagonal. Por outro lado, o *retardo temporal* $\tau = 60$ determina um grande desacoplamento dos estados, provocando um desdobramento muito esparso.

3.3.2 Dimensão de imersão

Uma extensão do *teorema de imersão de Takens*, proposta por Sauer, Yorke e Casdagli (1991), aponta as condições suficientes para a preservação da estrutura topológica do *espaço de fase* quando $m_r \geq 2m_a$, onde m_a é a *dimensão fractal* do atrator e m_r é a *dimensão de imersão*, a dimensão mínima para a reconstrução do atrator (SAUER; YORKE; CASDAGLI, 1991; KANTZ; SCHREIBER, 2004). Vale salientar que a *dimensão fractal* não é um número inteiro, sendo tipicamente relacionada a atratores caóticos (DEVANEY, 1992; KANTZ; SCHREIBER, 2004; SAVI, 2006). Além disso, dependendo da aplicação, uma *dimensão de imersão* menor pode ser suficiente, desde que $m_r > m_a$ (KANTZ; SCHREIBER, 2004). Assim, vários métodos foram propostos para determinar o menor valor para a *dimensão de imersão* m_r (GRASSBERGER; PROCACCIA, 1983; BROOMHEAD; KING, 1986; KENNEL; BROWN; ABARBANEL, 1992; CAO, 1997). Dentre esses métodos, o False Nearest Neighbours (traduzido para o português, Falsos Vizinhos Próximos) (FNN) (KENNEL; BROWN; ABARBANEL, 1992) é um dos mais utilizados na literatura, e fundamenta o método de Cao (CAO, 1997), utilizado nesta tese.

A Figura 13 ilustra a ideia do procedimento, na qual um segmento do atrator (na cor preta) do Sistema de Lorenz no espaço $\mathbb{R}^{d=3}$ ($m_r = 3$) é projetado no plano XY (na cor cinza), ou seja, no espaço $\mathbb{R}^{d=2}$ ($m_r = 2$). Embora ocorram sobreposições na trajetória projetada,

Figura 13 – Ilustração do método FNN para a reconstrução do *espaço de fase*. Na cor preta, um segmento do atrator do Sistema de Lorenz no espaço $\mathbb{R}^{d=3}$. No plano XY , na cor cinza, o mesmo segmento no espaço $\mathbb{R}^{d=2}$.



Fonte: Elaborada pelo autor (2021).

o seguimento do espaço original não apresenta cruzamentos. Além disso, a distância d^{m_r} (linha pontilhada) e a localização dos estados $\mathbf{v}_p^{m_r}$ e $\mathbf{v}_j^{m_r}$ são distorcidas na projeção, sendo identificadas por d^{m_r-1} , $\mathbf{v}_p^{m_r-1}$ e $\mathbf{v}_j^{m_r-1}$, respectivamente. Ao considerar um raio de vizinhança $\varepsilon = d^{m_r-1}$, $\mathbf{v}_p^{m_r-1}$ e $\mathbf{v}_j^{m_r-1}$ serão vizinhos na dimensão $m_r = 2$, mas não com $m_r = 3$, em que $\varepsilon = d^{m_r-1} < d^{m_r}$. Desse modo, por meio de incrementos sucessivos da dimensão m_r , *falsos vizinhos* são identificados.

Considerando uma vizinhança $N(\mathbf{v}_p^{m_r}, \varepsilon)$, definida na Equação 3.11 por um raio ε em um *espaço de fase* de dimensão m_r , a quantidade de supostas projeções $\mathbf{v}_j^{m_r-k}$, referentes aos estados $\mathbf{v}_j^{m_r}$ vizinhos de qualquer outro estado $\mathbf{v}_p^{m_r}$, tendem a aumentar na medida em que k decrementos sucessivos forem realizados na dimensão m_r . Inversamente, essas projeções $\mathbf{v}_j^{m_r-k}$ são contabilizadas como *verdadeiros vizinhos próximos* enquanto permanecerem dentro da vizinhança $N(\mathbf{v}_p^{m_r-k}, \varepsilon)$ ao longo de sucessivos incrementos dimensionais. Caso algum estado $\mathbf{v}_j^{m_r}$ não seja vizinho de fato de um estado $\mathbf{v}_p^{m_r}$ em alguma dimensão superior m_{r+k} obtida em algum incremento dimensional k , conclui-se, portanto, que as supostas projeções

$\mathbf{v}_j^{m_r-k}$ são de fato *falsos vizinhos*.

$$N(\mathbf{v}_p^{m_r}, \varepsilon) = \{\mathbf{v}_j^{m_r} : \|\mathbf{v}_j^{m_r} - \mathbf{v}_p^{m_r}\| \leq \varepsilon\}. \quad (3.11)$$

Desse modo, a estimação da dimensão m_r por meio do método FNN consiste em contabilizar a redução de *falsos vizinhos* a cada incremento dimensional dentro do raio de vizinhança ε . A ideia é reduzir dobras e cruzamentos nas órbitas do *atrator* reconstruído, e progressivamente atingir uma dimensão m_r , idealmente a menor, onde o *atrator* encontre-se significativamente desembaraçado. Por meio desse procedimento, as distâncias em relação a vizinhança $N(\mathbf{v}_p^{m_r}, \varepsilon)$ são comparadas sistematicamente com as distâncias da vizinhança $N(\mathbf{v}_p^{m_{r+1}}, \varepsilon)$ na dimensão seguinte até ocorrer uma diminuição consistente do número de *falsos vizinhos* (KENNEL; BROWN; ABARBANEL, 1992; SMALL, 2005).

A diferença das distâncias ($d^{m_{r+1}} - d^{m_r}$) entre pares de estados de dimensões consecutivas pode ser expressa considerando apenas a distância entre as últimas componentes dos estados da dimensão mais alta, logo

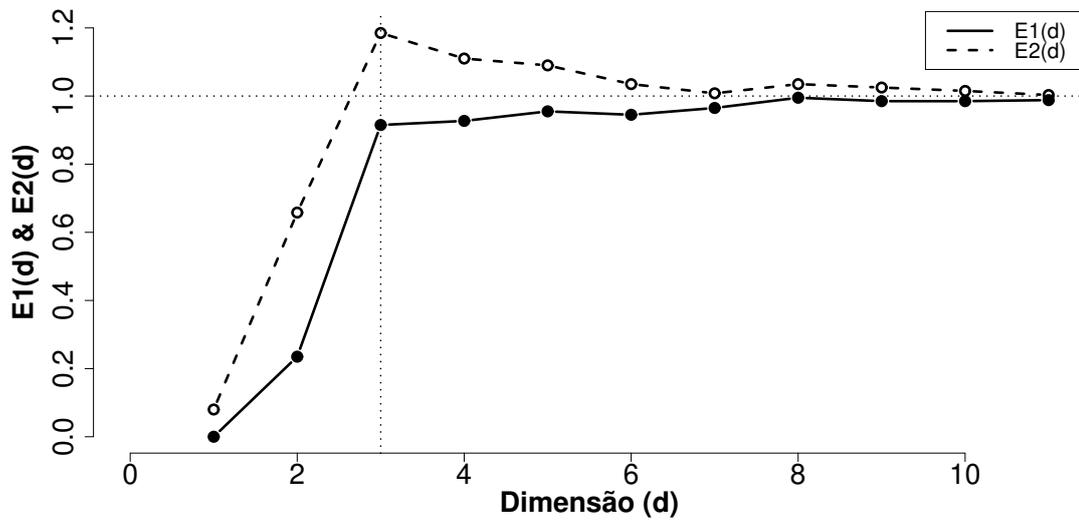
$$(d^{m_{r+1}})^2 - (d^{m_r})^2 = \|\mathbf{v}_p^{m_{r+1}} - \mathbf{v}_j^{m_{r+1}}\|^2 - \|\mathbf{v}_p^{m_r} - \mathbf{v}_j^{m_r}\|^2 = (x_{p+(m_{r+1})\tau} - x_{j+(m_{r+1})\tau})^2. \quad (3.12)$$

Com essa diferença, a Equação 3.13 calcula a proporção das divergências entre as distâncias dos estados, em relação aos incrementos dimensionais, definindo quando dois estados são considerados *falsos vizinhos* a partir de um limiar δ_d . Para valores muito pequenos desse limiar, vizinhos verdadeiros serão contabilizados como falsos. Por outro lado, com um valor muito grande, *falsos vizinhos* serão contabilizados como verdadeiros. De qualquer modo, pares de estados realmente vizinhos permanecerão próximos com incrementos sucessivos da *dimensão de imersão* m_r , estabelecendo uma consistência do *atrator* em relação aos ajustes incrementais dessa dimensão (KENNEL; BROWN; ABARBANEL, 1992; CAO, 1997).

$$\frac{|x_{p+(m_{r+1})\tau} - x_{j+(m_{r+1})\tau}|}{\|\mathbf{v}_p^{m_r} - \mathbf{v}_j^{m_r}\|} \geq \delta_d. \quad (3.13)$$

O limiar δ_d depende da distribuição espacial dos estados, sendo usualmente determinado heurísticamente. Devido a essa subjetividade, Cao (1997) propôs um método prático e mais vantajoso para determinar a *dimensão de imersão* m_r . O método define dois indicadores: o primeiro, denominado $E1(d)$, indica a dimensão de reconstrução apropriada por meio de sua saturação caso a série temporal avaliada derive de algum sistema determinístico. O segundo

Figura 14 – Gráfico dos indicadores $E(1)$ e $E(2)$ obtidos do método de Cao aplicado sobre a série temporal da variável x do Sistema de Lorenz.



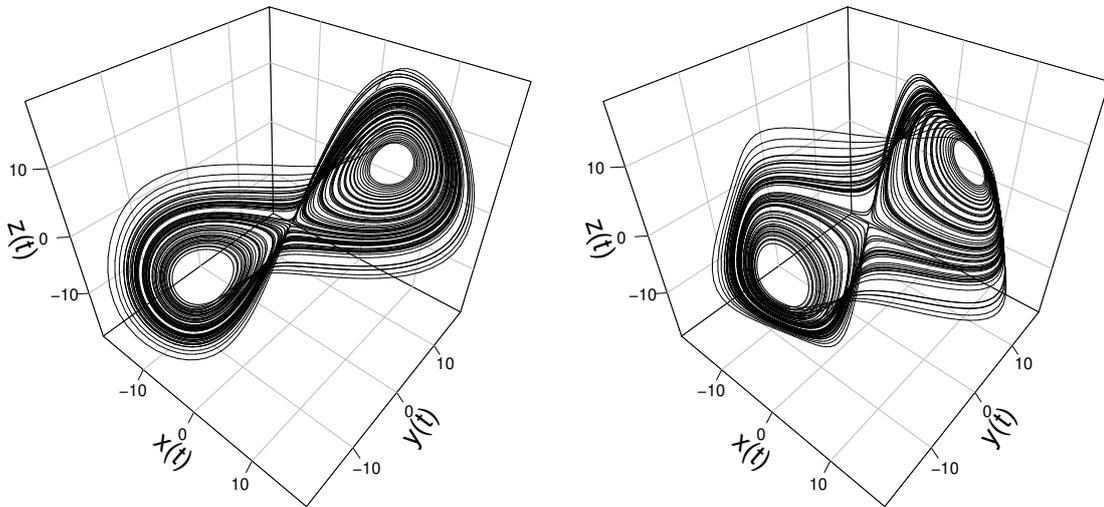
Fonte: Elaborada pelo autor (2021).

indicador, denominado $E2(d)$, identifica distorções na saturação do indicador $E1(d)$, principalmente devido a possibilidade do sistema subjacente ser um processo estocástico. Para séries temporais estocásticas $E2(d) = 1$ para qualquer dimensão d . Caso contrário, deve existir alguma dimensão d de modo que $E2(d) \neq 1$. De qualquer modo, deve-se calcular ambos os indicadores para determinar a *dimensão de imersão*, bem como para distinguir se a série temporal deriva de um sistema determinístico ou de um processo estocástico (CAO, 1997).

A Figura 14 exibe o gráfico do método de Cao aplicado sobre a série temporal da variável x do Sistema de Lorenz considerando o *retardo temporal* $\tau = 16$. No gráfico, ambos os indicadores $E(1)$ e $E(2)$ mudam bruscamente a inclinação ascendente a partir da dimensão $d = 3$ (linha vertical pontilhada). Além disso, os indicadores começam a convergir após essa dimensão, ambos saturando próximo ao valor 1,0 (linha horizontal pontilhada) a partir da dimensão $d = 6$.

Assim como na estimação do *retardo temporal* τ por meio da AMIF, a estimação da *dimensão de imersão* m_r por meio do método de Cao é bastante razoável. Com esses resultados, a dinâmica do *atrator* do Sistema de Lorenz (Figura 8a) pode ser reconstruída de forma aproximada, considerando as defasagens $\tau = 8$, $\tau = 16$, e dimensão $m_r = 3$, como ilustrado na Figura 15. Apesar das reconstruções apresentadas nas Figuras 15a e 15b não representarem fielmente o *atrator* original, suas estruturas ainda possibilitam inferir propriedades importantes do sistema subjacente (SCHREIBER, 1999).

Figura 15 – Reconstrução do atrator por meio da série temporal da variável x do Sistema de Lorenz, considerando a dimensão estimada $m_r = 3$.



(a) Atrator reconstruído com $\tau = 8$.

(b) Atrator reconstruído com $\tau = 16$.

Fonte: Elaborada pelo autor (2021).

3.4 DETECTANDO ALTERAÇÕES NA DINÂMICA DE SÉRIES TEMPORAIS NÃO LINEARES

Várias técnicas foram propostas para distinguir mudanças na dinâmica de séries temporais não lineares (TRULLA et al., 1996; BANDT; POMPE, 2002; CAO et al., 2004; MARWAN et al., 2007; DONNER et al., 2010). Em geral, essas técnicas estão relacionadas aos principais tipos de medidas de complexidade, como entropias, dimensões fractais, e expoentes de Lyapunov, referentes aos sistemas dinâmicos não lineares (KANTZ; SCHREIBER, 2004; SAVI, 2006).

Quando as equações que descrevem um sistema dinâmico não linear estão disponíveis, pode-se avaliar a sensibilidade às condições iniciais por meio da divergência exponencial entre trajetórias vizinhas. A partir de um modelo matemático, o algoritmo proposto por Wolf et al. (1985) pode ser usado para determinar os *expoentes de Lyapunov* (SAVI, 2006). O maior expoente indica mudanças na dinâmica do sistema por meio da diferenciação entre regimes caóticos e regulares ao longo do tempo (TRULLA et al., 1996; COSTA, 2017). Embora os expoentes de Lyapunov possam ser estimados a partir de uma série temporal (WOLF et al., 1985; ROSENSTEIN; COLLINS; De Luca, 1993; KANTZ, 1994), os resultados podem não ser satisfatórios, a depender da qualidade dos dados, do nível de ruído, da parametrização do método utilizado, e da representação da dinâmica do sistema subjacente (SCHREIBER, 1999; KANTZ; SCHREIBER, 2004). Essas restrições levaram ao desenvolvimento de técnicas alternativas para

avaliar mudanças na dinâmica de séries temporais não lineares. As próximas seções descrevem algumas das técnicas mais populares, visando identificar transições de regimes dinâmicos, as quais fundamentam os métodos propostos no Capítulo 4.

3.4.1 Análise de Quantificação de Recorrência

A inspeção visual de *espaços de fase* com alta dimensionalidade restringe-se a uma projeção no espaço $\mathbb{R}^{d=2}$ ou $\mathbb{R}^{d=3}$. Entretanto, a análise de *recorrências* possibilita avaliar a complexidade da dinâmica em espaços de fase com dimensões elevadas, principalmente devido à introdução do Diagrama de Recorrência (RP) (ECKMANN; KAMPHORST; RUELLE, 1987; MARWAN et al., 2007). *Recorrência* é uma propriedade fundamental dos sistemas dinâmicos que pode ser explorada para caracterizar séries temporais com dinâmicas distintas (MARWAN et al., 2007). Considerando um critério de vizinhança, o Diagrama de Recorrência (RP) mapeia as correlações entre os estados em uma representação gráfica de padrões binários bidimensional. Trata-se de uma ferramenta poderosa para descrever o comportamento do sistema abrangendo diferentes escalas, independentemente da estacionariedade e não linearidade dos dados, bem como da dimensão do *espaço de fase* (CASDAGLI, 1997; SCHREIBER, 1999; MARWAN et al., 2007).

Na proposta original (ECKMANN; KAMPHORST; RUELLE, 1987), somente os k-Nearest Neighbors (traduzido para o português, k-Vizinhos Mais Próximos) (k-NN) de um estado particular \mathbf{v}_p são considerados. Entretanto, uma forma alternativa para definir a vizinhança é usar um raio ε como limiar (MARWAN et al., 2007; DONNER et al., 2011):

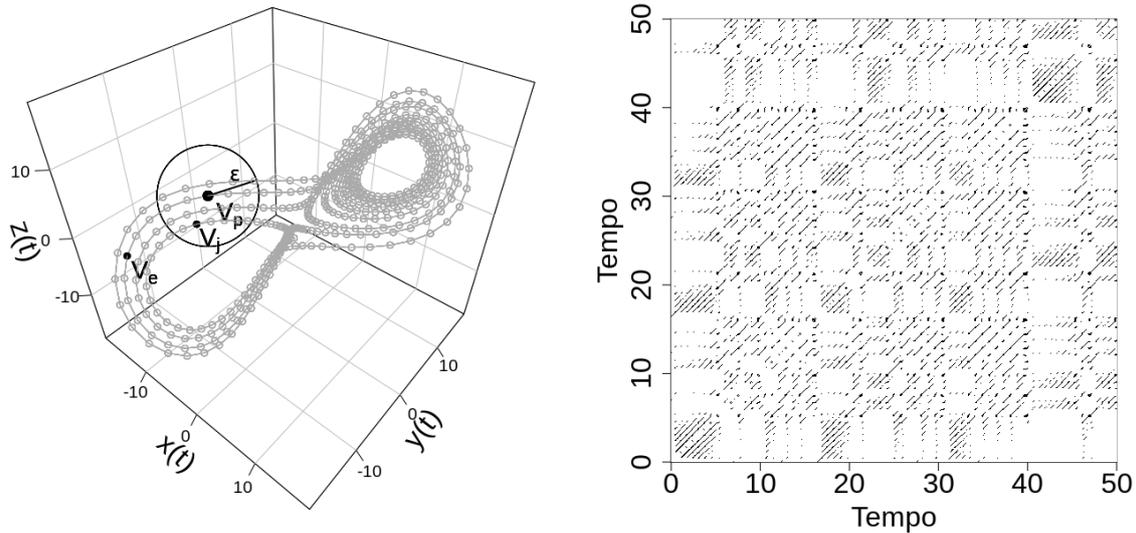
$$R_{p,j}(\varepsilon) = \Theta(\varepsilon - \|\mathbf{v}_p - \mathbf{v}_j\|), \quad (3.14)$$

onde $\Theta(\cdot)$ é a função degrau unitária

$$\Theta(r) = \begin{cases} 1, & (r \geq 0); \\ 0, & (r < 0). \end{cases} \quad (3.15)$$

Em geral, a análise requer alguma heurística para estimar o raio de vizinhança ε , em que um valor ótimo depende da aplicação (MARWAN et al., 2007). Caso ε seja muito pequeno, pode não haver nitidez do padrão de *recorrências* do sistema subjacente representado no RP. Por outro lado, para um valor muito grande, quase todos os pontos serão considerados vizinhos,

Figura 16 – Ilustração do RP sobre o atrator reconstruído a partir da série temporal da variável x do Sistema de Lorenz. A Figura (a) representa o processo de construção do RP apresentado na Figura (b).



(a) Ilustração da construção de um RP a partir de um segmento do atrator do Sistema de Lorenz reconstruído, usando o critério de raio de vizinhança.

(b) RP do atrator reconstruído do Sistema de Lorenz, considerando o *retardo temporal* $\tau = 8$, dimensão $m_r = 3$, e raio $\epsilon = 3,0$.

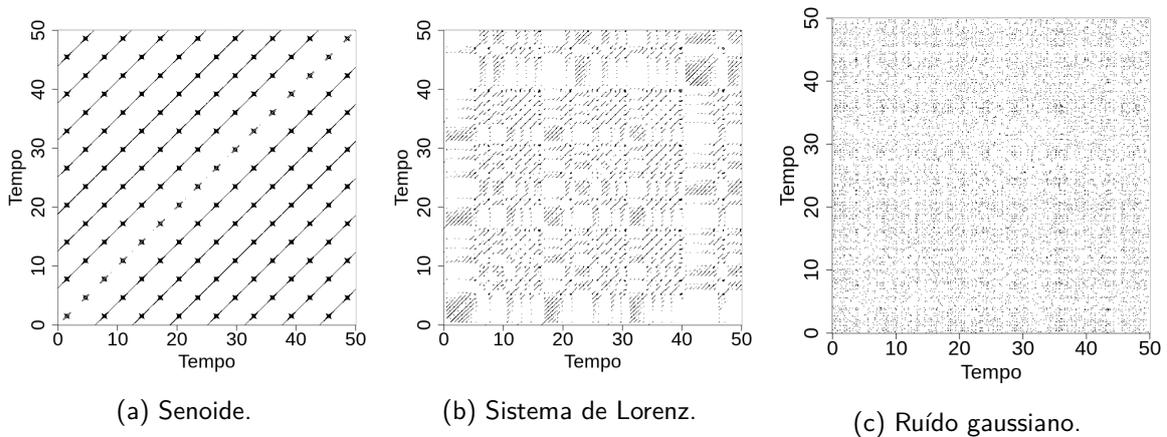
Fonte: Elaborada pelo autor (2021).

consequentemente, ocorrerão muitas *recorrências* espúrias. De acordo com a Equação 3.14, o RP consiste em uma matriz binária, na qual um ponto é colocado em $\mathbf{R}_{p,j}(\epsilon)$ sempre que v_j está suficientemente perto de v_p , considerando o limiar ϵ . Assim, $\mathbf{R}_{p,j}(\epsilon) = 1$ para cada estado v_j vizinho de v_p , do contrário, $\mathbf{R}_{p,j}(\epsilon) = 0$. Para a definição do RP, algumas heurísticas sugerem uma densidade de pontos em torno de 1%, e que o limiar ϵ não ultrapasse 10% do diâmetro do *espaço de fase* (MARWAN et al., 2007).

A Figura 16 ilustra o processo de construção de um RP, usando o critério de raio de vizinhança, a partir da série temporal da variável x do Sistema de Lorenz. A Figura 16a apresenta uma discretização de um segmento do atrator reconstruído, já ilustrado na Figura 15a, e uma região de vizinhança delimitada por um raio ϵ , centrada no estado v_p (ponto preto maior). Os estados contidos nessa vizinhança (circulo preto), como o estado v_j , são marcados com pontos pretos ($\mathbf{R}_{p,j}(\epsilon) = 1$) na Figura 17b. Do contrário, estados externos à região de vizinhança, como o estado v_e , não são marcados na Figura 17b ($\mathbf{R}_{p,j}(\epsilon) = 0$). A Figura 17b apresenta o RP do atrator reconstruído a partir da série temporal da variável x do Sistema de Lorenz (Figura 15a), considerando o *retardo temporal* $\tau = 8$, a dimensão $m_r = 3$, e definido com $\epsilon = 3,0$.

Por meio da construção de RPs, diferentes tipos de regimes podem ser caracterizados a

Figura 17 – Exemplos de diagramas de recorrência para um sistema com dinâmica periódica (a), caótica (b) e estocástica (c).



Fonte: Ilustração adaptada de Donner et al. (2011).

partir de padrões relacionados a correlações dos estados do sistema subjacente. Os padrões podem ser visualmente inspecionados para analisar a evolução de linhas diagonais ou verticais, o que permite identificar comportamentos periódicos, caóticos e estocásticos (MARWAN et al., 2007), como exemplificados na Figura 17. Mudanças no comportamento de séries temporais podem ser detectadas por meio da avaliação dos padrões presentes no RP, em macro ou em micro escala (TRULLA et al., 1996; MARWAN et al., 2007; DONNER et al., 2010; DONNER et al., 2011).

A subjetividade relacionada à análise de estruturas complexas por meio da inspeção visual dos padrões presentes no RP motivaram o desenvolvimento da Análise de Quantificação de Recorrência (RQA) (MARWAN et al., 2007). Várias medidas de complexidade para quantificar as características dessas estruturas popularizaram-se em decorrência da RQA (TRULLA et al., 1996; MARWAN et al., 2007; DONNER et al., 2010; DONNER et al., 2011). Essa popularidade está relacionada a sua eficiência em extrair informações acerca da complexidade de processos em diferentes domínios, com o objetivo de descrever transições na dinâmica de sistemas a partir de séries temporais (CASDAGLI, 1997; TRULLA et al., 1996; MARWAN et al., 2007; DONNER et al., 2010; DONNER et al., 2011). Além disso, a relação entre RPs com matrizes de adjacências possibilita a construção e a análise de *redes complexas* relacionadas ao *espaço de fase* do sistema dinâmico subjacente, principalmente para a identificação de propriedades topológicas em *espaços de fase* reconstruídos com alta dimensionalidade (DONNER et al., 2010; DONNER et al., 2011). Apesar da sua versatilidade, a construção de um RP requerer $\mathcal{O}(n^2)$ operações (COSTA et al., 2017).

3.4.2 Entropia de permutação

Assim como a RQA, a Entropia de Permutação (PE) tem sido aplicada em diversos domínios na análise da complexidade de séries temporais (BANDT; POMPE, 2002; CAO et al., 2004; CUESTA-FRAU, 2019). Trata-se de uma ferramenta conceitualmente simples, robusta a ruídos, e computacionalmente eficiente para quantificar a estrutura intrínseca dos dados por meio da dinâmica simbólica (BANDT; POMPE, 2002; CAO et al., 2004; AMIGÓ, 2010).

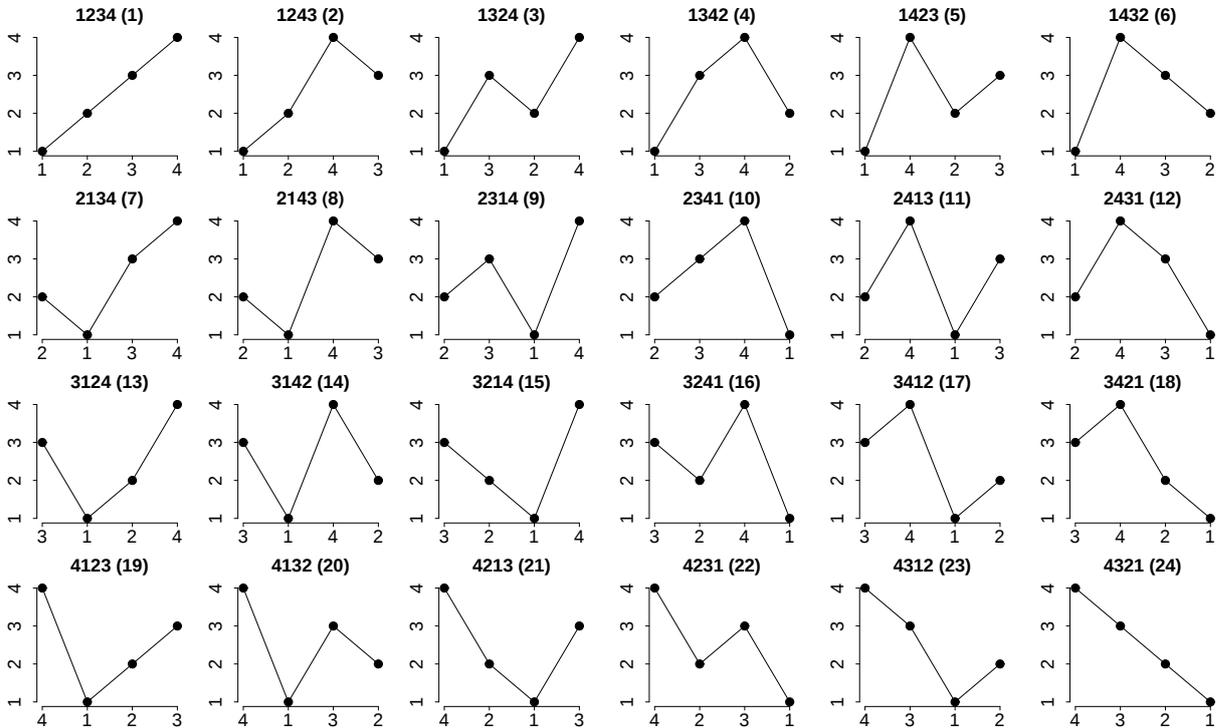
Essencialmente, um sistema dinâmico pode ser adequadamente representado e analisado como uma sequência de símbolos (CAO et al., 2004; AMIGÓ, 2010). Nessa perspectiva, as componentes $(x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$ de cada estado \mathbf{v}_i são interpretadas como permutações de padrões ordinais mapeados em $m!$ símbolos. Assim, diferentemente da maioria das ferramentas de análise do *espaço de fase*, fundamentadas em medidas de dissimilaridades entre os estados (SCHREIBER, 1999; MARWAN et al., 2007; DONNER et al., 2010), PE tem por base a distribuição estatística desses padrões, ou símbolos (CAO et al., 2004). Basicamente, o conjunto de padrões ordinais representa um histograma em que o ordenamento das componentes de cada estado são contabilizados como um símbolo rotulado entre 1 e $m!$. Após considerar todas as permutações ordinais, as distribuições das frequências relativas P_1, P_2, \dots, P_k são obtidas para cada símbolo distinto k , onde $k \leq m!$. Por fim, a medida de complexidade é definida na Equação 3.16 como a *Entropia de Shannon* (SHANNON, 2001) sobre os símbolos k (CAO et al., 2004):

$$H(m) = - \sum_{i=1}^k P_i \log_2 P_i. \quad (3.16)$$

O limite superior para a Equação 3.16 ocorre quando os dados são uniformemente aleatórios (uma série temporal i.i.d.), onde todos os $m!$ possíveis símbolos possuem a mesma frequência estatística c , de modo que $P_i = \frac{c}{c \times m!} = \frac{1}{m!}$. Logo, $H(m) < \log_2 m!$ sugere que a série temporal apresenta algum nível de determinismo (BANDT; POMPE, 2002).

Como exemplo, a Figura 18 apresenta os padrões de todas as ordenações possíveis para $m = 4$ ($4! = 24$ símbolos). Em cada padrão, o eixo das abscissas corresponde aos índices das permutações após a ordenação ascendente das componentes $(x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$ de qualquer estado \mathbf{v}_i no espaço $\mathbb{R}^{d=4}$. No topo de cada padrão, uma sequência numérica determina a ordenação, e um índice (em parênteses) identifica os respectivos símbolos. Nesta pesquisa, os padrões ordinais também foram utilizados na rotulação de estados de modo a viabilizar a aplicação de modelos adaptativos supervisionados a fim de detectar mudanças na dinâmica do

Figura 18 – Padrões de todas as ordenações possíveis para $m = 4$ ($4! = 24$ símbolos). O eixo das abscissas corresponde aos índices de permutação das componentes $(x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$ para qualquer estado v_i no espaço $\mathbb{R}^{d=4}$.



Fonte: Elaborada pelo autor (2021).

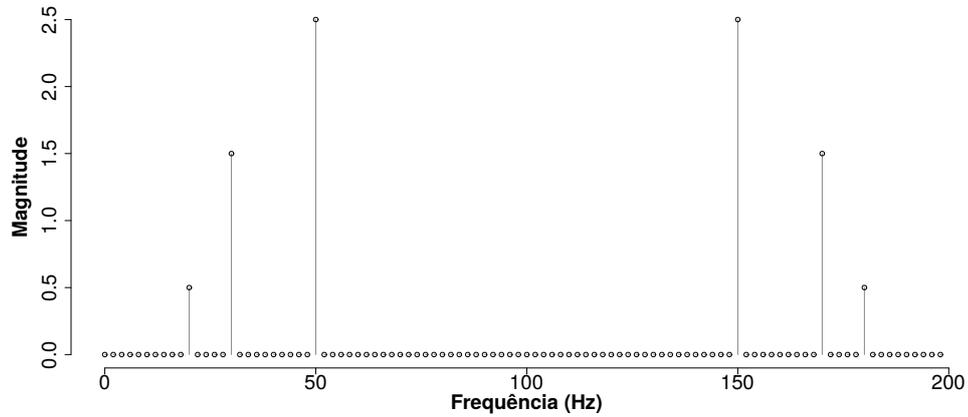
processo subjacente aos dados. O procedimento para o cálculo da PE requer $\mathcal{O}(n)$ operações, onde n é o número de elementos de uma série temporal discreta (COSTA et al., 2017).

3.4.3 Entropia espectral

A informação presente em uma série temporal pode ser descrita por meio de entropias (KANTZ; SCHREIBER, 2004). No domínio temporal, caso as observações sejam independentes, a *Entropia de Shannon* pode ser determinada empiricamente a partir da distribuição de probabilidade das observações. Analogamente, a Entropia Espectral (SE) (POWELL; PERCIVAL, 1979) é uma medida de entropia no domínio espectral que permite quantificar o espalhamento espectral obtido a partir do Espectro de Potência (PS) (POWELL; PERCIVAL, 1979).

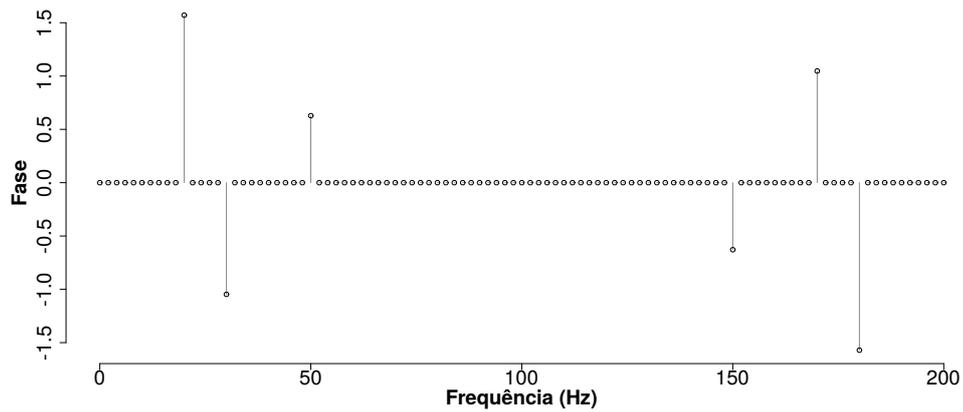
O PS não contém qualquer informação de fase ϕ_ξ , mas contém informações das magnitudes A_ξ e das frequências ξ da série temporal original. Dado que ambos os componentes espectrais $\{A_\xi, \phi_\xi\}$ podem ser determinados usando a Fast Fourier Transform (traduzido para o português, Transformada Rápida de Fourier) (FFT) (COOLEY; TUKEY, 1965), as Figuras 19 e 20 apresentam os gráficos das magnitudes (A_ξ) e ângulos de fase (ϕ_ξ) obtidos da de-

Figura 19 – Magnitudes (A_ξ) relacionadas às frequências (ξ) obtidas como componentes espectrais da série temporal apresentada na Figura 1.



Fonte: Elaborada pelo autor (2021).

Figura 20 – Ângulos de fase (ϕ_ξ) em radianos, relacionados às respectivas magnitudes (A_ξ) apresentadas na Figura 19.



Fonte: Elaborada pelo autor (2021).

composição da série temporal ilustrada na Figura 1 usando a FFT. Note que as magnitudes A_ξ encontram-se espelhadas em torno de 100 Hz. De modo semelhante, na Figura 20 os ângulos de fase encontram-se espelhado em torno de 100 Hz, mas também estão rebatidos. Considerando apenas as informações das magnitudes A_ξ , o PS é definido como $PS_\xi = (A_\xi)^2$.

$$S_{\xi_i} = \frac{PS_{\xi_i}}{\sum_{j=1}^N PS_{\xi_j}} \quad (3.17)$$

A Equação (3.17) define a normalização S_{ξ_i} dos componentes espectrais PS_{ξ_j} , onde N é o número de intervalos do histograma de frequências ξ_i , assumindo intervalos de uma unidade espectral. Assim, os componentes S_{ξ_i} são usados como termos da formulação da *Entropia de Shannon* na Equação (3.18), onde o fator de normalização $\log_2(N)$ representa a entropia de um PS uniformemente distribuído (INOUYE et al., 1991; SLEIGH et al., 2004).

$$H(\xi) = \sum_{i=1}^N H_{\xi_i}, \text{ onde } H_{\xi_i} = -\frac{S_{\xi_i} \log_2 S_{\xi_i}}{\log_2(N)}. \quad (3.18)$$

Finalmente, a Equação (3.18) define a SE como a soma de todos os termos H_{ξ_i} . Essa medida foi estendida por Inouye et al. (1991) para quantificar a irregularidade do eletroencefalograma (EEG), sendo comumente usada para monitorar, em tempo real, a atividade elétrica cerebral. Além disso, a SE está de acordo com o Surrogate Stability Concept (SSC) proposto por Vallim e Mello (2014) para a detecção de *mudanças de conceito*.

O SSC tem por base o Theiler's Surrogate Fourier Transformed Method (TSFTM) (THEILER et al., 1992), uma das ferramentas da NLTA mais utilizadas para testes de não linearidade em séries temporais, sendo formalizado em termos das divergências de modelos baseados em *séries substitutas* (do Inglês *surrogates*), as quais são novas observações de dados gerados usando a série temporal original como entrada, mas preservando as amplitudes associadas a cada frequência (VALLIM; MELLO, 2014). Os autores assumem que as observações não são necessariamente i.i.d., e afirmam que qualquer algoritmo que produz um PS como modelo está de acordo com esse conceito de estabilidade. Além disso, com o uso da FFT, a SE pode ser calculada com complexidade assintótica de $\mathcal{O}(n \log_2 n)$. Portanto, SE é um método rápido e eficiente para quantificar o PS, de modo que um PS espalhado implica em uma SE próxima de um, enquanto valores próximo de zero sugerem um PS condensado em poucas frequências.

3.5 RESUMO DO CAPÍTULO

Este capítulo apresentou os conceitos da NLTA como abordagem alternativa para lidar com o problema da dependência temporal ao invés de presumir a hipótese de observações i.i.d., com base na interpretação de séries temporais como fluxo de dados. As seções finais apresentaram os métodos para avaliar alterações da dinâmica do processo gerador dos dados por meio do espaço de fase, considerando recorrências (Diagrama de Recorrência (RP), Análise de Quantificação de Recorrência (RQA)) e a dinâmica simbólica (Entropia de Permutação (PE)), bem como, por meio da análise espectral (Entropia Espectral (SE)).

4 PROPOSTA DE DETECTORES E PSEUDO-ROTULAÇÃO SIMBÓLICA

Este capítulo descreve as abordagens propostas para a detecção de *mudança de conceito* fundamentadas na NLTA, bem como a estratégia para rotulação automática por meio da dinâmica simbólica. As abordagens de detecção abrangem algoritmos baseados na RQA, PE, e SE. A estratégia de rotulação utiliza a relação entre os estados em um *espaço de fase* reconstruído com padrões ordinais, viabilizando a aplicação de classificadores a fluxo de dados unidimensionais como séries temporais com a finalidade de identificar *mudanças de conceito*.

4.1 AVALIAÇÃO ENTRE JANELAS DE OBSERVAÇÕES

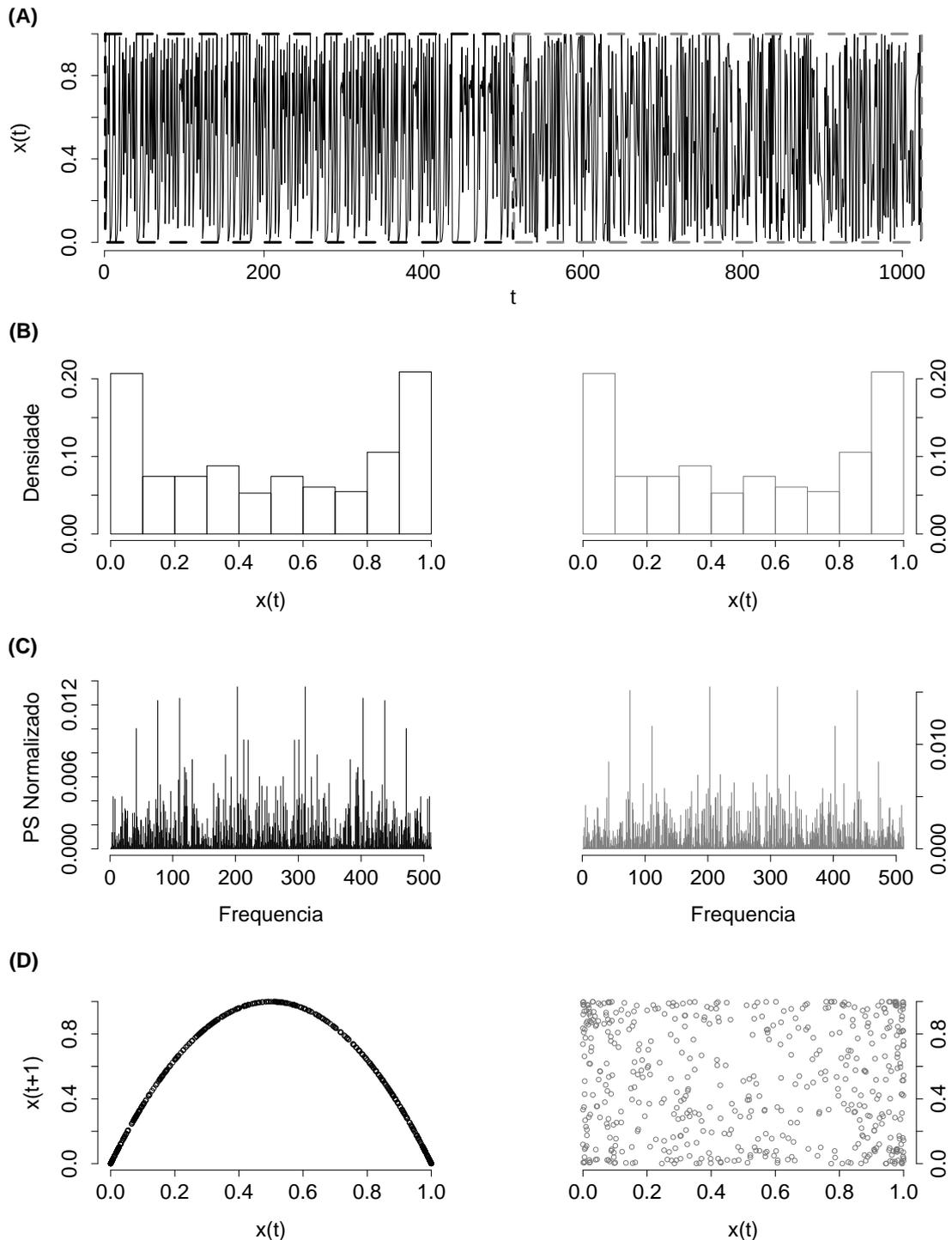
Como descrito na Seção 2.2.1.3, uma das formas mais utilizadas para se avaliar mudanças no comportamento de um *fluxo de dados* é realizar comparações de amostras consecutivas dos dados entre duas janelas de observações. O critério de avaliação das amostras entre as janelas é definido por algum teste estatístico, o qual determina o quanto as distribuições são significativamente distintas ao ponto de confirmar uma *mudança de conceito*. Entretanto, outras formas de avaliações são possíveis, as quais podem revelar informações não acessíveis quando apenas as distribuições dos dados são utilizadas.

Como exemplo, a Figura 21 ilustra o uso de um histograma, do Espectro de Potência (PS), e do *espaço de fase*, para avaliar mudanças em um *fluxo de dados*. A Figura 21 (A) mostra um fluxo sintético com duas janelas de observações consecutivas. A primeira janela, destacada com linhas tracejadas pretas, apresenta um fluxo com dependências temporais, produzidas usando o Mapa Logístico, definido como $x_{t+1} = r \times x_t(1 - x_t)$, onde $x_0 = 0,7$, e $r = 4,0$. A segunda janela, destacada com linhas tracejadas em cinza, apresenta um novo fluxo obtido ao distribuir aleatoriamente os dados da primeira janela.

A Figura 21 (B) apresenta os histogramas produzidos usando suas respectivas janelas de dados. Observe que não é possível distinguir qualquer mudança avaliando apenas as distribuições. De modo semelhante, a Figura 21 (C) apresenta os respectivos PS normalizados, onde diferenças sutis das distribuições das magnitudes entre as janelas podem ser identificadas. Finalmente, a Figura 21 (D) mostra os respectivos *espaços de fase*, os quais são muito diferentes.

Nota-se que informações relevantes da dependência temporal entre as observações presen-

Figura 21 – (A) Um fluxo de dados sintético com duas janelas de observações consecutivas, destacadas com linhas tracejadas na cor preta e cinza, respectivamente; (B) seus respectivos histogramas (barras pretas e cinzas) em que não é possível identificar qualquer diferença; (C) espectros de potência normalizados, os quais apresentam diferenças sutis entre as janelas; Finalmente, (D) os respectivos espaços de fase, os quais são significativamente diferentes.



Fonte: Elaborada pelo autor (2021).

tes na Figura 21 (D) são desconsideradas quando apenas as distribuições são avaliadas, como ilustrado na Figura 21 (B). Além disso, como ilustrado ao longo da Figura 21, mesmo que as distribuições sejam semelhantes, a dinâmica do *fluxo de dados* pode não ser mais a mesma. Outro ponto é que embora o PS não apresente diferenças significativas entre as janelas de avaliações, a sua aplicação não está restrita à observações i.i.d..

4.2 MAPEAMENTO DO FLUXO DE ERROS EM UMA SÉRIE TEMPORAL ESCALAR

Apesar do fato de que a taxa de erro pode ser influenciada pela dependência temporal, as abordagens apresentadas na Seção 2.2.1 assumem que o fluxo de erro das predições é um conjunto de observações estatisticamente independentes. Consequentemente, nenhuma garantia teórica pode ser fornecida por essas abordagens diante de um cenário onde esse fluxo apresente dependência temporal (ŽLIOBAITĖ et al., 2015; BIFET, 2017).

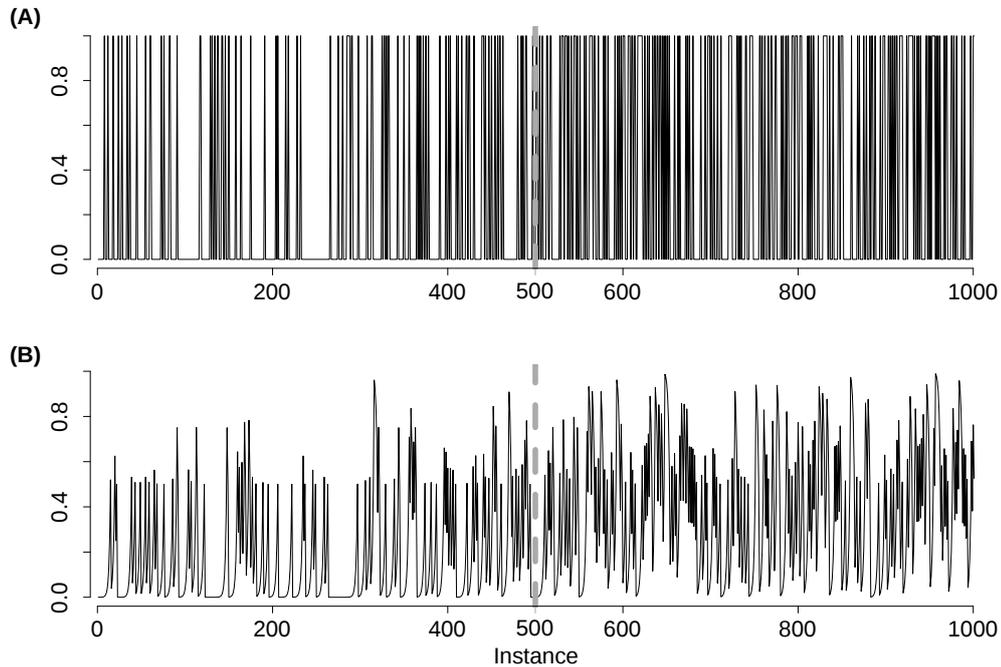
4.2.1 Aplicação do Mapa de Bernoulli

Como apresentado na Seção 3.2.1.1, o Mapa de Bernoulli (OTT, 2002) descreve algumas propriedades relacionadas aos sistemas dinâmicos caóticos. Embora a Equação (3.5) represente cada valor x_i por uma expansão infinita, aplicações reais estão limitadas a trabalhar apenas com expansões finitas em um computador digital. Devido a essa restrição, os detectores propostos realizam cada aplicação do Mapa de Bernoulli sobre o fluxo de erros considerando uma resolução de L bits, de forma que a Equação (3.5) torna-se:

$$\begin{aligned} x_0 = 0.b_1b_2\dots b_L &= \sum_{j=1}^L 2^{-j}b_j = \frac{b_1}{2^1} + \dots + \frac{b_L}{2^L} = \frac{b_12^{L-1}}{2^L} + \frac{b_22^{L-2}}{2^L} + \dots + \frac{b_L2^0}{2^L} = \\ &= \frac{1}{2^L} \sum_{j=1}^L 2^{(j-1)}b_{L-(j-1)}. \end{aligned} \quad (4.1)$$

Desse modo, a aplicação sucessiva do mapa sobre o fluxo de erro $b_1b_2b_3\dots$ produz a série temporal $x_0, x_1, x_2, x_3, \dots$, onde cada $x_i = 0.b_{i+1}b_{i+2}\dots b_{(L+i-1)}b_{(L+i)}$. Apesar da restrição, a Equação 4.1 é uma aproximação do mapa de Bernoulli realizado em computadores digitais em que os bits menos significativos (LSB) são substituídos por uma nova sequência de bits (ÖZTÜRK; KILIÇ, 2018). Como exemplo, $x_0^C \cong 0,2764 \equiv 0.0100011011000010001001101\dots$ obtido com a Equação 3.5 na Seção 3.2.1.1, é expresso como $x_0^C = \frac{0}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{0}{2^4} + \frac{0}{2^5} + \frac{1}{2^6} + \frac{1}{2^7} + \frac{0}{2^8} + \frac{1}{2^9} + \frac{1}{2^{10}} = 0,2763671875$, de acordo com a Equação 4.1, onde $L = 10$.

Figura 22 – (A) Amostra do fluxo de erro com o ponto de mudança na instância 500 (linha cinza tracejada); (B) A série temporal obtida após a aplicação do Mapa de Bernoulli.



Fonte: Elaborada pelo autor (2021).

Embora as iterações digitais do Mapa de Bernoulli sempre convirjam para zero, por causa do limite da unidade de memória de L -bits, esta pesquisa utiliza o *método de extensão* LSB (ÖZTÜRK; KILIÇ, 2018) considerando o fluxo de erro como entrada ao invés de uma sequência de bits obtida de um *Gerador de Número Aleatório* (RNG). Assim, após cada operação de deslocamento de bit à esquerda (denotado por $\ll 1$), o bit menos significativo (LSB) em uma unidade de memória de L -bits é substituído pela avaliação da predição atual (0 se estiver correta ou 1 se incorreta).

A Figura 22 ilustra esse mapeamento. A Figura 22 (A) apresenta um segmento do fluxo de erro com o ponto de mudança na instância 500 (linha tracejada cinza), enquanto a Figura 22 (B) mostra a série temporal obtida pela aplicação o Mapa de Bernoulli usando uma resolução $L = 32$ bits.

Uma vez que o fluxo de erro pode apresentar padrões de bits aleatórios ou regulares ao longo do tempo, pode-se obter aproximações das órbitas caóticas ou periódicas do mapa. Assim, por meio desse mapeamento, transforma-se o fluxo de erro em uma série temporal escalar com o objetivo de aplicar abordagens mais elaboradas que levem em conta a dependência temporal e não sejam restritas à suposições de dados i.i.d.

4.2.1.1 Estimador da Entropia Espectral

Como já descrito na Seção 3.4.3, a SE é uma medida do domínio espectral, obtida a partir do PS normalizado. Essa medida foi estendida por Inouye et al. (1991) para quantificar irregularidades em sinais de Eletroencefalograma (EEG), sendo comumente usada para o monitoramento em tempo real da atividade elétrica cerebral (SLEIGH et al., 2004).

SE está de acordo com o SSC proposto por Vallim e Mello (2014) para a detecção de *mudanças de conceito*, o qual assume que as observações não são necessariamente i.i.d (VALLIM; MELLO, 2014) e tem por base o TSFTM (THEILER et al., 1992). Os autores afirmam que qualquer algoritmo que produza um PS como modelo será estável de acordo com o SSC. Devido a sua estabilidade em relação aos dados de entrada, o SSC foi usado nesta pesquisa com o objetivo de avaliar a sua contribuição na redução do número de alarmes falsos na detecção de *mudanças de conceito*.

Tabela 1 – Algoritmo para estimar a Entropia Espectral.

Algoritmo 1: Entropia Espectral.

Input: $\{x_1, x_2, \dots, x_N\}$

- 1 $h \leftarrow s \leftarrow 0.0$
- 2 **for** $\xi \leftarrow 1$ **to** N **do**
- 3 $A_\xi \leftarrow \phi_\xi \leftarrow 0.0$
- 4 **for** $\xi \leftarrow 1$ **to** $N/2 + 1$ **do**
- 5 $PS_\xi \leftarrow P_\xi \leftarrow 0.0$
- 6 $\{\mathbf{A}, \phi\} \leftarrow FFT(\{x_1, x_2, \dots, x_N\})$
- 7 $PS_1 \leftarrow (A_1)^2$
- 8 $PS_{N/2+1} \leftarrow (A_{N/2+1})^2$
- 9 $s \leftarrow PS_1 + PS_{N/2+1}$
- 10 **for** $\xi \leftarrow 2$ **to** $N/2$ **do**
- 11 $PS_\xi \leftarrow 2 \times (A_\xi)^2$
- 12 $s \leftarrow s + PS_\xi$
- 13 **for** $\xi \leftarrow 1$ **to** $N/2 + 1$ **do**
- 14 $P_\xi \leftarrow PS_\xi / s$
- 15 **if** $P_\xi > 0.0$ **then**
- 16 $h \leftarrow h - P_\xi \times (\log_2 P_\xi) / \log_2(N/2 + 1)$

Output: h

Fonte: Elaborado pelo autor (2021).

O Algoritmo 1 apresenta o pseudocódigo para calcular a SE. Para ilustrar esse procedimento, considere a série temporal $\{-2, 875; 0, 125; 2, 125; 3, 125; -0, 875; 4, 125; -3, 875;$,

$-1,875\}$, onde o número de observações N deve ser uma potência de 2, de acordo com o algoritmo da FFT (COOLEY; TUKEY, 1965).

As linhas 1 a 5 referem-se à inicialização de variáveis. Na linha 6, as magnitudes $\mathbf{A} = \{0, 0; 10, 721; 3, 606; 6, 86; 11, 0; 6, 86; 3, 606; 10, 721\}$ de cada frequência no domínio espectral são calculadas usando a FFT. Observe que as informações de fase (ϕ) são descartadas (não são usadas para calcular a SE). As magnitudes $(A_\xi)^2$ (elevadas ao quadrado) retornam o PS em dois segmentos simétricos de um vetor, com uma metade representando a energia exibida nos componentes de frequência à esquerda, e na outra metade, a energia exibida nos componentes à direita.

Em seguida, as linhas 7-12 descartam a segunda metade do vetor e multiplicam todas as magnitudes A_ξ por 2, exceto $PS_{N/2}$ e PS_0 , de modo a converter os dois segmentos do vetor em um único segmento para descrever o PS, em que $\mathbf{PS} = \{0, 0; 229, 880; 26, 006; 94, 119; 121, 0\}$. Então, a linha 14 normaliza esse PS unilateral de modo a obter o histograma $\mathbf{P} = \{0, 0; 0, 488, 0, 055; 0, 200; 0, 257\}$.

Finalmente, $h = -\frac{0,488 \log_2(0,488)}{\log_2(5)} - \frac{0,055 \log_2(0,055)}{\log_2(5)} - \frac{0,200 \log_2(0,200)}{\log_2(5)} - \frac{0,257 \log_2(0,257)}{\log_2(5)} = 0,734$ computa a entropia de Shannon, calculada nas linhas 15-16, onde $1/\log_2(N/2 + 1)$ é um fator de normalização.

4.2.1.2 Estimador do Determinismo da RQA

Como apresentado na Seção 3.4.1, a recorrência é uma propriedade fundamental dos sistemas dinâmicos que pode ser explorada para caracterizar o comportamento do sistema por meio do *espaço de fase* (MARWAN et al., 2007). Apesar da RQA (TRULLA et al., 1996) fornecer diversas medidas quantitativas sobre a matriz de recorrências, as quais podem ser aplicadas para detectar mudanças dos regimes dinâmicos em diferentes tipos de séries temporais (TRULLA et al., 1996; MARWAN et al., 2007), esta pesquisa restringe-se a calcular apenas o *determinismo percentual* RQA_{DET} , caracterizado como o número de pontos que formam linhas diagonais longas, as quais estão relacionadas com o determinismo do sistema avaliado (MARWAN et al., 2007). A ideia é quantificar e identificar a transição do processo do aprendizado entre a fase de possível convergência do modelo durante a estacionariedade do conceito e sua instabilidade devido à *mudança de conceito*.

O Algoritmo 2 descreve o procedimento para calcular o *determinismo percentual* RQA_{DET} . Na linha 2, n guarda o índice do último estado obtido da série temporal $\{x_0, x_1, x_2, \dots, x_{N-1}\}$.

Tabela 2 – Algoritmo para estimar o Determinismo com base na RQA.

Algoritmo 2: Análise de Quantificação de Recorrência - Determinismo.

Input: time series $\{x_0, x_1, x_2, \dots, x_{N-1}\}$; time delay τ ; embedding dimension m ; threshold ε

- 1 $DET \leftarrow 0.0$
- 2 $n \leftarrow N - (\tau \times (m - 1) + 1)$
- 3 $l_{min} \leftarrow 2$
- 4 **for** $i \leftarrow 0$ **to** n **do**
- 5 $D_i \leftarrow WS_i \leftarrow 0.0$
- 6 **for** $i \leftarrow 2$ **to** n **do**
- 7 $l \leftarrow 0$
- 8 $k \leftarrow (n - i) + 1$
- 9 **for** $j \leftarrow 0$ **to** $n - i$ **do**
- 10 $q \leftarrow i + j$
- 11 $\mathbf{v}_q^{m,\tau} \leftarrow (x_q, x_{q+\tau}, \dots, x_{q+(m-1)\tau})$
- 12 $\mathbf{v}_j^{m,\tau} \leftarrow (x_j, x_{j+\tau}, \dots, x_{j+(m-1)\tau})$
- 13 **if** $\|\mathbf{v}_q^{m,\tau} - \mathbf{v}_j^{m,\tau}\| \leq \varepsilon$ **then**
- 14 $l \leftarrow l + 1$
- 15 **else**
- 16 **if** $l > 0$ **then**
- 17 $D_l \leftarrow D_l + 1$
- 18 $l \leftarrow 0$
- 19 **if** $l > 0$ **then**
- 20 $D_l \leftarrow D_l + 1$
- 21 $WS_k \leftarrow k \times D_k + WS_{(k+1)}$
- 22 $DET \leftarrow WS_{(l_{min})}/WS_1$

Output: DET

Fonte: Elaborado pelo autor (2021).

As linhas 4-5 inicializam o histograma D e o vetor de soma ponderada acumulada WS . O histograma D é usado para calcular as frequências dos comprimentos das linhas diagonais, onde cada índice corresponde ao respectivo comprimento.

Os elementos WS_k correspondem à soma ponderada acumulada $WS_k \leftarrow k \times D_k + WS_{(k+1)}$ do histograma D_k em ordem decrescente do índice k da linha diagonal. Por exemplo, $WS_1 = 1 \times D_1 + WS_2$, $WS_2 = 2 \times D_1 + WS_3$, etc. Considerando a simetria da matriz de recorrência \mathbf{R} , apenas a matriz triangular inferior é utilizada, desconsiderando a *linha de identidade* (LOI) e o *movimento tangencial* (pontos consecutivos na trajetória) (MARWAN et al., 2007), representado pela linha diagonal abaixo da LOI. Para descartar a avaliação dessas duas linhas diagonais, a variável i é iniciada em 2 (índice da terceira maior linha diagonal, a LOI está no índice zero)

na linha 6. As linhas 6-18 calculam o número de recorrências em cada linha diagonal, da mais alta (próxima à LOI) até a mais baixa. As variáveis q e j definem as posições das recorrências $\mathbf{R}_{q,j}(\varepsilon)$ na diagonal de comprimento k . Caso a distância Euclidiana entre os estados $\mathbf{v}_q^{m,\tau}$ e $\mathbf{v}_j^{m,\tau}$ for menor ou igual a ε , o índice l , referente ao comprimento do segmento, é incrementado. Do contrário, o respectivo elemento l no histograma D_l é incrementado, e l é reiniciado em zero.

As linhas 19-20 evitam a perda na contagem dos segmentos diagonais que terminam no final da diagonal atual. Na linha 21, a soma cumulada ponderada de cada linha de comprimento k em cada elemento D_k é armazenada em ordem decrescente, no respectivo elemento WS_k . Finalmente, na linha 22, o *determinismo percentual* RQA_{DET} é calculado considerando l_{min} , onde apenas as linhas diagonais $l_{min} \geq 2$ foram consideradas, de acordo com Marwan et al. (2007).

4.2.1.3 Estimador da Entropia de Permutação

Em geral, os métodos que lidam com a topologia do *espaços de fase* para detectar mudanças em regimes dinâmicos são computacionalmente caros (CAO et al., 2004). Entropia de Permutação (PE) (BANDT; POMPE, 2002) usa uma abordagem simples e computacionalmente eficiente para contornar essa desvantagem. Trata-se de uma medida de complexidade para análise de séries temporais, tendo por base o mapeamento dos estados do *espaço de fase* em um conjunto de $m!$ padrões ordinais (ou símbolos) (CAO et al., 2004).

Considerando os fundamentos da Seção 3.4.2, o Algoritmo 3 apresenta o procedimento para calcular a PE. Por esse procedimento, o padrão ordinal inicial $\boldsymbol{\pi}^m = (0, 1, \dots, m-1)$ associado a cada estado será permutado em um dos $m!$ padrões ordinais distintos do conjunto $\boldsymbol{\Pi}^m = (\boldsymbol{\pi}_0^m, \boldsymbol{\pi}_1^m, \dots, \boldsymbol{\pi}_{m!-1}^m)$ (CUESTA-FRAU, 2019). Assim, quando os m componentes de cada estado $\mathbf{v}_i^{m,\tau}$ são classificados em ordem crescente, de modo que $x_{i+(\pi_{j0})\tau} \leq x_{i+(\pi_{j1})\tau} \leq \dots \leq x_{i+(\pi_{j(m-1)})\tau}$, pode-se obter a frequência de ocorrência de cada padrão do conjunto $\boldsymbol{\Pi}^m$. Após considerar todos os estados, as frequências relativas P_0, P_2, \dots, P_{k-1} são usadas na Equação 3.16.

Para descrever o procedimento, considere a série temporal $\{4, 7, 9, 10, 6, 11, 3\}$ (comprimento $N = 7$), $m = 3$, e $\tau = 1$. Portanto, existem $m! = 6$ padrões ordinais: $\boldsymbol{\Pi}_0^3 = \boldsymbol{\pi}_0^3 = (0, 1, 2)$, $\boldsymbol{\Pi}_1^3 = \boldsymbol{\pi}_1^3 = (0, 2, 1)$, $\boldsymbol{\Pi}_2^3 = \boldsymbol{\pi}_2^3 = (1, 0, 2)$, $\boldsymbol{\Pi}_3^3 = \boldsymbol{\pi}_3^3 = (1, 2, 0)$, $\boldsymbol{\Pi}_4^3 = \boldsymbol{\pi}_4^3 = (2, 0, 1)$, $\boldsymbol{\Pi}_5^3 = \boldsymbol{\pi}_5^3 = (2, 1, 0)$; e $N - \tau \times (m - 1) = 5$ estados: $\mathbf{v}_0^{3,1} = (4, 7, 9)$, $\mathbf{v}_1^{3,1} = (7, 9, 10)$,

Tabela 3 – Algoritmo para estimar a Entropia de Permutação.

Algoritmo 3: Entropia de Permutação.

Input: time series $\{x_0, x_1, x_2, \dots, x_{N-1}\}$; time delay τ ; embedding dimension m .

```

1  $h \leftarrow 0.0$ 
2  $n \leftarrow N - (\tau \times (m - 1) + 1)$ 
3  $\Pi^m \leftarrow (\pi_0^m, \pi_1^m, \dots, \pi_{m!-1}^m)$ 
4  $\pi^m \leftarrow (0, 1, \dots, m - 1)$ 
5 for  $i \leftarrow 0$  to  $m - 1$  do
6    $C_i \leftarrow P_i \leftarrow 0.0$ 
7 for  $i \leftarrow 0$  to  $n$  do
8    $\mathbf{v}_i^{m,\tau} \leftarrow (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$ 
9    $(\mathbf{w}_i^m, \pi_*^m) \leftarrow \text{sort}(\mathbf{v}_i^{m,\tau}, \pi^m)$ 
10  for  $j \leftarrow 0$  to  $m! - 1$  do
11    if  $\pi_*^m = \Pi_j^m$  then
12       $C_j \leftarrow C_j + 1$ 
13      break
14 for  $i \leftarrow 0$  to  $m! - 1$  do
15    $P_i \leftarrow C_i / (n + 1)$ 
16   if  $P_i > 0.0$  then
17      $h \leftarrow h + (-P_i \log_2 P_i) / \log_2(m!)$ 

```

Output: h

Fonte: Modificado de Cuesta-Frau (2019).

$\mathbf{v}_2^{3,1} = (9, 10, 6)$, $\mathbf{v}_3^{3,1} = (10, 6, 11)$, e $\mathbf{v}_4^{3,1} = (6, 11, 3)$, definido na linha 8. A ordenação das componentes desses estados na linha 9 resulta em: $\mathbf{w}_0^{3,1} = (4, 7, 9)$ e $\pi_*^m = (0, 1, 2)$, $\mathbf{w}_1^{3,1} = (7, 9, 10)$ e $\pi_*^m = (0, 1, 2)$, $\mathbf{w}_2^{3,1} = (6, 9, 10)$ e $\pi_*^m = (2, 0, 1)$, $\mathbf{w}_3^{3,1} = (6, 10, 11)$ e $\pi_*^m = (1, 0, 2)$, e $\mathbf{w}_4^{3,1} = (3, 6, 11)$ com $\pi_*^m = (2, 0, 1)$, em que o padrão ordinal inicial é $\pi^m = (0, 1, 2)$.

As linhas 10-12 calculam o número de ocorrências de cada padrão ordinal, a partir do qual os padrões Π_0^3 e Π_4^3 ocorrem duas vezes ($C_0 = 2$ e $C_4 = 2$), e o padrão Π_2^3 ocorre uma única vez ($C_2 = 1$). Assim, a linha 15 calcula as frequências relativas $P_0 = 0,4$, $P_1 = 0,0$, $P_2 = 0,2$, $P_3 = 0,0$, $P_4 = 0,4$ e $P_5 = 0,0$. Finalmente, a partir da Equação 3.16, as linhas 16-17 definem $H(6) = -(2 \times 0,4 \log_2(0,4) + 0,2 \log_2(0,2)) / \log_2(6) = 0,588$.

4.2.2 Heurística de detecção

A SE é comumente utilizada no monitoramento de transições entre o estado consciente e o inconsciente em avaliações da atividade elétrica cerebral (SLEIGH et al., 2004). De modo

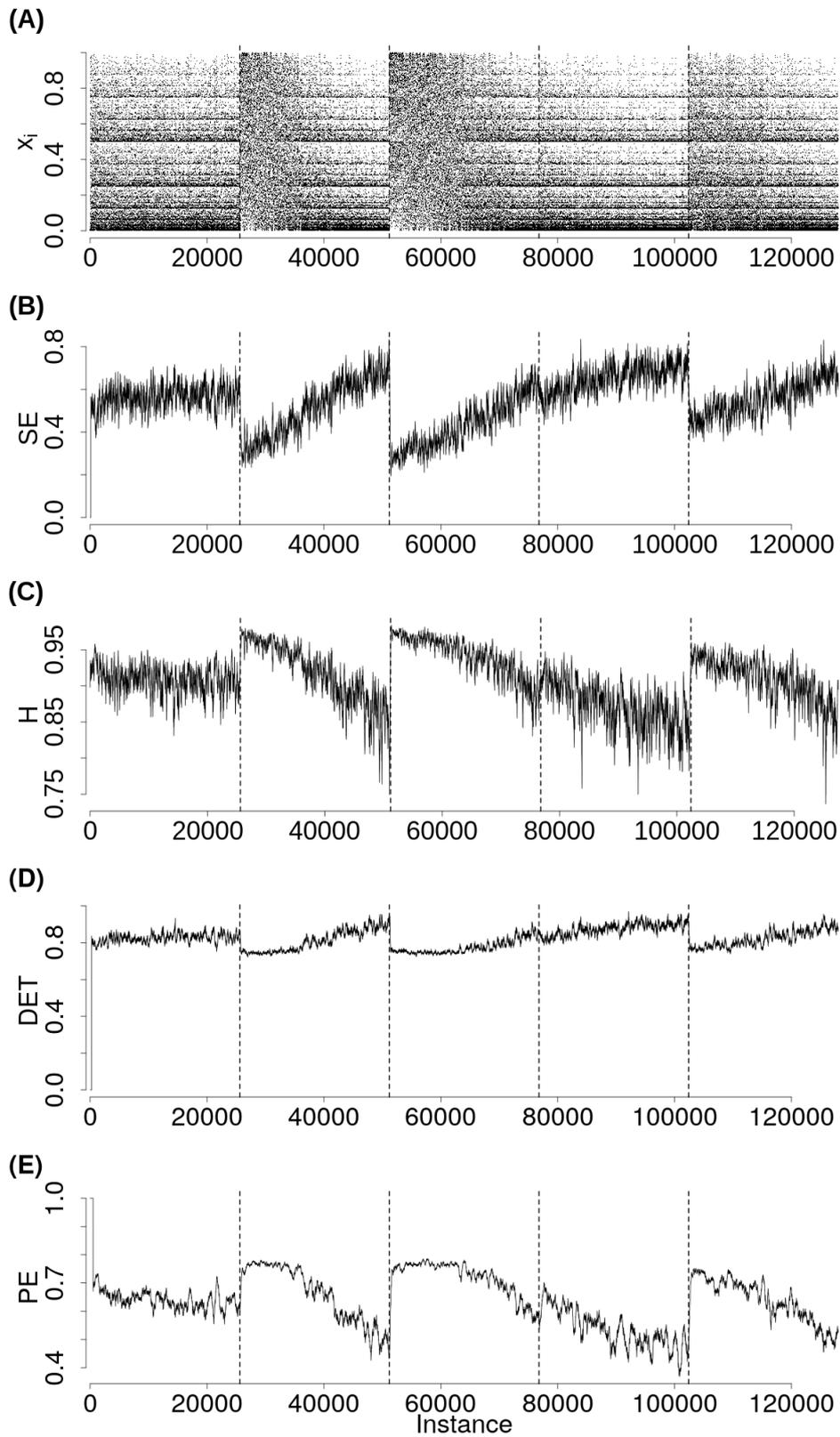
semelhante, a PE e a RQA são largamente usadas para quantificar a complexidade de séries temporais em diferentes campos de investigação (MARWAN et al., 2007; AMIGÓ, 2010). Nesta pesquisa, essas medidas são utilizadas para monitorar as transições entre os padrões obtidos do Mapa de Bernoulli ao longo do fluxo de erro, a fim de identificar *mudanças de conceito*. Assim, o Algoritmo 1 fundamenta o Spectral Entropy Drift Detector (SEDD), o Algoritmo 2 fundamenta o Recurrence Quantification Analysis Drift Detector (RQADD) com base no *determinismo percentual* RQA_{DET} , e de modo semelhante, o Algoritmo 3 fundamenta o Permutation Entropy Drift Detector (PEDD).

Ao computar essas medidas no retorno do Mapa de Bernoulli, verifica-se empiricamente que SE e o *determinismo percentual* RQA_{DET} flutuam em torno de uma média ou mesmo seguem uma tendência ascendente durante o processo de aprendizagem, enquanto diminui repentinamente durante a *mudança de conceito*. Por outro lado, a PE e a entropia de Shannon apresentam um comportamento inverso.

Para ilustrar o comportamento dessas medidas ao longo do processo preditivo, a Figura 23 (A) mostra o gráfico de uma série temporal obtida de um fluxo de erros por meio do Mapa de Bernoulli, onde quatro *mudanças de conceito* são sinalizadas com linhas tracejadas. Nota-se que o mapeamento destaca os padrões de transição sempre que ocorrem mudanças no *fluxo de dados*, exceto na terceira mudança, talvez porque a distribuição $P(\mathcal{Y}|\mathcal{X})$ não tenha mudado significativamente. Assim, esses padrões sugerem que as *mudanças de conceito* podem estar associadas a transições entre regimes do Mapa de Bernoulli. Mais especificamente, órbitas densas ou mesmo caóticas parecem estar relacionadas a mudanças em $P(\mathcal{Y}|\mathcal{X})$.

A Figura 23 (B) apresenta a flutuação da SE ao longo do processo de aprendizagem em relação aos padrões de transição. Nota-se que SE computa altos níveis de entropia para padrões regulares e limpos, e baixos níveis de entropia para padrões complexos e ruidosos. Esse comportamento é devido a SE usar apenas as informações de magnitude A_ξ do PS. Esse comportamento sugere que durante mudanças em $P(\mathcal{Y}|\mathcal{X})$, o processo de aprendizagem apresenta mais distúrbios nos componentes de fase ϕ_ξ do que nas magnitudes A_ξ . Essa perturbação é melhor avaliada por meio da entropia de Shannon, apresentada na Figura 23 (C), a qual se comporta de maneira oposta. Além disso, essa entropia apresenta variância diretamente proporcionais aos níveis da SE. A Figura 23 (D) apresenta o comportamento da entropia ao longo desta série temporal, onde PE flutua em torno de um platô ou mesmo decrescente durante o processo de aprendizagem, enquanto aumenta repentinamente durante os desvios de conceito. Finalmente, a Figura 23 (E) mostra o comportamento do RQA_{DET} , que flutua de forma

Figura 23 – (A) Série temporal do fluxo de erros obtida por meio do Mapa de Bernoulli; (B) Entropia Espectral; (C) Entropia de Shannon; (D) Determinismo; e (E) Entropia de Permutação. Linhas tracejadas marcam as *mudanças de conceito* do fluxo de dados.



Fonte: Elaborada pelo autor (2021).

semelhante à SE.

Com base nesses comportamentos, o Algoritmo 4 apresenta um fluxo abstrato de execução para lidar com a *mudança de conceito*. Observe que as linhas 6-8 realizam a aplicação do Mapa de Bernoulli após cada avaliação da predição corrente, expressa na linha 5. Na linha 6, uma operação de deslocamento de bit à esquerda ($\ll 1$) é sempre realizada na unidade de memória w antes que uma nova avaliação de erro b_i (0 se correto ou 1 caso contrário) seja adicionada à w na linha 7. Nessa caso, w pode ser interpretado como uma janela deslizante em que a avaliação mais recente é sempre armazenada no bit menos significativo (LSB), enquanto as avaliações anteriores são empurradas para o bit mais significativo (MSB). Depois disso, a linha 8 converte cada valor inteiro de w em um valor decimal em x_i de acordo com a Equação (4.1).

Tabela 4 – Algoritmo para identificar a mudança de conceito.

Algoritmo 4: Drift Detector.

Input: magnitude δ ; binary digit expansion length L ; window size N .

```

1  $upperBound \leftarrow lowerBound \leftarrow 0.0$ 
2  $w \leftarrow 0$ 
3  $s \leftarrow \{\}$ 
4 foreach instance in stream do
5    $b_i \leftarrow \text{prediction}(\textit{instance})$ 
6    $w \leftarrow w \ll 1$ 
7    $w \leftarrow w + b_i$ 
8    $x_i \leftarrow w/2^L$ 
9   enqueue  $x_i$  in  $s$ 
10  if  $\text{size}(s) = N$  then
11     $h_i \leftarrow \text{measure}(s)$ 
12    if  $h_i > 0.0$  and  $h_i < 1.0$  then
13      if  $h_i > upperBound$  then
14         $upperBound \leftarrow h_i$ 
15         $lowerBound \leftarrow upperBound$ 
16      if  $h_i < lowerBound$  then
17         $lowerBound \leftarrow h_i$ 
18    if  $upperBound - lowerBound \geq \delta$  then
19       $\text{alarmDrift}()$ 
20    drop the oldest element  $x_{i-N+1}$  from  $s$ 
```

Fonte: Elaborado pelo autor (2021).

Nas linhas 9-11, o valor x_i é inserido em uma fila (s) e um dos algoritmos descritos anteriormente (SEDD, PEDD, ou RQADD) é chamado na linha 11 (**measure(s)**) sempre que s contém dados suficientes. Observe que, na linha 20, o valor $x_{(i-N)+1}$ mais antigo será retirado

de s para que um novo valor de entrada x_{i+1} possa ser inserido e processado. As linhas 12-17 avaliam a flutuação da medida utilizada, armazenando os níveis superior e inferior nas variáveis *upperBound* e *lowerBound*, respectivamente. Uma mudança relevante ocorre quando a magnitude da diferença entre o *upperBound* e *lowerBound*, calculados anteriormente, é maior que um limiar δ . Devido a possíveis tendências ascendentes, *upperBound* é atualizado e *lowerBound* é redefinido sempre que um novo valor h_i for maior que *upperBound*. Além disso, os valores extremos (ou seja, 0,0 e 1,0) são desconsiderados para garantir atualizações contínuas das variáveis *upperBound* e *lowerBound*. Finalmente, as linhas 18-19 calculam a magnitude entre os níveis superior e inferior e disparam um alarme sempre que $upperBound - lowerBound \geq \delta$.

Por padrão, o Algoritmo 4 está relacionado com as medidas obtidas do SEDD e do RQADD, visto que a tendência da flutuação tratada na linha 13 do Algoritmo 4 está de acordo com o comportamento da SE e do *determinismo percentual RQA_{DET}*, como visto na 23 (B) e 23 (D), respectivamente. Entretanto, ao substituir **measure(s)**, na linha 11, por $1,0 - \mathbf{measure(s)}$, a condição da avaliação possibilita o uso do PEDD.

Ao contrário dos algoritmos citados na Seção 2.2.1, o SEDD, o PEDD, e o RQADD não fazem nenhuma suposição sobre a independência do fluxo binário de erros, ou seja, não está restrito ao pressuposto de dados i.i.d. É importante enfatizar que, quando há dependência temporal, as garantias teóricas de todos os algoritmos levantados na Seção 2.2.1 são violadas (ŽLIOBAITĖ et al., 2015). Por outro lado, o SEDD, o PEDD, e o RQADD levam em conta a dinâmica do processo de aprendizagem, e os alarmes são emitidos sempre que ocorrem transições de regimes do Mapa de Bernoulli, mais precisamente, transições entre regimes de baixa para alta complexidade. Nesse caso, a dependência temporal não invalida essas abordagens fundamentadas na NLTA. No entanto, é possível que o viés do classificador usado para estimar $P(\mathcal{Y}|\mathcal{X})$ possa afetar o desempenho da detecção.

4.3 PSEUDO-ROTULAÇÃO SIMBÓLICA

As abordagens supervisionadas aplicadas a *fluxos de dados* requerem instâncias rotuladas que podem ser difíceis de obter em muitas aplicações do mundo real. Entretanto, de acordo com Pagliosa e Mello (2017), modelos de classificação ou regressão podem ser induzidos ao considerar a reconstrução do *espaço de fase* para mapear dados com dependências temporais em instâncias i.i.d.. Os autores defendem que a partir de uma reconstrução apropriada do

espaço de fase, os estados obtidos representam as dependências temporais intrínsecas das observações, os quais podem ser amostrados de forma i.i.d..

Para ilustrar esse mapeamento, considere um modelo de regressão f em que cada retorno $x_{i+1} \in \mathcal{Y}$ depende de uma entrada $x_i \in \mathcal{X}$. Com base na reconstrução do *espaço de fase*, os estados $\mathbf{v}_i^{m,\tau} = (x_i, x_{i+1})$, $\mathbf{v}_{i+1}^{m,\tau} = (x_{i+1}, x_{i+2})$, $\mathbf{v}_{i+2}^{m,\tau} = (x_{i+2}, x_{i+3})$, \dots , podem ser uniformemente amostrados para compor um conjunto de treinamento, onde $m = 2$, e $\tau = 1$. Portanto, dado que a distribuição conjunta $P(\mathcal{X} \times \mathcal{Y})$ permaneça estacionária, as condições da SLT são atendidas. A Tabela (5) descreve o resultado desse mapeamento de acordo com Pagliosa e Mello (2017). Nas colunas \mathcal{X} , \mathcal{Y} , e $\mathbf{v}^{2,1}$, estão, respectivamente, as observações de entrada x_i , os retornos x_{i+1} e os estados. A partir dessa configuração, o mapeamento permite prever recursivamente as próximas observações com base em um horizonte de previsão predefinido (PAGLIOSA; MELLO, 2017).

Tabela 5 – Exemplo do mapeamento de uma série temporal em instâncias i.i.d. (coluna $\mathbf{v}^{2,1}$) obtidas após a reconstrução do *espaço de fase*, considerando os parâmetros $m = 2$, e $\tau = 1$.

\mathcal{X}	\mathcal{Y}	$\mathbf{v}^{2,1}$
\vdots	\vdots	\vdots
x_i	x_{i+1}	(x_i, x_{i+1})
x_{i+1}	x_{i+2}	(x_{i+1}, x_{i+2})
x_{i+2}	x_{i+3}	(x_{i+2}, x_{i+3})
\vdots	\vdots	\vdots

Fonte: Adaptada de Pagliosa e Mello (2017).

A previsão de estados em um *espaço de fase* reporta ao *Method of Analogues* de Lorenz (1969) (SCHREIBER, 1999; BRADLEY; KANTZ, 2015). A ideia é que, dado a relação $\mathbf{v}_{i+k}^{m,\tau} = \mathbf{f}^k(\mathbf{v}_i^{m,\tau})$, estados semelhantes evoluam para estados futuros também semelhantes devido ao determinismo (SCHREIBER, 1999), em um horizonte de k iterações. Com base nessa relação, a Tabela 6 apresenta uma generalização do mapeamento proposto em Pagliosa e Mello (2017), ilustrado na Tabela 5.

Na Tabela 6, os dados de entrada (\mathcal{X}) e retorno (\mathcal{Y}) são estados ($\mathbf{v}^{m,\tau}$) de um *espaço de fase* reconstruído no \mathbb{R}^m , enquanto as instâncias ($\mathbf{v}^{2m,\tau}$) são definidas em uma reconstrução no \mathbb{R}^{2m} . Note que a concatenação dos componentes dos estados $\mathbf{v}^{m,\tau}$ da entrada (\mathcal{X}) e do retorno (\mathcal{Y}) constituem as instâncias ($\mathbf{v}^{2m,\tau}$). Nessa configuração, a amostragem por meio da relação $\mathbf{v}_{i+m\tau}^{m,\tau} = \mathbf{f}^{(m\tau)}(\mathbf{v}_i^{m,\tau})$ entre os estados de entrada e de retorno deve ser definida em $m\tau$ iterações, de modo a evitar que esses estados ($\mathbf{v}^{m,\tau}$) apresentem sobreposições entre as suas

componentes, e garantir uma composição coerente de concatenação de modo a constituir as instâncias ($\mathbf{v}^{2m,\tau}$). Nessas condições, essas instâncias ($\mathbf{v}^{2m,\tau}$) podem ser amostradas de modo independente de acordo com o mapeamento sugerido por Pagliosa e Mello (2017).

Vale salientar que o horizonte de iterações definido por $f^{(m\tau)}(\cdot)$ pode enviesar a amostra dos estados usados para o processo indutivo. Considerando uma dimensão de reconstrução (m) fixa, um retardo temporal (τ) pequeno tende a favorecer uma amostra de estados propensos a induzir *overfitting* devido a uma forte dependência temporal na relação $\mathbf{v}_{i+m\tau}^{m,\tau} = \mathbf{f}^{(m\tau)}(\mathbf{v}_i^{m,\tau})$. Por outro lado, um retardo temporal (τ) muito grande pode favorecer a seleção de uma amostra de estados propensos a induzir *underfitting* devido a ausência de uma relação significativa entre o estado de entrada ($\mathbf{v}_i^{m,\tau} \in \mathcal{X}$) e de retorno ($\mathbf{v}_{i+m\tau}^{m,\tau} \in \mathcal{Y}$). Além disso, a dimensão (m) está diretamente relacionada com o número de classes ($m!$).

Tabela 6 – Exemplo da generalização do mapeamento, em que o conjunto de dados de entrada (\mathcal{X}) e retorno (\mathcal{Y}) estão definidos em uma dimensão m ($\mathcal{X}, \mathcal{Y} \in \mathbb{R}^m$), e as instâncias resultantes, na dimensão $2m$ ($\mathcal{X} \frown \mathcal{Y} \in \mathbb{R}^{2m}$).

$\mathcal{X} : \{\mathbf{v}^{m,\tau}\}$	$\mathcal{Y} : \{\mathbf{v}^{m,\tau}\}$	$\mathcal{X} \frown \mathcal{Y} : \{\mathbf{v}^{2m,\tau}\}$
⋮	⋮	⋮
$(x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})$	$(x_{i+m\tau}, x_{i+(m+1)\tau}, \dots, x_{i+(2m-1)\tau})$	$(x_i, \dots, x_{i+(m-1)\tau}, x_{i+m\tau}, \dots, x_{i+(2m-1)\tau})$
$(x_{i+m\tau}, x_{i+(m+1)\tau}, \dots, x_{i+(2m-1)\tau})$	$(x_{i+2m\tau}, x_{i+(2m+1)\tau}, \dots, x_{i+(3m-1)\tau})$	$(x_{i+m\tau}, \dots, x_{i+(2m-1)\tau}, x_{i+2m\tau}, \dots, x_{i+(3m-1)\tau})$
$(x_{i+2m\tau}, x_{i+(2m+1)\tau}, \dots, x_{i+(3m-1)\tau})$	$(x_{i+3m\tau}, x_{i+(3m+1)\tau}, \dots, x_{i+(4m-1)\tau})$	$(x_{i+2m\tau}, \dots, x_{i+(3m-1)\tau}, x_{i+3m\tau}, \dots, x_{i+(4m-1)\tau})$
⋮	⋮	⋮

Fonte: Elaborada pelo autor (2021).

O procedimento descrito nas Seções (3.4.2) e (4.2.1.3) para agrupar as observações de uma série temporal em um conjunto de símbolos considera possíveis dependências temporais, e preserva a correspondência entre a dinâmica real e a simbólica (AMIGÓ, 2010). Com base na relação intrínseca entre essas dinâmicas, pode-se descrever a generalização do mapeamento apresentado na Tabela 6 por meio de padrões ordinais, como ilustrado na Tabela 7. Nessa tabela, as relações entre os estados de entrada (coluna \mathcal{X}) com os estados de retorno (coluna \mathcal{Y}), estão acompanhadas com os respectivos símbolos na coluna \mathcal{W} , representando os padrões ordinais obtidos dos estados de retorno. Nesse caso, cada instância é construída por meio da associação de um estado de entrada e o símbolo relacionado ao estado de retorno, de acordo com a coluna $\mathcal{X} \frown \mathcal{W}$ da Tabela 7.

Como apontado em Pagliosa e Mello (2017), de acordo com o *teorema de Bayes*, cada atributo de entrada em \mathcal{X} deve ser significativo para estimar a classe de saída em \mathcal{Y} . Portanto, os atributos de entrada devem apresentar algum nível de dependência do resultado esperado

Tabela 7 – Exemplo da pseudo-rotulação simbólica, em que o conjunto de símbolos de retorno (\mathcal{W}) estão definidos na dimensão 1 ($\mathcal{W} \in \mathbb{N}$), e são obtidos a partir do conjunto \mathcal{Y} . Nesse caso, as instâncias resultantes estão definidas na dimensão $m + 1$ ($\mathcal{X} \cap \mathcal{W} \in \mathbb{R}^{m+1}$).

\mathcal{X}	\mathcal{Y}	\mathcal{W}	$\mathcal{X} \cap \mathcal{W}$
\vdots	\vdots	\vdots	\vdots
$\mathbf{v}_i^{m,\tau}$	$\mathbf{v}_{(i+m\tau)}^{m,\tau}$	$\pi_{(i+m\tau)}^m$	$(\mathbf{v}_i^{m,\tau}, \pi_{(i+m\tau)}^m)$
$\mathbf{v}_{(i+m\tau)}^{m,\tau}$	$\mathbf{v}_{(i+2m\tau)}^{m,\tau}$	$\pi_{(i+2m\tau)}^m$	$(\mathbf{v}_{(i+m\tau)}^{m,\tau}, \pi_{(i+2m\tau)}^m)$
$\mathbf{v}_{(i+2m\tau)}^{m,\tau}$	$\mathbf{v}_{(i+3m\tau)}^{m,\tau}$	$\pi_{(i+3m\tau)}^m$	$(\mathbf{v}_{(i+2m\tau)}^{m,\tau}, \pi_{(i+3m\tau)}^m)$
\vdots	\vdots	\vdots	\vdots

Fonte: Elaborada pelo autor (2021).

Tabela 8 – Algoritmo para gerar a pseudo-rotulação simbólica.

Algoritmo 5: Symbolic Labeling Adapter.

Input: time delay τ ; embedding dimension m ; base learner \mathcal{L} ; drift detector \mathcal{D} .

```

1  $n \leftarrow \tau \times m$ 
2  $\pi^m \leftarrow (0, 1, \dots, m - 1)$ 
3  $\Pi^m \leftarrow (\pi_0^m, \pi_1^m, \dots, \pi_{m-1}^m)$ 
4  $s \leftarrow \{\}$ 
5 foreach  $x_i$  in stream do
6   enqueue  $x_i$  in  $s$ 
7   if  $\text{size}(s) = n$  then
8      $\mathbf{v}_i^{m,\tau} \leftarrow (s_0, s_\tau, s_{2\tau}, \dots, s_{(m-1)\tau})$ 
9      $(\mathbf{w}_i^m, \pi_*^m) \leftarrow \text{sort}(\mathbf{v}_i^{m,\tau}, \pi^m)$ 
10    if  $p_{(i-1)} = \pi_*^m$  then
11       $\text{error} \leftarrow 0$ 
12    else
13       $\text{error} \leftarrow 1$ 
14     $\text{drift} \leftarrow \text{driftDetection}(\mathcal{D}, \text{error})$ 
15    if  $\text{drift} = \text{TRUE}$  then
16      replace( $\mathcal{L}$ )
17       $\text{drift} \leftarrow \text{FALSE}$ 
18     $p_i \leftarrow \text{prediction}(\mathcal{L}, \mathbf{v}_i^{m,\tau}, \Pi^m)$ 
19    train( $\mathcal{L}, (\mathbf{v}_{(i-1)}^{m,\tau}, \pi_*^m)$ )
20     $s \leftarrow \{\}$ 

```

Fonte: Elaborado pelo autor (2021).

(BISHOP, 2006). Assim, como em cada instância $(\mathbf{v}_i^{m,\tau}, \pi_{(i+m\tau)}^m)$ o estado de entrada $(\mathbf{v}_i^{m,\tau})$ tem algum nível de dependência em relação ao seu respectivo símbolo $(\pi_{(i+m\tau)}^m)$, esse estado pode ser interpretado como um vetor de atributos, enquanto o símbolo representa o respectivo rótulo. Logo, pode-se induzir um classificador ao considerar a configuração das instâncias da

Tabela 7. Conseqüentemente, essa pseudo-rotulação simbólica viabiliza uma amostragem para fornecer instâncias rotuladas relacionadas à dinâmica subjacente do *fluxo de dados*, com o objetivo de detectar *mudanças de conceito*.

O Algoritmo 5, denominado Symbolic Labeling Adapter (SLA), fornece detalhes do procedimento para rotular automaticamente uma série temporal com a finalidade de detectar *mudanças de conceito* utilizando modelos adaptativos supervisionados aplicados a *fluxo de dados*. As linhas 6-7 insere cada observação x_i na fila s , e desdobra um estado $\mathbf{v}_i^{m,\tau}$ sempre que s contém dados suficientes, considerando um intervalo de $m \times \tau$ observações. Semelhante ao Algoritmo 4, a linha 20 limpa a fila s , de modo que novas observações possam ser inseridas e processadas.

Após a ordenação dos elementos de cada estado $\mathbf{v}_i^{m,\tau}$ na linha 9, define-se o respectivo símbolo (padrão ordinal) π_*^m , o qual será usado como rótulo. As linhas 10-13 retornam o erro após a avaliação da predição anterior \mathbf{p}_{i-1} com relação ao rótulo atual (símbolo π_*^m). As linhas 14-17 trata as detecções de *mudança de conceito*, onde o fluxo de erros das predições é avaliado pelo detector \mathcal{D} na linha 14. A linha 16 troca o classificador \mathcal{L} sempre que um alarme é emitido na linha 15 e redefine a variável *drift* na linha 17. A linha 18 realiza as predições usando o classificador \mathcal{L} e o estado atual $\mathbf{v}_i^{m,\tau}$, de modo que \mathbf{p}_i representa um dos padrões ordinais do conjunto Π^m . Finalmente, a linha 19 treina o classificador \mathcal{L} usando a instância $(\mathbf{v}_{(i-1)}^{m,\tau}, \pi_*^m)$. Observe que em cada instância os atributos são os elementos do estado anterior $\mathbf{v}_{(i-1)}^{m,\tau}$, e o rótulo, o símbolo atual π_*^m .

4.4 RESUMO DO CAPÍTULO

Neste capítulo, foram explicados os métodos para a detecção de *mudanças de conceito* fundamentadas na NLTA, bem como a estratégia para a pseudo-rotulação simbólica por meio da reconstrução do espaço de fase. Esses métodos constituem formas alternativas para avaliar o comportamento de *fluxo de dados*, as quais podem revelar informações não acessíveis quando apenas as distribuições dos dados são consideradas. Com base no mapeamento do fluxo de erros das predições em uma série temporal escalar, três detectores foram desenvolvidos: Spectral Entropy Drift Detector (SEDD), Permutation Entropy Drift Detector (PEDD) e Recurrence Quantification Analysis Drift Detector (RQADD). A estratégia de pseudo-rotulação simbólica, o Symbolic Labeling Adapter (SLA), viabiliza amostragens de estados de modo a fornecer instâncias rotuladas relacionadas à dinâmica subjacente do *fluxo de dados*, com o objetivo de

detectar *mudanças de conceito*. Essa rotulação busca contornar o problema da disponibilidade de instâncias rotuladas necessárias para o uso de modelos adaptativos supervisionados em muitas aplicações do mundo real.

5 ANÁLISE EMPÍRICA E ESTATÍSTICA

Este capítulo apresenta os resultados e as respectivas análises referentes aos experimentos usados na avaliação dos métodos propostos no Capítulo 4. As seções seguintes descrevem as bases de dados, os experimentos e suas configurações, os testes estatísticos, e a metodologia utilizada para avaliar os resultados.

5.1 BASES DE DADOS

Dois grupos de experimentos foram usados para a avaliação dos métodos propostos contra um grande número de detectores.

O primeiro grupo abrange as bases artificiais, e tem por finalidade avaliar aspectos relevantes em termos da acurácia e da qualidade da detecção de *mudança de conceito* em um ambiente controlado, envolvendo diversos cenários para avaliar a resiliência dos métodos em relação à distribuição desbalanceada de classes, níveis de ruídos, problemas multiclases, e dependência temporal. Também é investigado a viabilidade da utilização do Symbolic Labeling Adapter (SLA) para identificar mudanças em fluxos de dados não rotulados com o auxílio do Sistema de Lorenz Transiente.

O segundo grupo de experimentos busca generalizar a aplicação dos métodos propostos a cenários do mundo real. Neste sentido, foram utilizados um conjunto de 20 bases reais largamente utilizadas na literatura, e disponíveis livremente em diversos repositórios. Esse conjunto de bases apresenta diversas propriedades ligadas a problemas do mundo real. Um ponto a destacar está na impossibilidade de saber quando uma *mudança de conceito* vai ocorrer. Nesta mesma linha, também foi usada uma base de EEG, configurando mais um cenário apropriado para a avaliação do Symbolic Labeling Adapter (SLA).

5.1.1 Bases Artificiais

Os conjuntos de dados sintéticos fornecem um cenário controlado, o que permitem identificar os pontos exatos de *mudança de conceito*. Assim, experimentos foram realizados utilizando três geradores de bases artificiais com o objetivo de avaliar tanto o desempenho da classificação quanto as detecções, dado que os pontos de mudanças são conhecidos previamente. Vale

salientar que geradores de bases artificiais são largamente utilizados na literatura (ANDERSON et al., 2019; BARROS; SANTOS, 2018; DUONG; RAMAMPIARO; NØRVÅG, 2018; PESARANGHADER; VIKTOR; PAQUET, 2018; SEBASTIÃO; GAMA; MENDONÇA, 2017) e estão disponíveis livremente no Massive Online Analysis Framework (MOA) (BIFET et al., 2010).

5.1.1.1 Geradores de bases artificiais

Cada gerador produz conjuntos de dados com 256.000 instâncias, considerando *mudanças de conceito* abruptas e graduais. Quatro pontos de mudanças foram distribuídos em intervalos regulares a cada 51.200 instâncias (5 conceitos). As mudanças graduais foram configuradas com um nível de transição proporcional a 5% (2.560 instâncias) do tamanho do conceito. Nessas regiões de transição, as instâncias geradas têm cada vez mais probabilidade de pertencerem ao conceito mais recente.

Os geradores de conjunto de dados selecionados são:

- Agrawal (AGRAWAL; IMIELINSKI; SWAMI, 1993) armazena informações sobre salário, comissão, idade, escolaridade, CEP, valor da casa, etc., de pessoas classificadas em dois grupos (A ou B) com o objetivo de receber um empréstimo. Os autores propuseram dez funções configuradas com diferentes formas de avaliação e com possibilidade de adição de ruído. As últimas cinco funções (F6 a F10) foram utilizadas para gerar os respectivos cinco conceitos dos conjuntos de dados com 9 atributos e 2 classes. Duas configurações foram utilizadas, considerando classes balanceadas ($Agrawal_B$) e desbalanceadas ($Agrawal_D$). Os outros parâmetros foram deixados nas configurações padrão.
- LED (FRÍAS-BLANCO et al., 2015) representa os dígitos de um display LED de sete segmentos, sendo composto por 24 atributos categóricos (17 deles irrelevantes) e 10 classes (os respectivos dígitos). Cada atributo tem uma probabilidade de 10% de ser invertido (ruído), e as *mudanças de conceitos* são simuladas alterando a posição dos sete atributos relevantes. Duas configurações também foram usadas para avaliar o efeito do ruído. As configurações, referidas como $LED_{5\%}$ e $LED_{10\%}$, usam 5% e 10% de ruído, respectivamente. O número de atributos com *drift* muda de 1 para 5 ao longo dos conceitos, e os parâmetros restantes também foram deixados com o valor padrão.
- RandomRBF (HIDALGO; MACIEL; BARROS, 2019) usa uma distribuição Gaussiana para

determinar os valores de m atributos e n centroides para definir os rótulos das classes. Cada centroide também determina uma hipersfera normalmente distribuída com densidades variadas de exemplos. As *mudanças de conceito* são simuladas alterando as posições dos centroides. Os conjuntos de dados foram gerados com 5 atributos e 10 centroides no modelo, considerando a velocidade de mudança $s = 0,01$. Tanto o número de centroides com *mudanças de conceito* quanto a semente para a geração aleatória do modelo mudam de 1 a 5 ao longo dos conceitos. As configurações utilizadas consideram tarefa de classificação com duas e cinco classes, e são denominadas RBF_2 e RBF_5 , respectivamente.

Para cada gerador foram obtidas 30 bases artificiais, e 5 conceitos com as configurações de mudanças abruptas e graduais. Desse modo, esse conjunto de geradores possibilitou a criação de 360 fluxos de dados distintos, considerando classes balanceadas e desbalanceadas, diferentes níveis de ruídos, problemas com 2 e 5 classes. Embora esses geradores sejam projetados para produzir instâncias sem dependências temporais, eles foram utilizados nesta pesquisa para viabilizar a comparação dos detectores propostos em cenários já utilizados com a maioria dos detectores fornecidos na literatura.

5.1.1.2 Sistema de Lorenz Transiente

De modo semelhante às bases artificiais criadas com os geradores, o sistema transiente de Lorenz, descrito na Seção 3.2.2, possibilita a construção de fluxos de dados com pontos de mudanças em sua dinâmica. Esses pontos são identificados por meio do maior expoente de Lyapunov (λ_1), usado como medida de complexidade relacionada ao comportamento *caótico*. Nesta pesquisa, os pontos de mudanças na dinâmica dos *fluxos de dados* gerados pelo Sistema de Lorenz Transiente foram definidos seguindo outros experimentos semelhantes da literatura (TRULLA et al., 1996; CAO et al., 2004; COSTA et al., 2017). Assim, fluxos de dados foram gerados por meio da variação do estado inicial $\mathbf{v}_0 = (k, k, k)$, onde k foi definido no intervalo dos inteiros [5,35].

Como variações do estado inicial desse sistema caótico geram orbitas completamente distintas, foram gerados 30 fluxos de dados considerando a variação do parâmetro $\rho(t)$ de 28,0 até 268,0 com incrementos de 0,0002. Com essa configuração, foram obtidos fluxos de dados com 1.200.000 observações da variável $x(t)$ do sistema. Nesse caso, como as transições entre

os regimes ocorrem de forma gradual, os pontos de mudança da dinâmica nos 30 fluxos de dados foram considerados de modo aproximado nas observações 360.000, 585.000, 690.000, e 932.000. Esses fluxos de dados foram utilizados na avaliação do procedimento de rotulação simbólica descrito na Seção 4.3.

5.1.2 Bases do mundo real

Com o intuito de suprir requisitos para aplicações mais realistas, conjuntos de dados do mundo real foram selecionados para complementar a avaliação dos métodos propostos. As bases reais também abrangem experimentos com bases rotuladas e não-rotuladas.

5.1.2.1 Bases reais rotuladas

A Tabela 9 apresenta um resumo das características de todas as bases reais utilizadas nos experimentos. Vale salientar que essas bases são utilizadas para avaliar os detectores num cenário onde não é possível identificar previamente os pontos de *mudança de conceito*. Por estarem relacionadas a contextos do mundo real, apresentam diferentes graus de dependência temporal. Desse modo, a utilização do *Classificador Temporalmente Aumentado* como um referencial faz-se necessária para uma comparação criteriosa dos detectores avaliados. Além disso, foram selecionadas por estarem disponíveis livremente nos repositório apresentados abaixo, e serão descritas a seguir.

- CSD (<http://mlkd.csd.auth.gr/concept_drift.html>);
- Github (<<https://github.com/vlosing/driftDatasets/>>);
- Kaggle (<<https://www.kaggle.com/>>);
- OpenML (<<https://www.openml.org/d/>>);
- UCI (<<https://archive.ics.uci.edu/ml/datasets/>>); e
- Waikato (<<https://moa.cms.waikato.ac.nz/datasets/>>).

Bank Marketing (MORO; CORTEZ; RITA, 2014): essa base possui 45.211 instâncias com 17 atributos de campanhas de marketing de uma instituição bancária portuguesa. O objetivo da classificação é prever se um cliente fará um depósito a prazo.

Tabela 9 – Características das bases de dados reais utilizadas nos experimentos com seus respectivos repositórios.

Base de Dados	#instâncias	#Atributos	#Classes	Repositório
Bank Marketing	45.211	17	2	UCI
Census-Income (KDD)	299.285	40	2	UCI
Chess (King-Rook vs. King)	28.056	7	17	UCI
Connect-4 Data Set	67.557	42	3	UCI
Forest Covertypes	581.012	54	7	Waikato
EEG Eye State	14.980	15	2	UCI
Electricity	45.312	9	2	Waikato
Give me Some Credit	150.000	12	2	Kaggle
Letter Recognition	20.000	17	26	UCI
Nomao Challenge	34.465	120	2	UCI
Outdoor Objects	4.000	21	40	Github
Poker-Hand	829.201	11	10	Waikato
Santander Customer Satisfaction	76.020	371	2	Kaggle
Sick	3.772	30	2	OpenML
Spam Data	9.324	500	2	CSD
Usenet 1 e 2	3.000	99	2	CSD
Weather	18.159	9	2	Github

Fonte: Elaborada pelo autor (2021).

Census-Income (KDD) (OZA; RUSSELL, 2001): contém dados do censo ponderado extraídos das pesquisas populacionais de 1994 e 1995 conduzidas pelo U.S. Census Bureau. A base possui 299285 instâncias com 40 atributos, e o atributo alvo refere-se à renda total do indivíduo.

Chess (King-Rook vs. King) (KOHAVI, 1997): dados de final de uma partida de xadrez reduzido a três peças: um rei e uma torre para um jogador, e apenas o rei do outro. O conjunto de dados consiste de 28.056 instâncias e 7 atributos, em que cada instância representa as posições das jogadas no tabuleiro. O atributo alvo é o número mínimo de jogadas para as brancas vencerem (0 a 16 jogadas), ou empatarem. Também foi utilizada uma versão com as instâncias embaralhadas.

Connect-4 (BARROS; HIDALGO; CABRAL, 2018): é composto por 42 atributos e 67.557 instâncias. Ele contém todas as posições legais de 8 camadas no jogo de conectar-4 em que nenhum dos jogadores ganhou ainda e em que o próximo movimento não é forçado. A classe de resultado é o valor teórico do jogo para o primeiro jogador: vitória, derrota ou empate.

Forest Covertypes (versão normalizada) (ŽLIOBAITĖ et al., 2015) : contém células de informação que correspondem a uma cobertura florestal de 30x30 metros determinadas pelo Sistema de Informações de recursos da Região 2 do Serviço Florestal dos Estados Unidos. Essa base possui 581.012 instâncias, 7 classes e 54 atributos, e o objetivo é prever o tipo de cobertura florestal a partir das variáveis cartográficas. Também foi utilizada uma versão com as instâncias embaralhadas.

EEG Eye State (DUONG; RAMAMPIARO; NØRVÅG, 2018): registros de Eletroencefalograma (EEG) com 14.980 instâncias e 15 atributos referentes as condições de olhos abertos e fechados.

Electricity (versão normalizada) (GAMA et al., 2004; ŽLIOBAITĖ et al., 2015): tem 45.312 instâncias e oito atributos de dados coletados do mercado australiano de eletricidade de New South Wales, onde os preços dependem da demanda e oferta do mercado. O preço varia a cada cinco minutos e o objetivo do problema é prever se aumentará ou diminuirá em relação a uma média móvel das últimas 24 horas. Também foi utilizada uma versão com as instâncias embaralhadas.

Give Me Some Credit (ZIEBA; HÄRDLE, 2018): conjunto de dados usado com o objetivo de melhorar a pontuação de crédito, prevendo a probabilidade de alguém passar por dificuldades financeiras nos próximos dois anos. No experimento desta pesquisa esses dados foram usados para prever se uma pessoa ficou inadimplente. Essa base possui 150.000 instâncias e 12 atributos.

Letter Recognition (FREY; SLATE, 1991): possui 20000 instâncias e 16 atributos, cujo o objetivo é identificar letras maiúsculas do alfabeto representadas por pixels retangulares em preto e branco.

Nomao Challenge (CANDILLIER; LEMAIRE, 2013): informações de diversos lugares obtidos de várias fontes que precisam ser agregadas corretamente. A tarefa consiste em detectar quais dados se referem ao mesmo lugar. O conjunto de dados tem 34.465 instâncias e 120 atributos.

Outdoor Objects (LOSING; HAMMER; WERSING, 2015): consiste em conjunto de imagens de objetos gravadas por um celular. A tarefa é classificar 40 objetos diferentes, abordados dez vezes sob diferentes condições de iluminação. O conjunto de dados tem um total de 4.000 instâncias com 21 atributos.

Pokerhand (BARROS et al., 2017): representa o problema de identificar o valor de uma mão contendo cinco cartas num jogo de Poker. As cartas são ordenadas pelos respectivos valores e naipes e as duplicações foram removidas. Essa base tem 829.201 instâncias e 11 atributos, e

trata-se também da versão normalizada.

Santander Customer Satisfaction (FERREIRA et al., 2018): o objetivo é identificar o cliente insatisfeito do banco Santander. Este conjunto de dados contém 76.020 instâncias com 371 atributos para prever se um cliente está satisfeito ou insatisfeito com relação ao Santander.

Sick (QUINLAN, 1986): contém 3.772 registros com 30 atributos relacionados à doenças da tireoide, o atributo alvo representa duas classes: doente, ou negativo.

Spam (KATAKIS; TSOUMAKAS; VLAHAVAS, 2010): consiste em 500 atributos e 9324 instâncias obtidas de emails da coleção Spam Assassin. As características dessas mensagens de spam mudam gradualmente ao longo do tempo, provocando uma mudança de conceito.

Usenets (BARROS; HIDALGO; CABRAL, 2018): foi construído a partir dos conjuntos de dados Usenet1 e Usenet2 (KATAKIS; TSOUMAKAS; VLAHAVAS, 2010; FRÍAS-BLANCO et al., 2015), que são baseados em 20.000 coleções de vinte newsgroups, ambos com 99 atributos e 1.500 instâncias. Eles simulam fluxos de mensagens apresentadas sequencialmente a um usuário, que as rotula como interessantes ou lixo, de acordo com seus interesses pessoais. Como esses dois conjuntos de dados são muito pequenos, eles foram concatenados para criar um conjunto de dados maior.

Weather (ELWELL; POLIKAR, 2011): representa medidas meteorológicas adquiridas ao longo de 50 anos tais como velocidade do vento, pressão, temperatura, visibilidade, nível do oceano e ponto de condensação da água. Trata-se de uma base binária com 9 atributos e 18.159 instâncias, em que as predições devem indicar se existe possibilidade de chuva ou não.

5.1.2.2 Bases de Eletroencefalograma (EEG)

Esta base de dados contém registros de EEG obtidos de recém-nascidos com as respectivas marcações de convulsões epiléticas definidas por três especialistas humanos (STEVENSON et al., 2018). Os registros foram obtidos a partir de 79 recém-nascidos da Unidade de Terapia Intensiva Neonatal do Hospital Universitário de Helsinque. A mediana da duração desses registros de EEG é de 74 minutos, com intervalo interquartil de 64 a 96 minutos. No conjunto de dados, cada especialista identificou uma média de 460 crises epiléticas. Por consenso, 39 neonatais tiveram crise convulsiva, enquanto que 22 não apresentaram crise. Os registros de EEG selecionados são apresentados na Tabela 10, bem como os respectivos pontos de mudanças marcados de forma aproximada a partir do consenso dos especialistas.

Esta base de dados pode ser usada como uma referência para o desenvolvimento de métodos

Tabela 10 – Bases de dados de EEG selecionadas com os respectivos pontos de mudanças.

Registro	Observações onde ocorrem mudanças devido a convulsões
EEG04	263.168, 488.960, 517.632, 528.640
EEG05	32.512, 65.536, 227.072, 249.344, 386.048, 429.824, 648.192, 680.448, 982.528
EEG07	4.096, 24.064, 28.672, 250.624, 284.672, 639.232, 675.328, 734.976, 777.984, 868.864, 907.008
EEG08	168.448, 176.640
EEG11	1.431.552, 1.436.928, 1.600.256, 1.608.704, 1.678.592, 1.690.112
EEG13	74.752, 204.800, 331.776, 2.178.816, 2.207.488, 2.262.784, 2.322.432, 2.398.720, 2.434.048
EEG33	508.928, 657.920
EEG51	900.608, 904.704, 961.024, 965.120, 1.090.816, 1.096.960, 1.114.880, 1.199.104
EEG63	82.176, 101.120, 413.952, 441.600, 620.032, 646.656, 787.456, 798.720, 916.736, 927.744
EEG68	376.832, 385.280
EEG75	315.904, 551.424
EEG76	99.840, 111.616, 202.496, 312.832
EEG79	37.376, 48.384, 93.440, 98.304, 144.384, 158.720, 547.584, 554.240, 624.640, 638.464

Fonte: Elaborada pelo autor (2021).

automatizados de detecção de convulsões neonatais e outras análises de EEG (STEVENSON et al., 2018). Nesta pesquisa, foram selecionados 13 dos 79 registros de EEG, apresentados na Tabela 10 com os respectivos pontos de mudanças definidos a partir das informações dos três especialistas. Para o experimento, apenas os sinais correspondentes a dois eletrodos da linha média (Pz e Fz) foram utilizados, ou seja, um total de 26 fluxos de dados. De acordo com o Sistema 10-20 (KLEM et al., 1999), esses eletrodos encontram-se na região entre os dois hemisférios cerebrais, de modo que foram selecionados devido a apresentarem maior dificuldade para os detectores em relação aos sinais obtidos por eletrodos dispostos sobre os hemisférios esquerdo ou direito. O desafio é que esses fluxos de dados de EEG não possuem rótulos, consequentemente, não é possível aplicar diretamente os métodos supervisionados de detecção. Entretanto, este é um cenário apropriado para a aplicação da rotulação simbólica descrita na Seção 4.3 com a finalidade de detectar *mudanças de conceito* relacionadas à alterações dos sinais de EEG.

5.2 CONFIGURAÇÕES DOS EXPERIMENTOS

Esta seção descreve as configurações dos experimentos realizados para avaliar os métodos apresentados no Capítulo 3 em comparação com uma seleção de detectores apresentados no Capítulo 2. Mais especificamente, a Tabela 11 apresenta os detectores selecionados e utilizados nos experimentos. A primeira linha ($\#Param.$) da tabela apresenta a quantidade de parâmetros

de cada detector, os quais foram mantidos com a configuração padrão do MOA. A segunda linha (#Observ.) apresenta o tamanho das janelas de observações para cada detector, com exceção do ADWIN, o qual possui janelas de tamanho variável, e do HDDMA, que computa as estatísticas de modo incremental, ou seja, sem necessidade do armazenamento prévio das observações.

Tabela 11 – Detectores utilizados nos experimentos com, respectivamente, o número de parâmetros, e o tamanho das janelas de observações.

	ADWIN	CUSUM	DDM	ECDD	EDDM	FHDDM	FTDD	HDDMA	MDDMA	PHT	RDDM	SEED	SEQDR2	STEPD	WSTD
#Param.	2	3	3	2	3	2	3	3	3	4	6	5	2	3	4
#Observ.	-	30	30	30	30	25	30	-	100	30	129	32	200	30	30

Fonte: Elaborada pelo autor (2021).

Para a avaliação da acurácia, foi utilizada a metodologia *Prequential* (GAMA et al., 2014), onde ambos os classificadores incrementais Naïve Bayes (NB) (JOHN; LANGLEY, 1995) e o Very Fast Decision Tree (VFDT) (HULTEN; SPENCER; DOMINGOS, 2001) (referenciado apenas como Hoeffding Tree (HT) daqui em diante) foram considerados porque são os mais frequentemente usados em experimentos da literatura e estão disponíveis no MOA.

No *Prequential*, cada instância é usada primeiramente para teste e em seguida para treinamento. Essa metodologia presume que os dados estão distribuídos de forma independente (BIFET et al., 2013), podendo levar a conclusões errôneas se considerada de modo isolado em cenários onde as classes das instâncias são distribuídas de forma desbalanceada, ou os rótulos apresentarem dependência temporal. Devido a essa limitação, também foram utilizados o classificador ingenuo denominado *Classificador de Classe Majoritária*, e o *Classificador Temporalmente Aumentado* (do Inglês, Temporally Augmented Classifier), como recomendado em Žliobaitė et al. (2015). Assim, dado que os exemplos não sejam Independente e Identicamente Distribuídas (i.i.d.), de modo que os rótulos contenham alguma dependência temporal, a acurácia do *Classificador Temporalmente Aumentado* será maior que a do *Classificador de Classe Majoritária* (ŽLIOBAITĖ et al., 2015). É importante ressaltar que esses dois classificadores são utilizados como referenciais para a avaliação dos demais detectores com relação a bases desbalanceadas ou com dependência temporal.

Os experimentos com as bases artificiais obtidas com os geradores do MOA foram realizados para avaliar a resiliência dos detectores em termos da detecção de *mudanças de conceito* e da acurácia em diversos cenários, envolvendo a distribuição desbalanceada das classes, níveis de ruídos, problemas multiclases, e dependência temporal. Além disso, essas bases também foram utilizadas para avaliar o uso de recursos computacionais em termos de memória e tempo

de processamento. Desse modo, essas avaliações buscam responder as questões de pesquisas **QP 1** e **QP 2**. Além dessas bases, os fluxos de dados gerados por meio do Sistema de Lorenz Transiente foram utilizados para avaliar o Symbolic Labeling Adapter (SLA) em diferentes reconstruções do *espaço de fase*, abrangendo as avaliações das questões **QP 3** e **QP 4**.

As bases reais foram utilizadas com objetivo de generalizar os resultados obtidos com as bases artificiais a cenários mais realistas. Vale destacar que nesse contexto não há garantia de que os dados sejam i.i.d., ou saber quando uma *mudança de conceito* pode ocorrer. Com o intuito de simular essas condições, as diversas bases da Seção 5.1.2 foram usadas para avaliar a acurácia e a detecção de *mudança de conceito* com o objetivo de identificar quais detectores apresentam o melhor equilíbrio entre essas duas avaliações. De modo semelhante à avaliação com fluxos de dados gerados por meio do Sistema de Lorenz Transiente, as bases de EEG também foram utilizadas para avaliar o SLA. O desafio é que o fluxo de dados de EEG também não possui rótulos, conseqüentemente, não é possível aplicar diretamente os métodos supervisionados de detecção nessas bases. Com esses experimentos, busca-se generalizar as respostas das questões de pesquisa **QP 1** a **QP 4** em aplicações do mundo real.

As implementações dos métodos foram realizadas no MOA, e os experimentos foram preparados e analisados com o uso da ferramenta MOAManager (MACIEL; SANTOS; BARROS, 2020) (disponível em <<https://github.com/brunom4ciel/moamanager/wiki>>) em conjunto com o ambiente para computação estatística R (disponível em <<https://www.R-project.org/>>). Todos os testes foram executados em um PC equipado com Core i7-Processador 8550U, 32 GB de RAM e sistema operacional Ubuntu Desktop 18.04 LTS de 64 bits.

5.2.1 Métricas de avaliação das detecções de mudanças de conceito

Com exceção das bases reais da Seção 5.1.2, as demais utilizadas nos experimentos possibilitam identificar os pontos de *mudanças de conceito*. Portanto, seguindo o protocolo descrito em Bifet (2017), antes que uma mudança de fato aconteça, todos os alarmes são considerados alarmes falsos. Após a ocorrência de uma mudança verdadeira, a primeira detecção sinalizada é considerada um alarme verdadeiro. Depois dessa e antes que uma nova mudança ocorra, as detecções subsequentes são consideradas alarmes falsos. Caso nenhuma detecção ocorra entre duas mudanças verdadeiras, será contabilizado uma detecção perdida (falso negativo). Para avaliar a qualidade das detecções com base nesse protocolo, as seguintes métricas foram utilizadas (BASSEVILLE; NIKIFOROV, 1993; BIFET, 2017):

- Mean Time Between False Alarms (traduzido para o português, Tempo Médio Entre Alarmes Falsos) (MTFA):

$$MTFA = \mathbb{E}(i_a - i_0 | p_{\theta_0}), \quad (5.1)$$

onde i_0 é o índice da primeira instância e i_a indica a quantidade de instâncias até o alarme. O MTFA define a frequência com que um detector dispara alarmes falsos, ou seja, alarmes que ocorrem quando não há mudança de fato. O MTFA está relacionado com a *taxa de alarmes falsos*, definida como $1/MTFA$. Logo, um bom detector deve ter um MTFA elevado.

- Mean Time for Detection (traduzido para o português, Tempo Médio de Detecção) (MTD):

$$MTD = \mathbb{E}(i_a - i_k | p_{\theta_1}), \quad (5.2)$$

define a latência ($i_a - i_k$) de alarme do detector após a ocorrência de uma *mudança de conceito* ($p_{\theta_0} \neq p_{\theta_1}$) ocorrida no instante i_k . Idealmente o MTD deve ser o menor possível.

- Missed Detection Rate (traduzido para o português, Taxa de Detecções Perdidas) (MDR) refere-se a probabilidade de não alarmar uma *mudança de conceito* quando esta de fato ocorre. Na prática o MDR deve ser pequeno ou zero, pois mede o percentual de falsos negativos.
- Average Running Length (traduzido para o português, Comprimento Médio de Sequência) (ARL):

$$ARL(\theta = \theta_0) = MTFA, \quad ARL(\theta = \theta_1) = MTD, \quad (5.3)$$

define o número médio em instâncias sem que o detector dispare um alarme, onde θ é o parâmetro relacionado à magnitude da mudança. Trata-se de uma generalização das métricas MTFA e MTD.

- Mean Time Ratio (traduzido para o português, Razão de Tempo Médio) (MTR):

$$MTR = (1 - MDR) \times MTFA/MTD, \quad (5.4)$$

reflete o equilíbrio entre detecções rápida e alarmes falsos por meio da razão entre MTFA e MTD ponderada pela probabilidade de detecção de uma *mudança de conceito*.

Os detectores que têm um MTR elevado em um mesmo cenário de *mudança de conceito* devem ser preferidos (BIFET, 2017). Vale ressaltar que nos experimentos $MTD = \ell$ e $MTFA = \ell$ sempre que uma detecção perdida é emitida, onde ℓ é o comprimento do conceito.

5.2.2 Testes estatísticos

Na análise dos resultados, todas as medidas estatísticas foram calculadas considerando um intervalo de confiança de 95%. Para a comparação dos resultados, o teste estatístico F_F , fundamentado no teste não-paramétrico de Friedman (DEMSAR, 2006), foi utilizado para identificar diferenças estatísticas significativas considerando o nível de significância α de 5%. Nesse teste, a hipótese nula afirma que não há diferença entre os métodos avaliados, e a sua refutação significa que existe uma ou mais diferenças estatísticas entre os detectores. Nesse caso, o teste *post-hoc* de Nemenyi (NEMENYI, 1962) é aplicado para comparar todos os detectores contra os demais usando uma Critical Differences (traduzido para o português, Diferença Crítica) (CD). Esse teste é exibido em uma notação gráfica, denominada daqui por diante como *gráfico de CD*, em que a hipótese nula é representada por uma barra conectando métodos que não são estatisticamente diferentes.

O *p-value* obtido do teste de Friedman é apresentado como P no canto superior direito de cada gráfico de CD. Assumindo um nível de significância de 5%, pode-se rejeitar com segurança a hipótese nula quando $P < 0,05$. Nesse caso, o teste *post-hoc* de Nemenyi pode ser considerado para identificar quais algoritmos são diferentes, onde o nível de significância também é definido em 5%. A CD obtida no teste *post-hoc* de Nemenyi é apresentado no canto superior esquerdo, e o ranqueamento do teste está na parte inferior do gráfico, em ordem crescente, relacionado a cada método localizado desde o canto superior esquerdo ao canto inferior direito do *gráfico de CD*.

O critério de qualidade adotado está relacionado com a barra de Critical Differences (traduzido para o português, Diferença Crítica) (CD) ligada ao topo do ranque. Por esse critério, são bem avaliados os métodos ligados à barra conectada àquele com o melhor desempenho em termos da acurácia (ACC), MDR, MTD, MTFA, ou MTR. O título nos *gráficos de CD* correspondem a cada métrica com um subscrito HT ou NB referentes ao respectivo classificador, além dos subscritos (A) e (G), para os testes nas bases com mudanças abruptas e graduais, respectivamente. Com base nesse critério de qualidade, apenas os algoritmos ligados à barra

de CD do topo do ranqueamento serão considerados como competitivos.

5.2.3 Análise de sensibilidade dos parâmetros

Os detectores Recurrence Quantification Analysis Drift Detector (RQADD), Spectral Entropy Drift Detector (SEDD), e Permutation Entropy Drift Detector (PEDD), possuem três parâmetros em comum: comprimento da expansão de dígitos binários L , tamanho da janela deslizante N , e magnitude δ . Ambos os detectores RQADD e PEDD ainda precisam definir os parâmetros da dimensão m e do tempo de defasagem τ . Além desses, RQADD necessita de um valor apropriado para o raio de vizinhança ε . Experimentos preliminares foram realizados a fim de entender a influência de cada um desses parâmetros nos resultados dos detectores, bem como propor seus respectivos valores padrões.

Os valores do parâmetro L foram avaliados no conjunto $\{4, 8, 16, 32\}$. Embora um L grande esteja relacionado a uma melhor resolução da simulação do Mapa de Bernoulli, o qual pode produzir verdadeiras órbitas periódicas com períodos arbitrariamente longos, ou mesmo aproximações de órbitas caóticas, $L = 8$ apresentou o melhor compromisso em nossa proposta para a detecção de *mudança de conceito* para SEDD. Para RQADD e PEDD, os valores $\{8, 16, 32\}$ não apresentaram diferenças significativas, ficando definido como $L = 32$ em ambos. Esse valor define maior granularidade numérica, e conseqüentemente, favorece o aumento da resolução dos Diagrama de Recorrência (RP) obtidos.

O tamanho N da janela deslizante afeta a variância do sinal de retorno dos detectores, o tempo de execução, e o consumo de memória. Mais especificamente, aumentar o valor de N reduz a variância ao custo de aumentar o tempo de execução e o consumo de memória. Por outro lado, um valor pequeno de N reduz a resolução espectral e pode não produzir um Espectro de Potência (PS), ou um *espaço de fase*, adequadamente. Nos experimentos preliminares, os valores testados para N foram $\{64, 128, 256, 512\}$. O valor $N = 128$ apresentou o melhor compromisso entre uma resolução espectral apropriada e o custo computacional para o SEDD. No caso da PEDD, o valor $N = 512$ foi utilizado para suprir uma amostragem para a dimensão $m = 4$, ou seja, $m! = 4! = 24$ padrões. Para o RQADD, o valor $N = 256$ foi usado devido ao elevado custo computacional da Análise de Quantificação de Recorrência (RQA). Como ilustrado na Figura 23, o aumento do tamanho N da janela deslizante resulta em um gráfico mais suavizado, como exibindo nas Figuras 23 (D) e (E).

Para o PEDD e o RQADD, a defasagem foi definida usando $\tau = 1$, e o raio de vizinhança

ε foi definido dinamicamente para o RQADD considerando uma densidade fixa de 25% do RP. Vale salientar que foram avaliadas densidades no intervalo $\{1\%, \dots, 25\%\}$, em que a densidade de 25% apresentou melhores resultados em termos de detecção e evitou a quebra de algumas simulações nos experimentos.

A magnitude δ está relacionada ao atraso na detecção. Valores mais baixos levam a alarmes mais rápidos, mas ao custo de uma taxa maior de falsos positivos. Semelhante aos sinais das entropias e do determinismo, δ está definido no intervalo $[0, 1]$. Nos experimentos preliminares, o valor δ foi ajustado eurísticamente até alcançar os valores apropriados para cada um dos detectores. Assim, os valores padrões propostos para o algoritmo SEDD são $L = 8$, $N = 128$ e $\delta = 0,37$; para o algoritmo PEDD são $L = 32$, $N = 512$, $m = 4$, $\tau = 1$ e $\delta = 0,16$; e para o RQADD são $L = 32$, $N = 256$, $m = 2$, $\tau = 1$, $\delta = 0,17$, e ε definido dinamicamente para uma densidade fixa de 25% do RP.

5.3 RESULTADOS E ANÁLISES DO EXPERIMENTOS COM BASES ARTIFICIAIS

Esta seção apresenta os resultados experimentais e a avaliação estatística para a comparação dos detectores em termos da acurácia e das detecções considerando as configurações do conjunto de dados artificiais obtidas com os geradores descritos na Seção 5.1.1.1. Também são discutidos os aspectos relevantes dos experimentos e a análise de seus resultados.

As Tabelas 13 a 18 apresentam os resultados obtidos usando o *Classificador de Classe Majoritária*, o *Classificador Temporalmente Aumentado*, o classificador HT isolado e com os detectores descritos na Seção 5.2, considerando as bases artificiais $Agrawal_B$, $Agrawal_D$, $LED_{5\%}$, $LED_{10\%}$, RBF_2 , e RBF_5 , respectivamente. Como apresentado nas tabelas, os resultados levam em conta as configurações de mudanças abruptas e graduais. As colunas das tabelas indicam os resultados obtidos por cada detector com respeito à acurácia e às métricas referentes à detecção de *mudança de conceito*: MTD, MTFA, MDR e MTR. Nas tabelas, o melhor resultado é mostrado em **negrito**. No caso de resultados com os mesmos valores, destaca-se como o melhor detector aquele com o menor intervalo de confiança. Vale salientar que detecções perdidas fornecem médias mais altas das medidas MTD e MTFA. Isso ocorre devido a convenção adotada nesta tese, conforme explicado na Seção 5.2, o $MTD = \ell$ e o $MTFA = \ell$ sempre que uma detecção perdida é emitida, onde ℓ é o número de instancias do conceito.

As Figuras 24 a 35 apresentam os gráficos de CD obtidos dos testes estatísticos relacionados ao resultados resumidos nas tabelas 13 a 18. Os testes referem-se às comparações com base

na acurácia (ACC), MTD, MTFA, MDR e MTR, considerando ambos os conjuntos de dados artificiais com mudanças abruptas (A) e graduais (G). As variáveis $\overline{MTD} = \ell - MTD$ e $\overline{MDR} = 1 - MDR$ foram definidas para atender a ordem do ranqueamento dos testes estatísticos, em que valores mais altos significam melhores desempenhos. De forma semelhante, as Tabelas 26 a 31 do Apêndice A referem-se aos resultados com base no classificador NB. As Figuras 46 a 57 apresentam os resultados dos testes estatísticos relacionados aos dados resumidos nas Tabelas 26 a 31. Vale ressaltar que diferentes perspectivas em relação às métricas utilizadas na avaliação podem ser consideradas, e isso depende da aplicação alvo. Por exemplo, para algumas aplicações obter um MTD baixo pode ser mais importante que uma boa acurácia. Nesta pesquisa, considera-se que os melhores algoritmos são aqueles que apresentam um equilíbrio entre a acurácia e a qualidade da detecção de *mudança de conceito* fornecido pelo MTR e estão ligados à barra de CD do topo do ranqueamento (são os mais competitivos).

Os resultados das bases com as configurações $Agrawal_{(B)}$ e $Agrawal_{(D)}$ usando o HT são relatadas nas Tabelas 13 e 14. A agregação dos gráficos de CD relacionados a esses resultados são apresentados nas Figuras 24 e 25, referentes à configuração $Agrawal_{(B)}$, e as Figuras 26 e 27 resumem as avaliações estatísticas com a configuração $Agrawal_{(D)}$.

Com relação as bases $LED_{(5\%)}$ e $LED_{(10\%)}$, também usando o HT, a acurácia e as métricas de detecção dos algoritmos estão resumidas nas Tabelas 15 e 16. As agregações dos gráficos de CD obtidos dos resultados dos testes estatísticos são apresentadas nas Figuras 28 e 29, referentes à configuração $LED_{(5\%)}$, e nas Figuras 30 e 31, referentes à $LED_{(10\%)}$. De modo semelhante, os resultados do experimento das configurações $RBF_{(2)}$ e $RBF_{(5)}$ usando HT são resumidos nas Tabelas 17 e 18. A agregação dos gráficos de CD relacionados a esses resultados são apresentados nas Figuras 32 e 33, referentes à configuração $RBF_{(2)}$, e as Figuras 34 e 35 resumem as avaliações estatísticas da base $RBF_{(5)}$. A descrição detalhada dessas configurações referente aos resultados com o classificador NB encontra-se no Apêndice A. Para facilitar a exposição das informação relacionadas ao resultado de todas as configurações das bases deste experimento, considerando ambos os classificadores HT e NB, as seções a seguir condensam a análise dos resultados em relação às bases $Agrawal_{(B)}$ e $Agrawal_{(D)}$, $LED_{(5\%)}$ e $LED_{(10\%)}$, $RBF_{(2)}$ e $RBF_{(5)}$. Por fim, uma avaliação em termos do custo computacional encerra a discussão deste experimento.

5.3.1 Análise dos resultados com as bases artificiais

Esta seção apresenta um resumo dos resultados experimentais com todas as bases artificiais, considerando classes balanceadas e desbalanceadas, as configurações com 5% e 10% de ruído, e com 2 e 5 classes. As Seções 5.3.1.1, 5.3.1.2 e 5.3.1.3 apresentadas a seguir, detalham esses resultados. A Tabela 12 agrega as distribuições das frequências relativas à contagem em que cada detector não apresentou diferenças estatísticas significativas com aquele que ficou no topo do ranqueamento ao longo dos testes em cada métrica.

Trata-se da proporção em que um detector esteve conectado à barra de Critical Differences (traduzido para o português, Diferença Crítica) (CD) do topo do ranqueamento em uma dada métrica em relação aos demais, ou seja, o número de vezes em que um dado detector não apresentou diferenças estatísticas significativas com aquele que ficou no topo do ranqueamento, dividido pela soma total do número de vezes em que todos os detectores estiveram conectados a esta barra de Critical Differences (traduzido para o português, Diferença Crítica) (CD) em uma métrica específica. Assim, quanto maior o percentual, mais frequentemente o detector foi competitivo em relação aos demais.

Estes percentuais foram agrupados levando em consideração as configurações de mudanças abruptas e graduais, onde as maiores proporções estão destacadas em **negrito**. Vale salientar que o critério de qualidade adotado está relacionado com a barra de CD ligada ao topo do ranque (detectores competitivos). Por esse critério, são bem avaliados os detectores ligados à barra conectada àquele com o melhor desempenho em termos da acurácia (ACC), e das métricas MDR, MTD, MTFA, e MTR.

Considerando o menor número de detectores competitivos que soma próximo de 50% dos percentuais (incluindo percentuais repetidos) em relação às mudanças abruptas da Tabela 12, pode-se observar que os detectores CUSUM, SEDD, HDDMA, RDDM, e FTDD abrangem 54,44% de todos os resultados competitivos em termos da acurácia (ACC). Em relação às demais métricas, os detectores STEPD, MDDMA, SEDD, e ECDD somam juntos 46,91% de todos os resultados competitivos com a métrica MTD. Em relação ao MTFA, o PHT, SEDD, FTDD, HDDMA, CUSUM somam juntos 56,04% dos resultados competitivos. Em termos do MDR, os detectores PHT, HDDMA, CUSUM, RQADD, ADWIN, RDDM, PEDD, DDE, e STEPD abrangem 53,20% dos melhores resultados. E por fim, o SEDD, HDDMA, MDDMA, CUSUM somam juntos 50,70% dos resultados competitivos em termos do MTR.

Nas mudanças graduais, os detectores CUSUM, HDDMA, PHT, RDDM, e SEDD abrangem

Tabela 12 – Percentual das frequências em que os detectores ficaram na barra ligada ao topo do ranqueamento dos gráficos de CD.

<i>Mudança</i>	<i>Classif./Detect.</i>	#ACC (%)	#MTD (%)	#MTFA (%)	#MDR (%)	#MTR (%)
Abrupta	ADWIN	2,22%	7,41%	2,20%	5,91%	8,45%
	CUSUM	13,33%	2,47%	9,89%	5,91%	9,86%
	DDM	3,33%	0,00%	6,59%	5,42%	0,00%
	ECDD	0,00%	8,64%	0,00%	5,42%	0,00%
	EDDM	0,00%	0,00%	1,10%	4,43%	0,00%
	FHDDM	5,56%	7,41%	5,49%	4,93%	2,82%
	FTDD	7,78%	7,41%	10,99%	5,42%	9,86%
	HDDMA	11,11%	6,17%	10,99%	5,91%	12,68%
	MDDMA	7,78%	13,58%	4,40%	5,42%	12,68%
	PHT	5,56%	0,00%	13,19%	5,91%	1,41%
	RDDM	10,00%	3,70%	2,20%	5,91%	5,63%
	SEED	1,11%	7,41%	0,00%	5,91%	2,82%
	SEQDR2	2,22%	1,23%	5,49%	5,42%	2,82%
	STEPD	1,11%	14,81%	0,00%	5,91%	1,41%
	WSTD	5,56%	6,17%	6,59%	4,93%	8,45%
	PEDD	2,22%	1,23%	2,20%	5,91%	0,00%
	RQADD	6,67%	2,47%	7,69%	5,91%	5,63%
SEDD	12,22%	9,88%	10,99%	5,42%	15,49%	
Gradual	ADWIN	2,27%	13,70%	2,22%	8,63%	12,31%
	CUSUM	11,36%	0,00%	11,11%	2,88%	0,00%
	DDE	0,00%	0,00%	0,00%	0,00%	0,00%
	DDM	6,82%	2,74%	8,89%	3,60%	9,23%
	ECDD	0,00%	13,70%	0,00%	7,91%	0,00%
	EDDM	0,00%	5,48%	1,11%	5,76%	3,08%
	FHDDM	4,55%	5,48%	4,44%	4,32%	6,15%
	FTDD	6,82%	1,37%	11,11%	3,60%	4,62%
	HDDMA	11,36%	2,74%	11,11%	3,60%	10,77%
	MDDMA	7,95%	2,74%	3,33%	5,04%	3,08%
	PHT	10,23%	9,59%	13,33%	3,60%	13,85%
	RDDM	10,23%	0,00%	1,11%	7,91%	0,00%
	SEED	1,14%	15,07%	0,00%	8,63%	9,23%
	SEQDR2	3,41%	2,74%	3,33%	7,91%	6,15%
	STEPD	1,14%	10,96%	0,00%	7,91%	0,00%
	WSTD	0,00%	0,00%	5,56%	3,60%	1,54%
	PEDD	2,27%	8,22%	3,33%	7,19%	1,54%
RQADD	7,95%	4,11%	8,89%	4,32%	9,23%	
SEDD	10,23%	1,37%	11,11%	3,60%	9,23%	

Fonte: Elaborada pelo autor (2021).

53,41% dos resultados competitivos em termos de acurácia (ACC). Em relação ao MTR, o SEED, ADWIN, ECDD, e STEPDP somam 53,42% dos resultados competitivos. Em termos do MTFA, os detectores PHT, HDDMA, SEDD, FTDD, e CUSUM abrangem 57,78% dos melhores resultados. Juntos, o ADWIN, SEED, SEQDR2, RDDM, ECDD, STEPDP, e o PEDD somam 56,12% dos resultados competitivos em termos de MDR. Finalmente, os detectores PHT, ADWIN, e HDDMA abrangem 36,92% dos melhores resultados em termos de MTR.

Considerando mudanças abruptas, o CUSUM foi o mais competitivo em 13,33% dos testes entre todos os detectores com relação à acurácia, conseqüentemente, essa é a proporção do número de vezes em que o CUSUM esteve conectado à barra de CD ligada ao topo do ranqueamento relativo a todas as bases artificiais, considerando classes balanceadas e desbalanceadas, as configurações com 5% e 10% de ruído, e com 2 e 5 classes, em comparação com a soma total do número de vezes em que todos os detectores também estiveram conectados à barra de CD do topo do ranqueamento. Do mesmo modo, o SEDD foi o mais competitivo em 15,49% dos testes relativos ao MTR, em comparação aos demais detectores. Em termos de MTD, o STEPDP foi o mais competitivo com um percentual de 14,81% da soma entre todos os resultados competitivos obtidos em conjunto com os demais detectores. Em relação ao MTFA, o PHT obteve 13,19% de todos os resultados competitivos entre todos os detectores. Finalmente, em termos de MDR, os detectores ADWIN, CUSUM, HDDMA, PHT, RDDM, SEED, STEPDP, PEDD, e RQADD conseguiram, cada um, 5,98% de todos os resultados competitivos obtidos entre todos os detectores.

Mudando o foco para as mudanças graduais, o SEED obteve 15,07% dos resultados competitivos entre todos os detectores com relação ao MTD, e também foi responsável por 8,63% dos resultados competitivos entre todos os obtidos pelos demais detectores em termos do MDR. O PHT conseguiu 13,33% dos resultados competitivos entre todos os detectores, em relação ao MTFA, e foi competitivo em 13,85% dos resultados com MTR, entre todos os resultados competitivos obtidos pelos demais detectores. O CUSUM e o HDDMA conseguiram obter, cada um, 11,36% de todos os resultados competitivos relacionados à acurácia (ACC), entre todos os detectores. Por fim, o ADWIN conseguiu obter 8,63% dos resultados competitivos em termos da métrica MDR.

As análises dos testes estatísticos sugerem que o desempenho dos detectores nas bases com mudanças graduais foram inferiores em relação as bases abruptas em termos de acurácia e da qualidade da detecção de *mudança de conceito*. A taxa de detecções perdidas MDR foi significativamente maior nas bases com mudanças graduais.

Os resultados experimentais também apontam que os métodos propostos apresentaram menor desempenho com problemas multiclases. Outro ponto a destacar é que nenhum dos detectores apresentou desempenho superior, em termos de acurácia, em comparação ao *Classificador Temporalmente Aumentado* nas configurações onde a dependência temporal dos rótulos foi mais acentuada. Entretanto, os detectores CUSUM, SEDD, HDDMA, RDDM, PHT, e FTDD foram os mais competitivos em relação a esse classificador na maioria das avaliações.

Por fim, os resultados apontam que o SEDD apresentou os melhores resultados entre os três detectores propostos, seguido do RQADD. O PEDD não foi competitivo na maioria das avaliações. Logo, os detectores fundamentados na Nonlinear Time Series Analysis (traduzido para o português, Análise de Séries Temporais Não Lineares) (NLTA) não apresentaram desempenho superior ao estado da arte em termos de acurácia e detecção neste experimento. Entretanto, o SEDD apresentou o melhor desempenho geral na detecção de mudanças abruptas, sendo bastante competitivo em relação a acurácia aos demais detectores da literatura.

5.3.1.1 Resultados e análises considerando classes balanceadas e desbalanceadas

Os resultados experimentais com classes balanceadas ($Agrawal_{(B)}$) mostram que o *Classificador Temporalmente Aumentado* obteve a melhor acurácia em relação ao *Classificador de Classe Majoritária*, ao classificador HT isolado, e a todos os detectores, considerando tanto mudanças abruptas quanto graduais, como apresentado na Tabela 13. A Tabela 26 do Apêndice A mostra que esse classificador também apresentou melhor acurácia que todos os detectores usando o NB. Um ponto a destacar é que os detectores com HT obtiveram uma acurácia superior quando comparados com o NB.

A vantagem da acurácia do *Classificador Temporalmente Aumentado* sobre os outros classificadores e detectores usando o HT nas bases com classes balanceadas ($Agrawal_{(B)}$), considerando mudanças abruptas, é confirmada no gráfico de CD $ACC_{HT(A)}$ (Figura 24). Nessa configuração, apenas o SEDD foi competitivo com *Classificador Temporalmente Aumentado* entre os detectores propostos. Também foram competitivos, os detectores CUSUM, FTDD, HDDMA, e MDDMA. No gráfico de CD $ACC_{HT(G)}$ (Figura 25), referentes aos testes estatísticos obtidos da configuração com mudanças graduais, o *Classificador Temporalmente Aumentado* também obteve vantagem na performance em relação à acurácia sobre os demais classificadores e detectores usando o HT. Do mesmo modo, apenas o SEDD foi competitivo com *Classificador Temporalmente Aumentado* entre os detectores propostos. Os detectores

Tabela 13 – Resultados obtidos usando HT, considerando a base artificial Agrawal balanceada, com um intervalo de confiança de 95%.

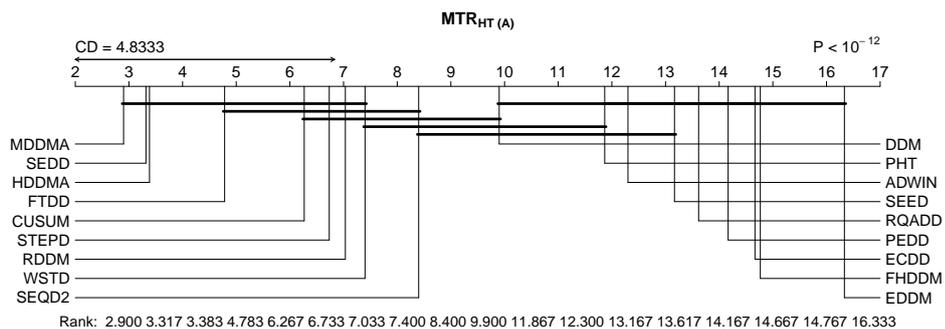
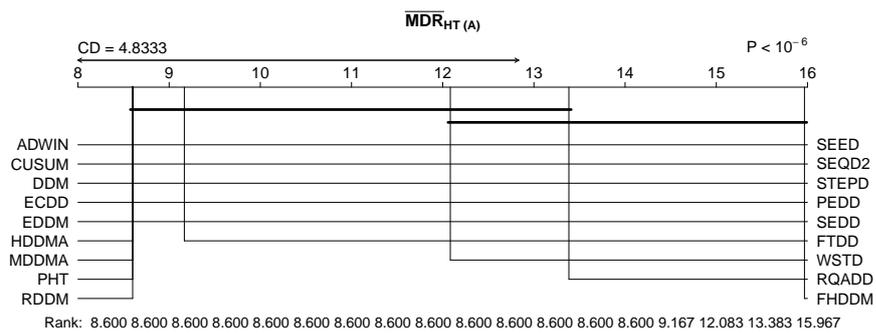
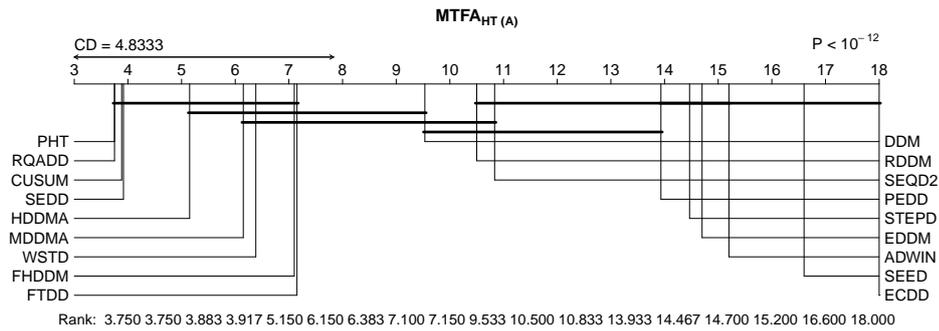
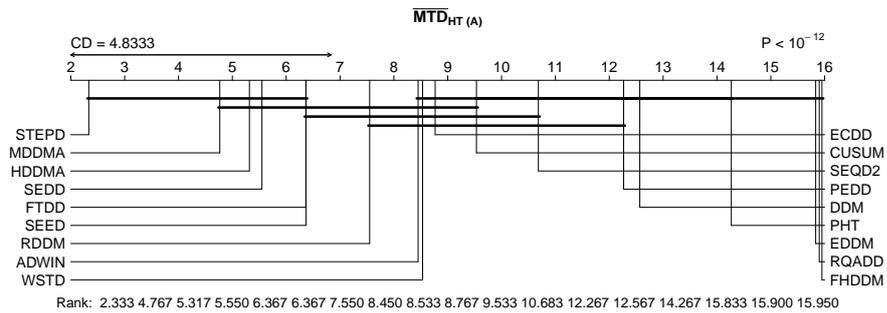
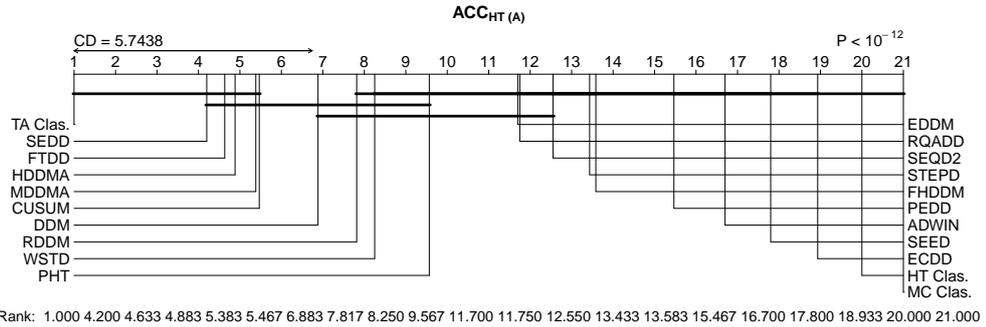
$Agrawal_{(B)}$	$Classificador/Detector_{(HT)}$	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Hoeffding Tree	82,01 ± 0,25					
	Majority Class	39,90 ± 0,00					
	Temporally Augmented Classifier	99,97 ± 0,00					
	ADWIN	88,46 ± 0,06	202,93 ± 114,54	4743,71 ± 135,14	0,00 ± 0,00	41,50 ± 7,84	
	CUSUM	93,88 ± 0,06	168,44 ± 4,00	50888,05 ± 638,00	0,00 ± 0,00	303,40 ± 8,58	
	DDM	93,81 ± 0,05	375,81 ± 7,43	35436,21 ± 4044,22	0,00 ± 0,00	94,41 ± 10,78	
	ECDD	85,25 ± 0,06	1658,54 ± 1510,63	733,46 ± 14,07	0,00 ± 0,00	24,93 ± 12,82	
	EDDM	92,14 ± 0,20	3236,86 ± 104,42	8301,18 ± 2505,74	0,00 ± 0,00	2,57 ± 0,76	
	FHDDM	90,37 ± 0,99	17811,66 ± 4621,10	44366,73 ± 1699,70	0,34 ± 0,09	87,66 ± 79,05	
	FTDD	93,83 ± 0,17	968,25 ± 1203,58	43600,54 ± 3061,15	0,02 ± 0,02	747,77 ± 234,38	
	HDDMA	93,88 ± 0,09	86,12 ± 22,41	48212,38 ± 1896,29	0,00 ± 0,00	719,66 ± 110,40	
	MDDMA	93,80 ± 0,13	67,38 ± 6,67	46072,96 ± 1822,12	0,00 ± 0,00	728,31 ± 73,76	
	PHT	93,43 ± 0,06	1079,07 ± 29,25	51200,00 ± 0,00	0,00 ± 0,00	47,69 ± 1,28	
	RDDM	93,66 ± 0,10	478,99 ± 767,86	30945,62 ± 4099,83	0,00 ± 0,00	298,98 ± 49,78	
	SEED	87,59 ± 0,11	529,87 ± 875,84	2457,51 ± 149,20	0,00 ± 0,00	31,34 ± 6,30	
	SEQDRIFT2	91,96 ± 0,14	205,00 ± 10,23	31561,83 ± 1884,96	0,00 ± 0,00	155,50 ± 10,41	
	STEPD	91,12 ± 0,36	241,69 ± 319,13	6205,12 ± 1058,04	0,00 ± 0,00	334,19 ± 81,05	
	WSTD	92,87 ± 0,47	5153,82 ± 2378,82	45684,75 ± 2057,51	0,10 ± 0,05	742,78 ± 246,80	
	Gradual	PEDD	89,36 ± 0,09	1039,39 ± 476,99	6009,79 ± 225,64	0,00 ± 0,00	19,02 ± 6,73
		RQADD	91,72 ± 0,69	10261,31 ± 3773,26	51200,00 ± 0,00	0,18 ± 0,08	40,86 ± 22,22
SEDD		93,95 ± 0,06	76,22 ± 4,60	50882,97 ± 648,41	0,00 ± 0,00	682,33 ± 37,22	
Hoeffding Tree		81,61 ± 0,20					
Majority Class		31,14 ± 0,00					
Temporally Augmented Classifier		98,91 ± 0,01					
ADWIN		87,91 ± 0,07	2194,67 ± 1213,93	4723,39 ± 167,93	0,00 ± 0,00	3,56 ± 0,55	
CUSUM		93,17 ± 0,07	43274,66 ± 2672,34	50901,73 ± 360,13	0,26 ± 0,03	0,90 ± 0,09	
DDM		93,28 ± 0,06	13031,64 ± 3055,45	46067,45 ± 1758,00	0,23 ± 0,05	34,66 ± 25,20	
ECDD		84,87 ± 0,07	3345,14 ± 678,24	778,94 ± 18,48	0,00 ± 0,00	0,36 ± 0,12	
EDDM		92,10 ± 0,18	5690,36 ± 2040,98	15688,43 ± 3541,99	0,06 ± 0,04	4,08 ± 1,30	
FHDDM		88,17 ± 0,91	33428,24 ± 4645,25	45832,12 ± 1446,51	0,64 ± 0,09	0,78 ± 0,32	
FTDD		93,00 ± 0,18	25082,13 ± 3104,94	47619,26 ± 1530,71	0,45 ± 0,06	1,28 ± 0,31	
HDDMA		93,13 ± 0,09	39211,68 ± 3430,11	48802,72 ± 1365,02	0,25 ± 0,05	1,00 ± 0,15	
MDDMA		93,03 ± 0,14	22884,34 ± 3468,56	47784,36 ± 1201,73	0,35 ± 0,05	1,66 ± 0,33	
PHT		93,14 ± 0,07	11218,86 ± 1947,60	51172,39 ± 7,52	0,20 ± 0,04	12,92 ± 7,74	
RDDM		92,93 ± 0,09	38868,12 ± 3640,92	26375,17 ± 2360,02	0,06 ± 0,04	0,69 ± 0,11	
SEED		87,06 ± 0,12	1836,00 ± 1184,13	2545,86 ± 187,47	0,00 ± 0,00	2,93 ± 0,70	
SEQDRIFT2		91,25 ± 0,14	16416,67 ± 3921,05	26029,82 ± 2273,13	0,06 ± 0,04	5,05 ± 3,21	
STEPD		90,91 ± 0,29	16386,17 ± 3594,02	14516,61 ± 2461,24	0,02 ± 0,03	1,41 ± 0,84	
WSTD	85,16 ± 0,69	42399,87 ± 2422,46	47054,95 ± 2227,83	0,80 ± 0,06	0,25 ± 0,08		
PEDD	88,77 ± 0,10	5881,63 ± 764,33	6204,52 ± 258,24	0,00 ± 0,00	1,17 ± 0,14		
RQADD	91,71 ± 0,47	23455,05 ± 2570,59	51126,24 ± 16,75	0,45 ± 0,05	3,78 ± 5,12		
SEDD	93,22 ± 0,06	29942,69 ± 2265,43	50835,70 ± 345,46	0,45 ± 0,04	0,95 ± 0,04		

Fonte: Elaborada pelo autor (2021).

CUSUM, DDM, FTDD, HDDMA, MDDMA, e PHT, também foram competitivos.

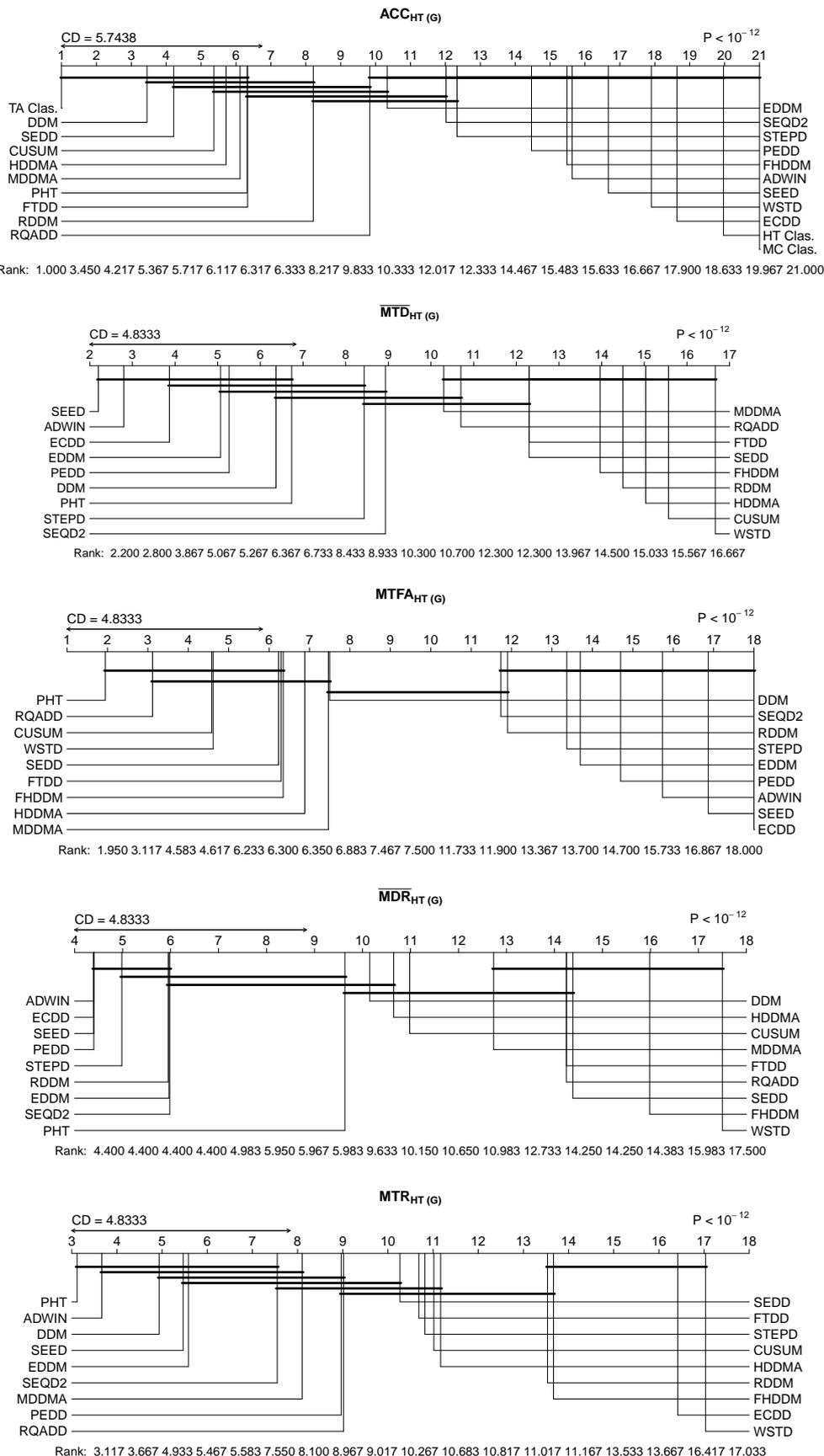
Ainda na configuração com a base de dados $Agrawal_{(B)}$, o *Classificador Temporalmente Aumentado* também fica no topo do ranqueamento com mudanças abruptas, considerando o NB, como indicado no gráfico de CD $ACC_{NB(A)}$ da Figura 46 do Apêndice A. Nessa

Figura 24 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 25 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

configuração, apenas o SEDD foi competitivo entre os três detectores propostos. Nesse caso, os detectores ADWIN, CUSUM, FTDD, WSTD, também foram competitivos. Esse classificador também fica no topo do ranqueamento no gráfico de CD $ACC_{NB(G)}$ da Figura 47 do Apêndice A, na configuração com mudanças graduais, considerando o NB. Nesse caso, nenhum dos detectores propostos foi competitivo. Entretanto, os detectores ADWIN, FTDD, PHT, SEED, e SEQDR2 não apresentaram diferenças estatísticas significativas em relação ao *Classificador Temporalmente Aumentado*.

Com relação aos resultados experimentais com classes desbalanceadas ($Agrawal_{(D)}$), a Tabela 14 e a Tabela 27 do Apêndice A mostram que o *Classificador Temporalmente Aumentado* também obteve desempenho superior em relação ao *Classificador de Classe Majoritária*, e a ambos os classificadores HT e NB usados sem detectores, considerando tanto mudanças abruptas quanto graduais.

Nesse cenário com classes desbalanceadas ($Agrawal_{(D)}$), o gráfico de CD $ACC_{HT(A)}$ na Figura 26 mostra que o SEDD obteve a melhor acurácia, não sendo estatisticamente diferente do RQADD usando o HT e considerando mudanças abruptas. No caso com mudança gradual, o RQADD apresentou a melhor acurácia, como mostrado no gráfico de CD $ACC_{HT(A)}$ da Figura 27, não sendo estatisticamente diferente do SEDD. Nessas configurações, os detectores CUSUM, DDM, FHDDM, FTDD, HDDMA, MDDMA, PHT, e RDDM também foram competitivos com ambos os classificadores HT e NB. Além desses, e o WSTD também foram competitivos com o HT.

Em relação ao classificador NB, o FTDD obteve a melhor acurácia em mudanças abruptas como apresentado no gráfico de CD $ACC_{NB(A)}$ da Figura 48 do Apêndice A, não sendo estatisticamente diferente do SEDD. Nessa mesma configuração, os detectores ADWIN, CUSUM, FTDD, HDDMA, RDDM, SEED, e WSTD também não apresentaram diferenças estatísticas significativas. No caso de mudanças graduais com esse mesmo classificador, o gráfico de CD $ACC_{NB(G)}$ da Figura 49 do Apêndice A mostra que o PHT obteve a melhor acurácia, mas não foi estatisticamente diferente do RQADD e do SEDD. Nesse caso, o CUSUM, DDM, FTDD, HDDMA, PHT, e RDDM também foram competitivos em relação ao *Classificador Temporalmente Aumentado*.

Em relação ao MTD usando o HT com classes balanceadas ($Agrawal_{(B)}$) e mudanças abruptas, o STEPD apresentou o menor MTD como mostrado nos gráficos de CD $\overline{MTD}_{HT(A)}$ (Figura 24), em que apenas o SEDD dos três detectores propostos não foi estatisticamente diferente do STEPD. Entre os demais detectores, o FTDD, HDDMA, MDDMA, SEED, e

STEPD, foram competitivos em relação ao STEPD. Na configuração com mudanças graduais, o SEED obteve o melhor MTD, e o PEDD foi o único dos três detectores propostos que não apresentou diferenças estatísticas significativas em relação ao SEED, como mostrado nos gráficos de CD $\overline{MTD}_{HT(G)}$ (Figura 25). Nesse caso, os detectores ADWIN, DDM, ECDD, EDDM, e PHT também foram competitivos.

Com relação ao MTD, usando o NB com classes balanceadas ($Agrawal_{(B)}$) e mudanças abruptas, o WSTD apresentou o menor MTD, não sendo estatisticamente diferente dos detectores ADWIN, FTDD, MDDMA, e STEPD, como mostrado nos gráficos de CD $\overline{MTD}_{NB(A)}$ (Figura 46 do Apêndice A). Nessa configuração, apenas o SEED entre os três detectores propostos foi competitivo em relação ao WSTD. Analisando a configuração com mudanças graduais, o SEED obteve o melhor MTD. Nesse caso, o PEDD foi o único dos três detectores propostos que não apresentou diferenças estatísticas significativas em relação ao SEED, como mostrado no gráfico de CD $\overline{MTD}_{NB(G)}$ (Figura 47 do Apêndice A). Essa figura também indica que os detectores ADWIN, ECDD, EDDM, FTDD, e PHT foram competitivos em relação ao SEED.

Nas bases com classes desbalanceadas ($Agrawal_{(D)}$) e mudanças abruptas usando o HT, o WSTD apresentou o melhor desempenho como mostrado nos gráfico de CD $\overline{MTD}_{HT(A)}$ (Figura 26). Nessa configuração, nenhum dos detectores propostos foi competitivo em relação ao WSTD. Entre os demais detectores, o FHDDM, FTDD, HDDMA, MDDMA, e STEPD também foram competitivos. Nas bases com mudanças graduais, o SEED obteve o melhor desempenho com o HT. Nesse caso, o PEDD foi o único dos três detectores propostos que não apresentou diferenças estatísticas significativas em relação ao SEED, como mostrado no gráfico de CD $\overline{MTD}_{HT(G)}$ (Figura 27). Entre os demais detectores, apenas o ADWIN, ECDD, EDDM, e SEQDR2 não apresentaram diferenças estatísticas significativas em relação ao SEED.

Nas bases com classes desbalanceadas ($Agrawal_{(D)}$) e mudanças abruptas usando o NB, o STEPD apresentou o melhor desempenho como indicado nos gráficos de CD $\overline{MTD}_{NB(A)}$ (Figura 48 do Apêndice A). Nessa configuração, apenas o SEED dos três detectores propostos não foi estatisticamente diferente do STEPD, e entre os demais detectores, o ADWIN, ECDD, FHDDM, FTDD, MDDMA, SEED, e WSTD também foram competitivos. Nas bases com mudança graduais, o SEED também obteve o melhor desempenho com o HT. Nesse caso, o PEDD também foi o único dos três detectores propostos que não apresentou diferenças estatísticas significativas em relação ao SEED, como mostrado no gráfico de CD $\overline{MTD}_{NB(G)}$ (Figura 49 do Apêndice A). Nessa configuração, os detectores ADWIN, ECDD, e EDDM também foram

competitivos.

Analisando o MTFA nas bases com classes balanceadas ($Agrawal_{(B)}$) e mudanças abruptas, o PHT apresentou o maior MTFA nas avaliações com o HT, como mostrado nos gráficos de CD $MTFA_{HT(A)}$ (Figura 24). Nas mudanças graduais, o gráfico de CD $MTFA_{HT(G)}$ (Figura 25) mostra que o PHT também obteve o maior MTFA. Nesse caso, os detectores CUSUM, FHDDM, FTDD, HDDMA, e WSTD foram competitivos em relação ao PHT, tanto em mudanças abruptas quanto nas graduais. O MDDMA foi competitivo apenas em mudanças abruptas, e somente o SEDD e o RQADD dos três detectores propostos não foram estatisticamente diferentes do PHT.

Analisando o MTFA nas bases com classes balanceadas ($Agrawal_{(B)}$) e mudanças abruptas, o PHT apresentou o maior MTFA nas avaliações com o NB, como indicado no gráfico de CD $MTFA_{NB(A)}$ (Figura 46 do Apêndice A).

Nas mudanças graduais, o gráfico de CD $MTFA_{NB(G)}$ (Figura 47 do Apêndice A) mostra que o PHT também obteve o maior MTFA. Nessas configurações, os detectores CUSUM, DDM, FTDD, HDDMA, e WSTD não apresentaram diferenças estatísticas significativas em relação ao PHT, considerando ambas as mudanças abruptas e graduais. Nessas configurações, apenas o SEDD e o RQADD foram competitivos entre os três detectores propostos.

Nas bases com classes desbalanceadas ($Agrawal_{(D)}$) e mudanças abruptas usando HT, o CUSUM apresentou o melhor desempenho como mostrado no gráfico de CD $MTFA_{HT(A)}$ (Figura 26). Nessa configuração, apenas o SEDD e o RQADD foram competitivos entre os detectores propostos. Também foram competitivos os detectores FHDDM, FTDD, HDDMA, MDDMA, PHT, SEQDR2, e WSTD. Nas bases com mudanças graduais, o gráfico de CD $MTFA_{HT(G)}$ da Figura 27 indica que o RQADD obteve o melhor desempenho nessa configuração, não apresentando diferença estatística significativa em relação aos detectores CUSUM, DDM, FHDDM, FTDD, HDDMA, MDDMA, PHT, e WSTD. Nesse caso, apenas o SEDD foi competitivo entre os detectores propostos.

Nas bases com classes desbalanceadas ($Agrawal_{(D)}$) e mudanças abruptas usando o NB, o PHT apresentou o melhor desempenho como mostrado no gráfico de CD $MTFA_{NB(A)}$ (Figura 48 do Apêndice A), onde apenas o SEDD e o RQADD foram competitivos entre os três detectores propostos. Na configuração com mudança gradual, os resultados foram semelhantes aos de mudanças abruptas. Nesse caso, o PHT apresentou o melhor desempenho como mostrado no gráfico de CD $MTFA_{NB(G)}$ (Figura 49 do Apêndice A), onde apenas o SEDD e o RQADD foram competitivos entre os três detectores propostos. Tanto nas configurações com

mudanças abruptas, quanto nas graduais, os detectores CUSUM, FTDD, HDDMA, WSTD foram competitivos em relação ao PHT.

Em relação ao MDR com a base $Agrawal_{(B)}$ e mudanças abruptas usando o HT, o ADWIN obteve a menor taxa de detecções perdidas como aponta o gráfico de CD $\overline{MDR}_{HT(A)}$ (Figura 24). Nessa configuração, os três detectores propostos não apresentam diferenças estatísticas significativas em relação ao ADWIN, e apenas o FHDDM não foi competitivo. Nas mudanças graduais, o gráfico de CD $\overline{MDR}_{HT(G)}$ (Figura 25) mostra que o ADWIN também obteve a menor taxa de detecções perdidas, onde apenas o PEDD foi competitivo entre os três detectores propostos. Nesse caso, os detectores ECDD, EDDM, RDDM, SEED, SEQDR2, e STEPD, também foram competitivos.

No tocante ao MDR, nas bases com classes balanceadas ($Agrawal_{(B)}$) e mudanças abruptas usando o NB, o ADWIN também obteve a menor taxa de detecções perdidas como aponta o gráfico de CD $\overline{MDR}_{NB(A)}$ (Figura 46 do Apêndice A). Nessa configuração, os três detectores propostos também não apresentam diferenças estatísticas significativas em relação ao ADWIN, como também nenhum dos demais detectores. Note que a hipótese nula do teste F_F não foi rejeitada nessa avaliação ($P > 0,05$).

No caso da configuração com mudanças graduais, o gráfico de CD $\overline{MRD}_{NB(G)}$ (Figura 47 do Apêndice A) mostra que o ADWIN também obteve o melhor resultado com o MDR. Vale salientar que nessa configuração apenas o PEDD foi competitivo entre os três detectores propostos, em que o ECDD, EDDM, SEED, e o SEQDR2 também não apresentaram diferenças estatísticas significativas em relação ao ADWIN.

Nas bases com classes desbalanceadas ($Agrawal_{(D)}$) e mudanças abruptas usando HT, nenhum dos detectores avaliados apresentaram diferenças estatísticas significativas, como apresentados no gráfico de CD $\overline{MDR}_{HT(A)}$ (Figura 26), onde $P > 0,05$, ou seja, a hipótese nula do teste F_F não foi rejeitada. Com relação às bases com mudanças graduais, o ADWIN obteve a menor taxa de detecções perdidas como apresentados nos gráficos de CD $\overline{MDR}_{HT(G)}$ (Figura 27), em que apenas o PEDD foi competitivo entre três detectores propostos. Entre os demais detectores, apenas o ECDD, EDDM, RDDM, SEED, SEQDR2, e STEPD foram competitivos.

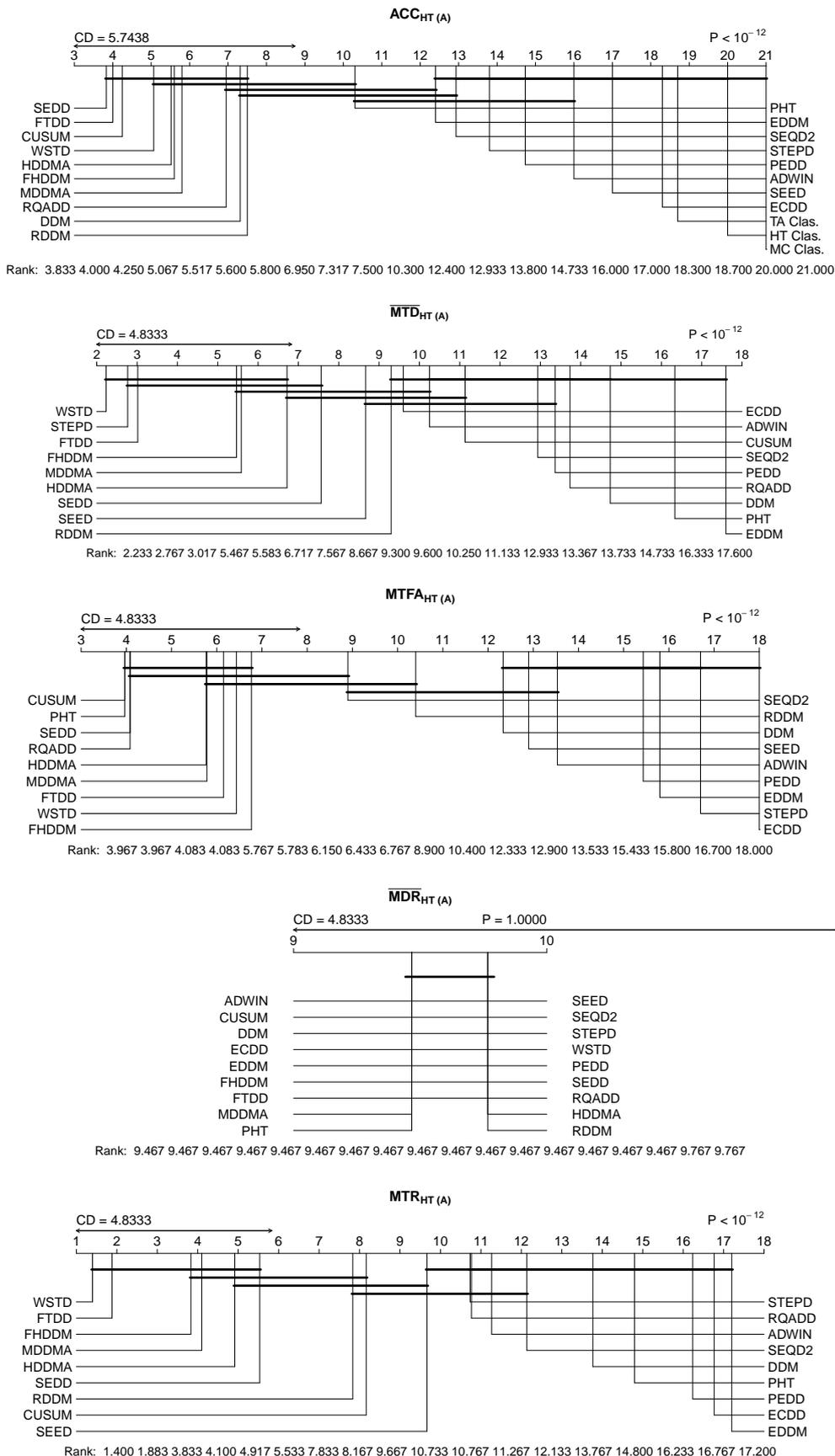
Nas bases com classes desbalanceadas ($Agrawal_{(D)}$) e mudanças abruptas usando NB, nenhum dos detectores avaliados apresentaram diferenças estatísticas significativas, como indicado no gráfico de CD $\overline{MDR}_{NB(A)}$ (Figura 48 do Apêndice A). Nesse caso, a hipótese nula do teste F_F não foi rejeitada ($P > 0,05$). Com relação às bases com mudanças graduais, o

Tabela 14 – Resultados obtidos usando HT, considerando a base artificial Agrawal desbalanceada, com um intervalo de confiança de 95%.

<i>Agrawal</i> _(D)	<i>Classificador/Detector</i> _(HT)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Hoeffding Tree	77,16 ± 0,33					
	Majority Class	66,94 ± 0,04					
	Temporally Augmented Classifier	86,08 ± 0,35					
	ADWIN	89,06 ± 0,07	113,60 ± 33,85	18438,66 ± 712,95	0,00 ± 0,00	220,78 ± 34,10	
	CUSUM	94,05 ± 0,06	112,78 ± 3,44	51200,00 ± 0,00	0,00 ± 0,00	456,94 ± 13,97	
	DDM	93,92 ± 0,07	251,60 ± 6,33	26522,76 ± 2915,78	0,00 ± 0,00	106,58 ± 12,46	
	ECDD	86,51 ± 0,04	1562,54 ± 1475,74	602,72 ± 8,74	0,00 ± 0,00	26,05 ± 13,01	
	EDDM	92,25 ± 0,19	2410,26 ± 125,86	6881,05 ± 1472,23	0,00 ± 0,00	2,92 ± 0,66	
	FHDDM	93,89 ± 0,12	35,80 ± 6,90	46357,17 ± 1841,00	0,00 ± 0,00	1560,83 ± 224,93	
	FTDD	94,01 ± 0,09	20,55 ± 1,74	47781,40 ± 1774,90	0,00 ± 0,00	2418,41 ± 194,99	
	HDDMA	93,89 ± 0,14	472,48 ± 871,51	48264,09 ± 1594,90	0,01 ± 0,02	1258,62 ± 263,62	
	MDDMA	93,81 ± 0,20	37,76 ± 2,25	47802,20 ± 1701,58	0,00 ± 0,00	1299,60 ± 96,84	
	PHT	93,54 ± 0,08	764,83 ± 26,31	51200,00 ± 0,00	0,00 ± 0,00	67,51 ± 2,39	
	RDDM	93,85 ± 0,09	496,95 ± 870,55	35726,87 ± 2714,38	0,01 ± 0,02	512,18 ± 64,10	
	SEED	88,44 ± 0,06	923,20 ± 1209,32	20781,49 ± 1012,05	0,00 ± 0,00	341,61 ± 60,87	
	SEQDRIFT2	92,02 ± 0,15	215,00 ± 15,62	40839,55 ± 1474,73	0,00 ± 0,00	194,53 ± 11,70	
	STEPD	91,13 ± 0,41	430,70 ± 641,36	4656,03 ± 1048,56	0,00 ± 0,00	328,98 ± 114,93	
	WSTD	93,91 ± 0,14	19,84 ± 2,01	47223,37 ± 1992,37	0,00 ± 0,00	2522,85 ± 240,51	
	PEDD	90,03 ± 0,12	746,15 ± 473,95	5566,48 ± 215,07	0,00 ± 0,00	30,61 ± 7,73	
	RQADD	93,93 ± 0,07	224,18 ± 3,11	50929,71 ± 552,81	0,00 ± 0,00	227,47 ± 3,90	
	SEDD	94,05 ± 0,05	56,03 ± 2,17	50895,01 ± 623,78	0,00 ± 0,00	918,08 ± 38,52	
	Gradual	Hoeffding Tree	76,76 ± 0,30				
		Majority Class	66,95 ± 0,04				
		Temporally Augmented Classifier	84,96 ± 0,28				
		ADWIN	88,42 ± 0,06	1306,67 ± 882,94	6459,22 ± 674,16	0,00 ± 0,00	7,87 ± 1,33
		CUSUM	93,24 ± 0,06	38090,24 ± 20,50	50947,61 ± 16,13	0,25 ± 0,00	1,00 ± 0,00
		DDM	93,28 ± 0,07	28021,12 ± 2881,04	49017,55 ± 1048,62	0,23 ± 0,02	1,48 ± 0,24
ECDD		85,99 ± 0,05	2366,78 ± 786,87	638,70 ± 17,58	0,00 ± 0,00	0,92 ± 0,42	
EDDM		92,17 ± 0,19	1919,68 ± 77,54	13125,69 ± 2222,22	0,00 ± 0,00	6,91 ± 1,17	
FHDDM		93,12 ± 0,12	13536,92 ± 3837,32	46919,45 ± 1659,48	0,23 ± 0,06	28,19 ± 18,63	
FTDD		93,24 ± 0,09	39983,96 ± 4258,65	49558,06 ± 1030,90	0,29 ± 0,04	0,99 ± 0,18	
HDDMA		93,06 ± 0,14	35016,20 ± 2356,12	48592,22 ± 1181,15	0,22 ± 0,03	1,16 ± 0,21	
MDDMA		93,01 ± 0,21	42568,44 ± 2248,39	48635,67 ± 1235,67	0,28 ± 0,03	0,83 ± 0,05	
PHT		93,16 ± 0,06	11683,15 ± 1634,01	51129,77 ± 8,58	0,22 ± 0,03	12,11 ± 9,00	
RDDM		92,97 ± 0,09	32786,78 ± 2621,59	23224,17 ± 1738,54	0,14 ± 0,05	0,67 ± 0,15	
SEED		87,81 ± 0,06	1124,27 ± 869,88	5084,78 ± 729,81	0,00 ± 0,00	8,69 ± 2,86	
SEQDRIFT2		91,25 ± 0,15	7220,00 ± 2681,62	32548,54 ± 1979,98	0,00 ± 0,00	23,96 ± 10,09	
STEPD		90,73 ± 0,40	16106,30 ± 4475,49	16021,14 ± 2520,76	0,13 ± 0,05	2,35 ± 1,43	
WSTD		89,12 ± 1,02	19348,49 ± 3560,48	48466,76 ± 1463,78	0,30 ± 0,07	4,12 ± 2,63	
PEDD		89,38 ± 0,12	5167,12 ± 709,88	5804,48 ± 217,49	0,00 ± 0,00	1,34 ± 0,27	
RQADD		93,29 ± 0,06	24805,21 ± 4497,89	50967,50 ± 349,73	0,37 ± 0,05	24,40 ± 46,59	
SEDD		93,25 ± 0,06	47130,57 ± 2508,36	50898,80 ± 345,53	0,27 ± 0,02	0,80 ± 0,04	

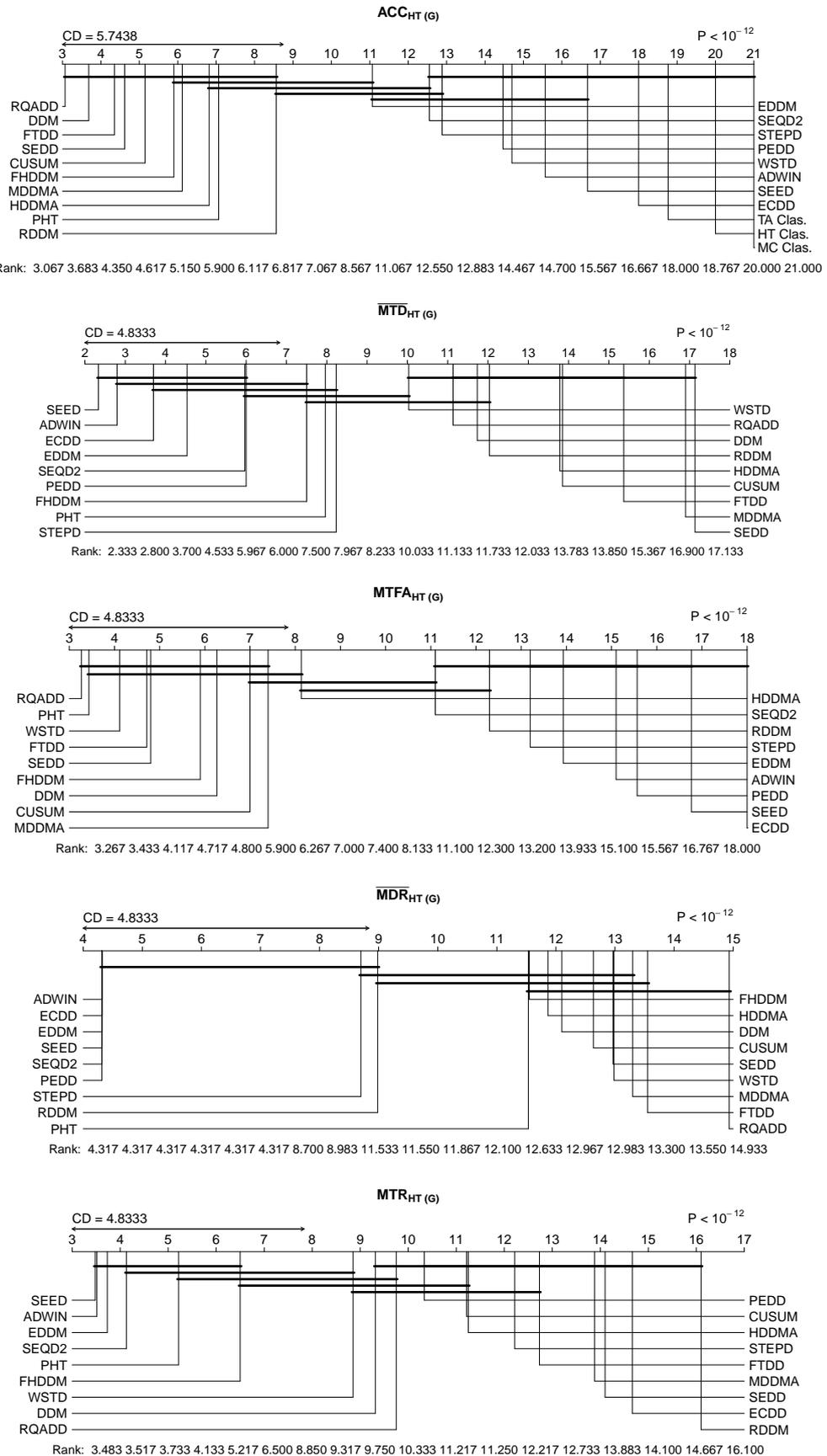
Fonte: Elaborada pelo autor (2021).

Figura 26 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base Agrawal desbalanceada, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 27 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal desbalanceada, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

ADWIN obteve a menor taxa de detecções perdidas como apresentados nos gráficos de CD $\overline{MDR}_{NB(G)}$ (Figura 49 do Apêndice A), em que também o PEDD foi o único competitivo entre três detectores propostos. Nessa configuração, os detectores ECDD, EDDM, RDDM, SEED, SEQDR2, e STEPD também foram competitivos.

Com relação ao MTR com classes balanceadas ($Agrawal_{(B)}$) usando o HT, o MDDMA obteve o melhor desempenho nessa métrica com mudanças abruptas, não sendo estatisticamente diferente do SEDD como mostrado no gráfico de CD $MTR_{HT(A)}$ (Figura 24). Nessa configuração, o PEDD e o RQADD não foram competitivos, e os detectores CUSUM, FTDD, HDDMA, RDDM, STEPD, e WSTD não apresentaram diferenças estatísticas significativas em relação ao MDDMA. Nas mudanças graduais, o gráfico de CD $MTR_{HT(G)}$ (Figura 25) mostra que o PHT obteve o melhor resultado com essa métrica, em que os detectores ADWIN, DDM, EDDM, SEQDR2, e SEED não apresentaram diferenças estatísticas significativas em relação ao PHT. Nesse caso, nenhum dos detectores propostos foi competitivo.

No tocante às bases com classes balanceadas ($Agrawal_{(B)}$) usando NB, o WSTD foi o melhor entre todos os detectores com relação ao MTR e o SEDD foi o único entre os detectores propostos que foi competitivo como apresentado no gráfico de CD $\overline{MTR}_{NB(A)}$ (Figura 46 do Apêndice A). Entre os demais detectores, apenas o ADWIN, FTDD, e MDDMA foram competitivos. No caso da configuração com mudanças graduais, o gráfico de CD $MTR_{NB(G)}$ (Figura 47 do Apêndice A) mostra que o ADWIN obteve o melhor resultado com o MTR. Nessa configuração, os detectores FHDDM, FTDD, PHT, SEQDR2, e SEED não apresentaram diferenças estatísticas significativas em relação ao ADWIN. Vale salientar que nessa configuração os detectores propostos não foram competitivos.

Nas bases com classes desbalanceadas ($Agrawal_{(D)}$) usando o HT, o WSTD obteve o melhor MTR com mudanças abruptas, não apresentando diferenças estatísticas significativas com os detectores FHDDM, FTDD, HDDMA, e MDDMA, em que apenas o SEDD entre os detectores propostos foi competitivo, como mostrado nos gráfico de CD $MTR_{HT(A)}$ (Figura 26). Na configuração com mudança gradual, o gráfico de CD $MTR_{HT(G)}$ (Figura 27) indica que o SEED obteve o melhor MTR com o HT. Nesse caso, nenhum dos detectores propostos foi competitivo. Por outro lado, o SEED não apresentou diferença estatística significativa em relação ao ADWIN, EDDM, FHDDM, PHT, e ao SEQDR2.

Nas configurações com classes desbalanceadas ($Agrawal_{(D)}$) usando NB, o FTDD foi o melhor entre todos os detectores com relação ao MTR e o SEDD foi o único entre os detectores propostos que foi competitivo como apresentado no gráfico de CD $MTR_{NB(A)}$

(Figura 48 do Apêndice A). Os demais detectores competitivos foram: ADWIN, FHDDM, HDDMA, MDDMA, SEED, e WSTD. No caso da configuração com mudanças graduais, o gráfico de CD $MTR_{NB(G)}$ (Figura 49 do Apêndice A) mostra que o ADWIN obteve o melhor resultado com o MTR. Nesse caso, apenas o PEDD foi competitivo entre os detectores propostos, e os detectores SEED, SEQDR2, e WSTD não apresentaram diferenças estatísticas significativas em relação ao ADWIN.

5.3.1.2 Resultados e análises considerando 5% e 10% de ruído

Os resultados experimentais com classes com 5% de ruído ($LED(5\%)$) mostram que o *Classificador Temporalmente Aumentado* obteve acurácia superior em relação ao *Classificador de Classe Majoritária*, mas não foi melhor que o classificador HT isolado. Além disso, não apresentou acurácia superior a maioria dos detectores, considerando tanto mudanças abruptas quanto graduais, como apresentado na Tabela 15 e 28 do Apêndice A.

A Tabela 16 e 29 do Apêndice A mostra que esse classificador também não foi melhor com 10% de ruído ($LED(10\%)$), em termos de acurácia, que a maioria dos detectores usando o NB, mas foi superior em relação ao *Classificador de Classe Majoritária*, e ao classificador NB isolado. Um ponto a destacar é que nesse caso a dependência temporal dos rótulos foi menos significativa para a acurácia que a mudança de conceito.

A acurácia dos classificadores e detectores usando o HT nas bases com 5% de ruído ($LED(5\%)$), considerando mudanças abruptas, é indicada no gráfico de CD $ACC_{HT(A)}$ (Figura 28). Nessa configuração, o MDDMA obteve a melhor acurácia, não apresentando diferença estatística significativa em relação ao CUSUM, FHDDM, HDDMA, e RDDM. No gráfico de CD $ACC_{HT(G)}$ (Figura 29), referente aos testes estatísticos obtidos da configuração com mudanças graduais, o RDDM obteve vantagem na performance em relação aos demais detectores, não apresentaram diferenças estatísticas significativas em relação ao CUSUM, FHDDM, HDDMA, e MDDMA, usando o HT. Em ambas as configurações com bases abruptas e graduais, apenas o SEDD e o RQADD foram competitivos entre os detectores propostos.

Ainda na configuração com a base de dados $LED(5\%)$ considerando o NB, o MDDMA fica no topo do ranqueamento com mudanças abruptas, não apresentando diferença estatística significativa em relação ao HDDMA, como indicado no gráfico de CD $ACC_{NB(A)}$ da Figura 50 do Apêndice A. Na configuração com mudanças graduais considerando o NB, o HDDMA fica no topo do ranqueamento no gráfico de CD $ACC_{NB(G)}$ da Figura 51 do Apêndice A, não

apresentando diferença estatística significativa em relação aos detectores MDDMA e PHT. Em ambas as configurações com bases abruptas e graduais, apenas o SEDD e o RQADD foram competitivos entre os detectores propostos. Nessas bases, os detectores CUSUM, RDDM e DDM também foram competitivos.

Com relação aos resultados experimentais com 10% de ruído ($LED_{(10\%)}$), a Tabela 16 e a Tabela 29 do Apêndice A mostram que o *Classificador Temporalmente Aumentado* também obteve desempenho superior em relação ao *Classificador de Classe Majoritária* e ao classificador NB sem os detectores, considerando tanto mudanças abruptas quanto graduais. Entretanto, o HT usado isoladamente e os detectores apresentaram uma acurácia superior ao *Classificador Temporalmente Aumentado* com essa configuração.

No cenário com 10% de ruído ($LED_{(10\%)}$), o gráfico de CD $ACC_{HT(A)}$ na Figura 30 mostra que o HDDMA obteve a melhor acurácia usando o HT, não apresentando diferença estatística significativa em relação ao RDDM. No caso com mudança gradual, o RDDM apresentou a melhor acurácia, não apresentando diferença estatística significativa em relação ao HDDMA, como mostrado no gráfico de CD $ACC_{HT(A)}$ da Figura 31. Entre os demais detectores, o CUSUM, MDDMA, e o PHT também foram competitivos em ambas as bases abruptas e graduais. Nessas mesmas configurações, os detectores propostos PEDD, RQADD, e SEDD também foram competitivos.

Em relação ao classificador NB, o HDDMA também obteve a melhor acurácia em mudanças abruptas como apresentado no gráfico de CD $ACC_{NB(A)}$ da Figura 52 do Apêndice A. Nesse caso, os detectores CUSUM, PHT, e RDDM também foram competitivos. No caso de mudanças graduais usando esse mesmo classificador, o gráfico de CD $ACC_{NB(G)}$ da Figura 53 do Apêndice A mostra que o RDDM obteve novamente a melhor acurácia, o qual não apresentou diferenças estatísticas significativas em relação aos detectores CUSUM, HDDMA, e PHT. Em ambas as configurações com bases abruptas e graduais, os três detectores propostos (PEDD, RQADD, e SEDD) também foram competitivos.

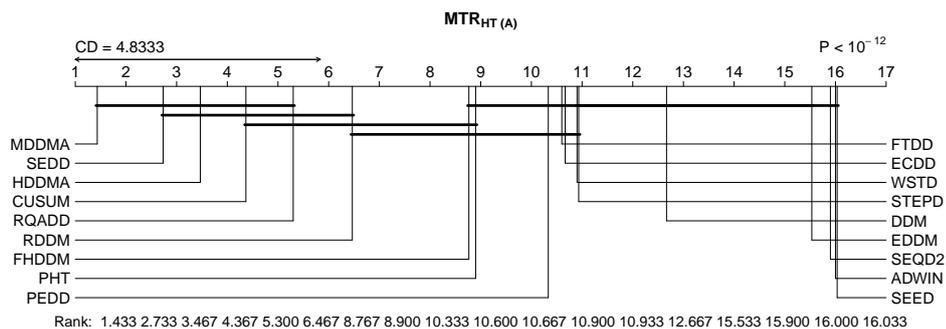
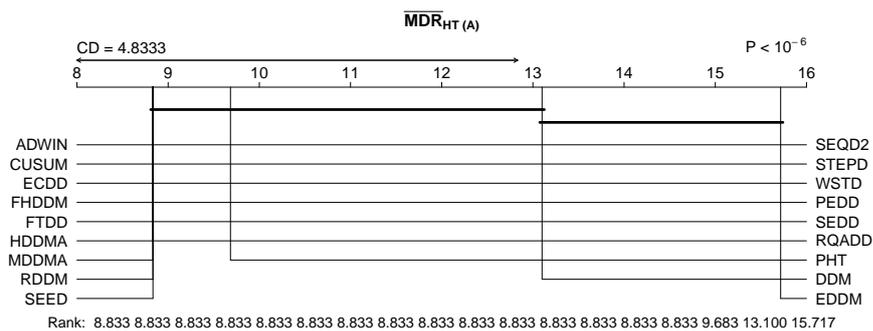
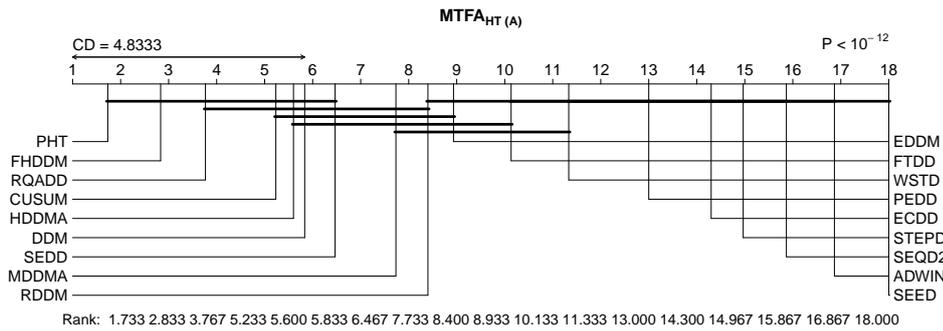
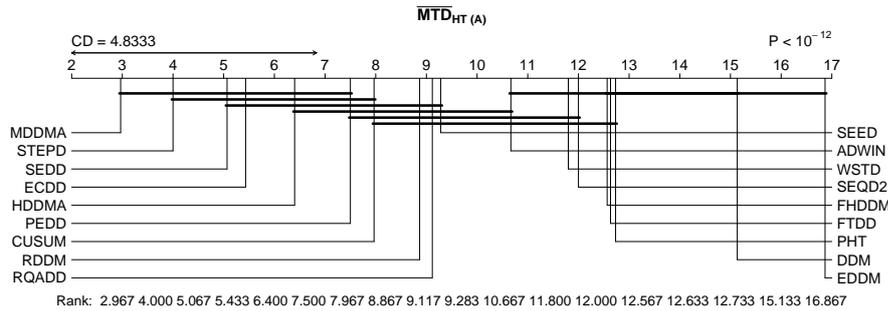
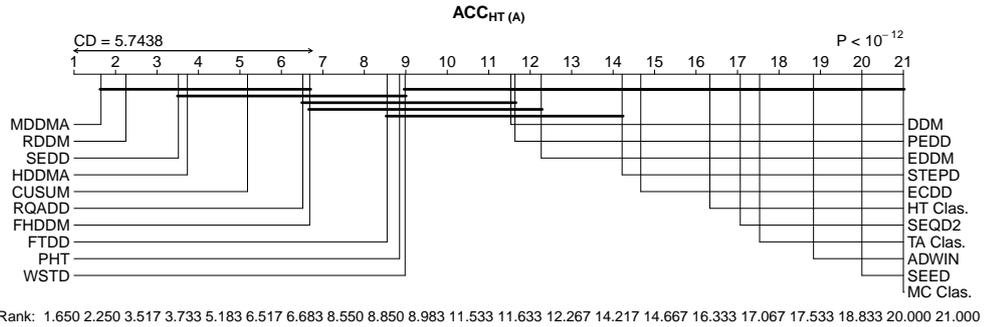
Em relação ao MTD usando o HT com 5% de ruído ($LED_{(5\%)}$) e mudanças abruptas, o MDDMA apresentou o menor MTD como mostrado nos gráficos de CD $\overline{MTD}_{HT(A)}$ (Figura 28), em que apenas o SEDD e o PEDD dos três detectores propostos não foi estatisticamente diferente do MDDMA. Nesse caso, os detectores ECDD, HDDMA, e STEPDP também foram competitivos. Na configuração com mudanças graduais, o ECDD obteve o melhor MTD, em que os três detectores propostos (PEDD, RQADD, e SEDD) foram competitivos, como mostrado no gráfico de CD $\overline{MTD}_{HT(G)}$ (Figura 29). Entre os demais detectores, o ADWIN,

Tabela 15 – Resultados obtidos usando HT, considerando a base artificial LED tendo 5% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.

$LED_{(5\%)}$	Classificador/Detector _(HT)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Hoeffding Tree	80,75 ± 0,37					
	Majority Class	10,11 ± 0,05					
	Temporally Augmented Classifier	80,23 ± 0,38					
	ADWIN	74,26 ± 0,88	3859,20 ± 3033,45	405,12 ± 64,96	0,00 ± 0,00	0,96 ± 0,32	
	CUSUM	86,30 ± 0,06	270,08 ± 18,91	35873,29 ± 126,86	0,00 ± 0,00	136,96 ± 8,66	
	DDM	84,73 ± 0,55	14429,76 ± 5677,00	35109,24 ± 3650,97	0,22 ± 0,10	12,28 ± 7,31	
	ECDD	82,93 ± 0,13	556,81 ± 306,72	1442,84 ± 45,29	0,00 ± 0,00	26,16 ± 13,56	
	EDDM	85,07 ± 0,14	19822,14 ± 3831,57	26141,41 ± 3895,22	0,29 ± 0,07	1,21 ± 0,35	
	FHDDM	86,07 ± 0,09	3133,01 ± 1021,62	37618,86 ± 520,80	0,00 ± 0,00	53,31 ± 35,79	
	FTDD	85,93 ± 0,11	3713,76 ± 1682,50	20677,22 ± 2291,76	0,00 ± 0,00	30,65 ± 19,01	
	HDDMA	86,33 ± 0,06	252,17 ± 62,67	34720,16 ± 1961,58	0,00 ± 0,00	195,51 ± 39,40	
	MDDMA	86,40 ± 0,06	675,29 ± 1199,28	35522,09 ± 248,93	0,00 ± 0,00	448,01 ± 68,83	
	PHT	85,91 ± 0,06	2584,36 ± 1414,99	40215,43 ± 1162,66	0,02 ± 0,03	29,04 ± 4,93	
	RDDM	86,39 ± 0,06	1163,73 ± 1012,39	26863,76 ± 3430,16	0,00 ± 0,00	95,59 ± 19,90	
	SEED	63,34 ± 0,12	9539,47 ± 6242,14	187,72 ± 1,03	0,00 ± 0,00	0,89 ± 0,29	
	SEQDRIFT2	78,80 ± 1,18	16266,67 ± 7772,23	644,25 ± 151,53	0,00 ± 0,00	1,14 ± 0,41	
	STEPD	83,45 ± 0,32	938,18 ± 1168,53	1186,29 ± 171,08	0,00 ± 0,00	18,56 ± 5,43	
	WSTD	85,85 ± 0,15	3573,47 ± 1066,84	15239,55 ± 2083,56	0,00 ± 0,00	36,52 ± 22,11	
	PEDD	85,29 ± 0,09	1000,86 ± 779,76	4121,59 ± 215,97	0,00 ± 0,00	22,32 ± 6,39	
	RQADD	86,22 ± 0,07	456,25 ± 172,02	36310,36 ± 249,30	0,00 ± 0,00	118,54 ± 17,82	
	SEDD	86,35 ± 0,06	194,67 ± 69,18	35324,60 ± 997,92	0,00 ± 0,00	244,59 ± 35,85	
	Gradual	Hoeffding Tree	80,63 ± 0,31				
		Majority Class	10,11 ± 0,05				
		Temporally Augmented Classifier	80,08 ± 0,31				
		ADWIN	74,32 ± 0,52	2524,53 ± 1593,21	426,57 ± 39,40	0,00 ± 0,00	0,69 ± 0,27
		CUSUM	86,11 ± 0,05	21379,78 ± 3117,57	40956,72 ± 1299,84	0,00 ± 0,00	2,22 ± 0,33
		DDM	85,04 ± 0,54	15170,35 ± 5097,34	34737,25 ± 3179,61	0,22 ± 0,10	13,31 ± 11,40
		ECDD	82,70 ± 0,11	914,92 ± 198,61	1440,90 ± 43,98	0,00 ± 0,00	2,28 ± 0,80
		EDDM	85,14 ± 0,12	17714,08 ± 3475,40	25698,55 ± 3304,98	0,26 ± 0,07	1,60 ± 0,64
		FHDDM	85,70 ± 0,08	6036,53 ± 1516,80	39290,31 ± 729,56	0,00 ± 0,00	10,31 ± 3,68
FTDD		85,52 ± 0,10	7145,63 ± 1427,23	21710,56 ± 2117,87	0,00 ± 0,00	5,27 ± 2,44	
HDDMA		86,17 ± 0,06	9150,03 ± 2969,38	36507,68 ± 1341,36	0,02 ± 0,02	23,18 ± 14,17	
MDDMA		86,15 ± 0,05	9598,75 ± 3184,27	37471,01 ± 711,97	0,00 ± 0,00	17,74 ± 9,06	
PHT		85,84 ± 0,05	2434,51 ± 1310,54	40947,70 ± 1402,81	0,02 ± 0,02	34,74 ± 9,80	
RDDM		86,19 ± 0,06	18784,87 ± 3054,83	31440,38 ± 1862,85	0,00 ± 0,00	2,06 ± 0,40	
SEED		63,21 ± 0,10	6937,60 ± 2999,63	188,31 ± 0,90	0,00 ± 0,00	0,93 ± 0,38	
SEQDRIFT2		79,05 ± 0,79	17711,67 ± 4840,50	1001,82 ± 222,78	0,00 ± 0,00	0,38 ± 0,29	
STEPD		83,35 ± 0,32	4344,43 ± 1185,69	1312,07 ± 188,03	0,00 ± 0,00	0,87 ± 0,58	
WSTD		85,03 ± 0,16	10891,54 ± 1954,04	16306,84 ± 2100,00	0,01 ± 0,02	2,35 ± 1,44	
PEDD		85,09 ± 0,07	3706,92 ± 611,47	4224,73 ± 197,50	0,00 ± 0,00	1,65 ± 0,63	
RQADD		85,94 ± 0,06	2599,00 ± 1273,98	37372,77 ± 355,40	0,00 ± 0,00	23,22 ± 3,74	
SEDD		86,03 ± 0,07	4420,55 ± 2128,71	36819,43 ± 736,06	0,00 ± 0,00	24,00 ± 7,68	

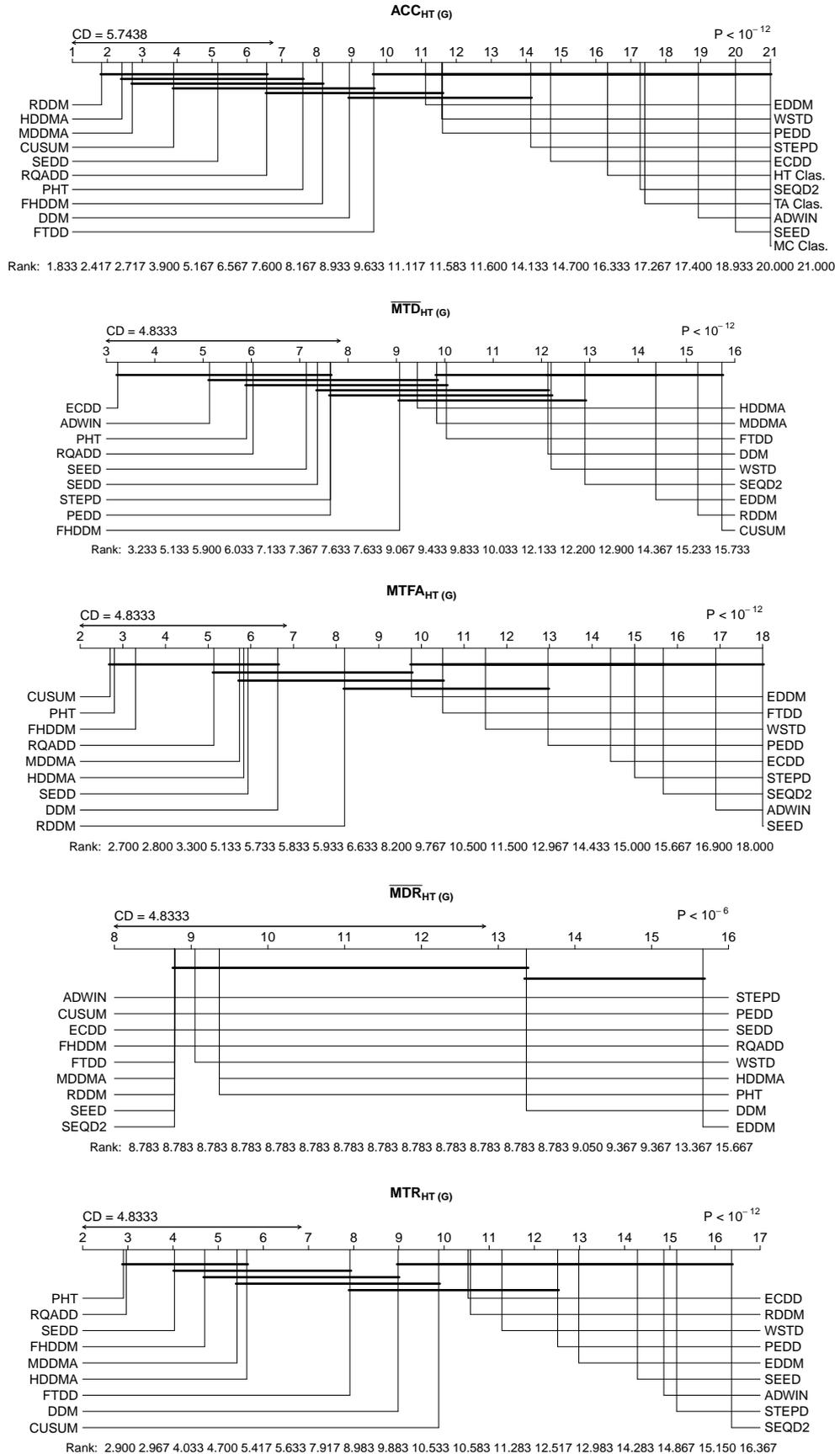
Fonte: Elaborada pelo autor (2021).

Figura 28 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 29 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

PHT, SEED, e STEPDP também foram competitivos nas bases com mudanças graduais.

Em relação ao MTD usando o NB com 5% de ruído ($LED_{(5\%)}$) e mudanças abruptas, o MDDMA apresentou o menor MTD como mostrado nos gráficos de CD $\overline{MTD}_{NB(A)}$ (Figura 50 do Apêndice A). Nessa configuração, apenas o SEDD, entre os três detectores propostos não foi estatisticamente diferente do MDDMA, onde os detectores ECDD e STEPDP também foram competitivos. Analisando a configuração com mudanças graduais, o ECDD obteve o melhor MTD. Dos detectores propostos, somente o PEDD não apresentou diferença estatística significativa em relação ao ECDD, como mostrado no gráfico de CD $\overline{MTD}_{NB(G)}$ (Figura 51 do Apêndice A). Entre os demais detectores, o ADWIN, DDM, PHT, SEED, e o STEPDP foram competitivos em relação às bases com mudanças graduais.

Nas bases com 10% de ruído ($LED_{(10\%)}$) e mudanças abruptas usando o HT, o MDDMA apresentou o melhor desempenho como mostrado nos gráfico de CD $\overline{MTD}_{HT(A)}$ (Figura 30). Nessa configuração, apenas o SEDD e o RQADD entre os detectores propostos foram competitivos em relação ao MDDMA, onde os detectores CUSUM, ECDD, e RDDM também foram competitivos. Nas bases com mudanças graduais, o ECDD obteve o melhor desempenho com o HT, não apresentando diferenças estatísticas significativas em relação ao ADWIN e ao PHT. Nesse caso, nenhum dos três detectores propostos foi competitivo em relação ao ECDD, como mostrado no gráfico de CD $\overline{MTD}_{HT(G)}$ (Figura 31). Nessas configurações, os detectores FHDDM, HDDMA, SEED, e STEPDP também foram competitivos em ambas as bases abruptas e graduais.

Nas bases com 10% de ruído ($LED_{(10\%)}$) e mudanças abruptas usando o NB, o MDDMA apresentou o melhor desempenho, não apresentando diferenças estatísticas significativas em relação aos detectores CUSUM, ECDD, e RDDM, como indicado no gráfico de CD $\overline{MTD}_{NB(A)}$ (Figura 52 do Apêndice A). Nessa configuração, apenas o RQADD e o SEDD, entre os três detectores propostos, foram competitivos em relação ao MDDMA. Nas bases com mudança graduais, o ECDD obteve o melhor desempenho com o HT, não apresentando diferenças estatísticas significativas em relação aos detectores PHT, e SEED. Nesse caso, nenhum dos detectores propostos foi competitivo, como mostrado no gráfico de CD $\overline{MTD}_{NB(G)}$ (Figura 53 do Apêndice A). Entre os demais detectores, o ADWIN, FHDDM, HDDMA, SEQDR2, e o STEPDP foram competitivos em ambas as configurações usando as bases com mudanças abruptas e graduais.

Analisando o MTFa nas bases com 5% de ruído ($LED_{(5\%)}$) e mudanças abruptas, o PHT apresentou o maior MTFa nas avaliações com o HT, não apresentando diferenças estatísticas

significativas em relação ao CUSUM, como mostrado nos gráficos de CD $MTFA_{HT(A)}$ (Figura 28). Nas mudanças graduais, o gráfico de CD $MTFA_{HT(G)}$ (Figura 29) mostra que o CUSUM também obteve o maior MTFA, não apresentando diferenças estatísticas significativas em relação ao PHT. Nessa configuração, o MDDMA também foi competitivo em relação ao CUSUM. Entre os três detectores propostos, apenas o SEDD e o RQADD foram competitivos, considerando ambas as bases abruptas e graduais. Entre os demais detectores, o DDM, FHDDM, e HDDMA também foram competitivos nessas bases.

Analisando o MTFA nas bases com 5% de ruído ($LED_{(5\%)}$) e mudanças abruptas, o FHDDM apresentou o maior MTFA nas avaliações com o NB, como indicado no gráfico de CD $MTFA_{NB(A)}$ (Figura 50 do Apêndice A). Nas mudanças graduais, o gráfico de CD $MTFA_{NB(G)}$ (Figura 51 do Apêndice A) mostra que o FHDDM também obteve o maior MTFA. Nessas configurações, o SEDD e o RQADD também foram competitivos entre os três detectores propostos. Entre os demais detectores, o CUSUM, DDM, FHDDM, HDDMA, MDDMA, e PHT foram competitivos em ambas as bases abruptas e graduais.

Nas bases com 10% de ruído ($LED_{(10\%)}$) e mudanças abruptas usando HT, o PHT apresentou o melhor desempenho como mostrado no gráfico de CD $MTFA_{HT(A)}$ (Figura 30). Nessa configuração, apenas o SEDD e o RQADD foram competitivos. Nas bases com mudanças graduais, o gráfico de CD $MTFA_{HT(G)}$ da Figura 31 indica que o PHT também obteve o melhor desempenho nessa configuração, em que todos os detectores propostos (PEDD, RQADD, e SEDD) foram competitivos. Entre os demais detectores, o CUSUM, DDM, EDDM, FTDD, HDDMA, e RDDM também não apresentaram diferenças estatísticas significativas em relação ao PHT, em ambas as bases com mudanças abruptas e graduais.

Nas bases com 10% de ruído ($LED_{(10\%)}$) e mudanças abruptas usando o NB, o PHT apresentou o melhor desempenho como indica o gráfico de CD $MTFA_{NB(A)}$ (Figura 52 do Apêndice A), onde o RDDM também foi competitivo. Na configuração com mudança gradual, o PHT apresentou novamente o melhor desempenho como mostrado no gráfico de CD $MTFA_{NB(G)}$ (Figura 53 do Apêndice A). Entre os três detectores apenas o SEDD e o RQADD foram competitivos nessas configurações, e entre os demais detectores, o CUSUM, DDM, FTDD, HDDMA também não apresentaram diferenças estatísticas significativas em relação ao PHT.

Em relação ao MDR nas bases com 5% de ruído ($LED_{(5\%)}$) e mudanças abruptas usando o HT, o ADWIN obteve a menor taxa de detecções perdidas como aponta o gráfico de CD $\overline{MDR}_{HT(A)}$ (Figura 28). Nas mudanças graduais, o gráfico de CD $\overline{MDR}_{HT(G)}$ (Figura 29)

Tabela 16 – Resultados obtidos usando HT, considerando a base artificial LED tendo 10% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.

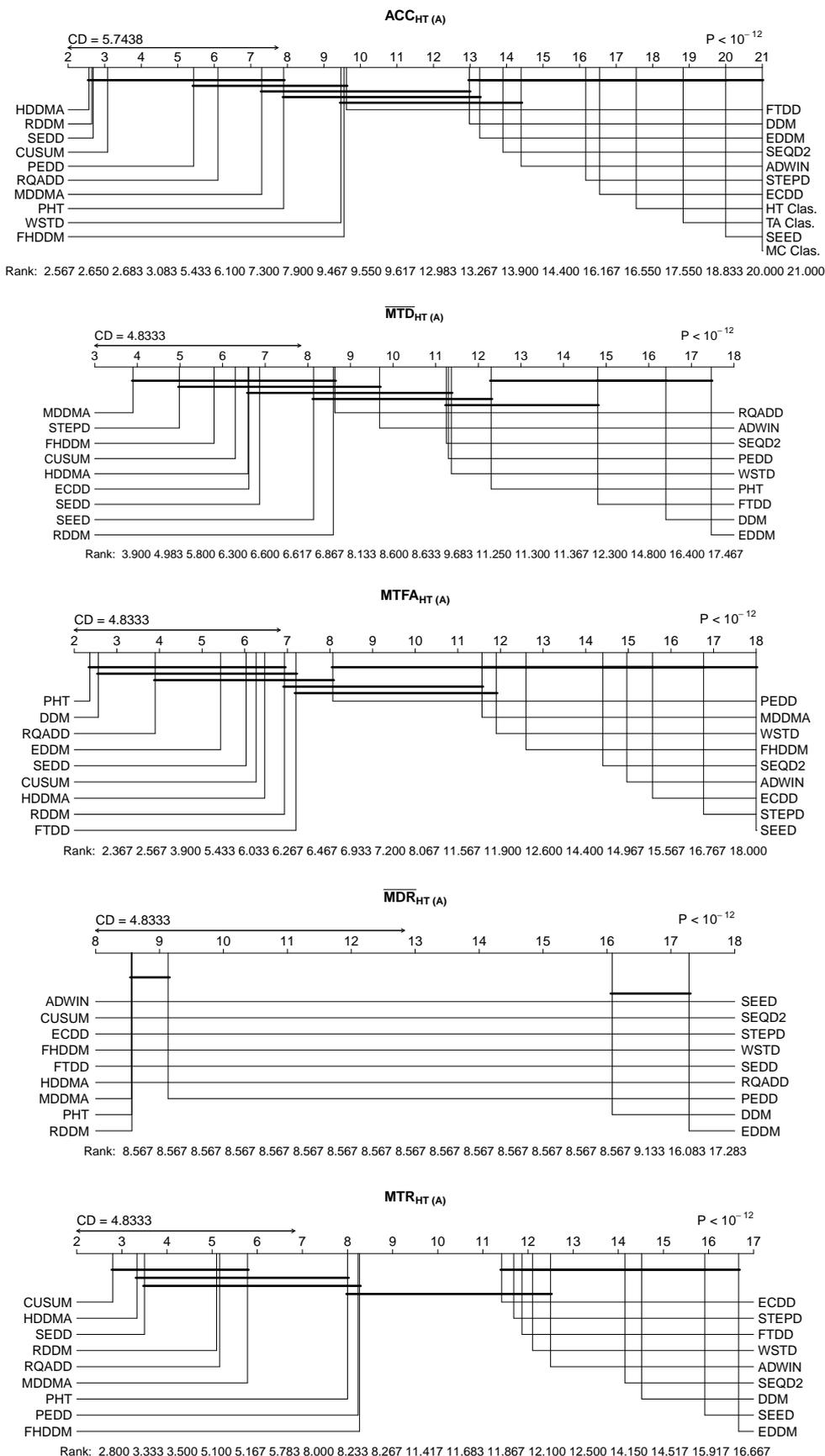
$LED_{(10\%)}$	Classificador/Detector _(HT)	Acurácia(%)	MTD	MTFA	MDR	MTR
Abrupta	Hoeffding Tree	68,35 ± 0,25				
	Majority Class	10,11 ± 0,05				
	Temporally Augmented Classifier	67,84 ± 0,25				
	ADWIN	70,62 ± 0,35	1269,33 ± 885,41	2204,63 ± 605,22	0,00 ± 0,00	4,11 ± 1,52
	CUSUM	73,54 ± 0,09	273,47 ± 24,58	34402,87 ± 1169,74	0,00 ± 0,00	134,56 ± 17,35
	DDM	71,38 ± 0,41	24338,73 ± 5883,55	41173,96 ± 2595,42	0,43 ± 0,12	3,16 ± 1,99
	ECDD	69,12 ± 0,15	594,33 ± 233,78	1301,52 ± 36,80	0,00 ± 0,00	13,70 ± 10,85
	EDDM	71,34 ± 0,14	31972,89 ± 2228,68	34378,20 ± 2379,75	0,48 ± 0,06	0,59 ± 0,10
	FHDDM	72,83 ± 0,14	953,33 ± 613,23	6818,15 ± 849,00	0,00 ± 0,00	37,75 ± 13,17
	FTDD	72,78 ± 0,16	9605,33 ± 1796,38	30326,40 ± 2573,77	0,00 ± 0,00	7,81 ± 8,30
	HDDMA	73,55 ± 0,08	353,03 ± 91,89	32162,11 ± 2005,23	0,00 ± 0,00	145,14 ± 40,60
	MDDMA	73,17 ± 0,09	1769,77 ± 1874,64	9407,78 ± 1101,90	0,00 ± 0,00	93,75 ± 26,19
	PHT	73,12 ± 0,10	1752,68 ± 164,54	39265,11 ± 1155,33	0,00 ± 0,00	24,19 ± 3,19
	RDDM	73,56 ± 0,09	414,96 ± 68,93	29493,81 ± 3199,27	0,00 ± 0,00	78,65 ± 10,17
	SEED	56,16 ± 0,18	5837,60 ± 4947,90	279,19 ± 3,37	0,00 ± 0,00	0,90 ± 0,23
	SEQDRIFT2	70,87 ± 0,52	8545,00 ± 5148,27	2753,95 ± 642,86	0,00 ± 0,00	4,75 ± 2,28
	STEPD	69,41 ± 0,54	1029,90 ± 799,11	903,95 ± 109,57	0,00 ± 0,00	7,55 ± 2,48
	WSTD	72,86 ± 0,16	5104,03 ± 1891,33	9053,52 ± 2115,36	0,00 ± 0,00	19,54 ± 18,28
	PEDD	73,37 ± 0,10	4021,19 ± 2064,44	25854,25 ± 3426,23	0,02 ± 0,02	40,41 ± 15,57
	RQADD	73,33 ± 0,11	1452,87 ± 800,79	36335,77 ± 1056,68	0,00 ± 0,00	94,22 ± 25,01
SEDD	73,54 ± 0,10	1301,92 ± 1259,11	34663,92 ± 1017,40	0,00 ± 0,00	148,87 ± 44,51	
Gradual	Hoeffding Tree	68,36 ± 0,28				
	Majority Class	10,11 ± 0,05				
	Temporally Augmented Classifier	67,83 ± 0,27				
	ADWIN	70,09 ± 0,38	2348,00 ± 1175,09	2247,03 ± 564,36	0,00 ± 0,00	2,52 ± 1,24
	CUSUM	73,41 ± 0,08	23430,22 ± 4371,42	41358,81 ± 1537,12	0,00 ± 0,00	13,72 ± 16,20
	DDM	71,20 ± 0,54	27395,13 ± 6812,89	40858,60 ± 2886,16	0,48 ± 0,14	4,93 ± 3,96
	ECDD	68,98 ± 0,15	867,88 ± 140,32	1310,76 ± 35,27	0,00 ± 0,00	1,90 ± 0,43
	EDDM	71,32 ± 0,17	33756,46 ± 2052,63	36594,98 ± 2636,71	0,52 ± 0,06	0,52 ± 0,08
	FHDDM	72,63 ± 0,14	3273,07 ± 629,88	7015,16 ± 790,94	0,00 ± 0,00	2,88 ± 0,77
	FTDD	72,58 ± 0,12	12806,64 ± 1714,35	31160,60 ± 3009,76	0,00 ± 0,00	2,70 ± 0,42
	HDDMA	73,42 ± 0,08	7196,24 ± 2876,09	32779,69 ± 2102,75	0,01 ± 0,02	24,56 ± 11,87
	MDDMA	72,99 ± 0,09	7023,64 ± 1933,88	11052,78 ± 1605,69	0,00 ± 0,00	2,74 ± 1,05
	PHT	73,08 ± 0,09	3082,61 ± 1635,09	40516,52 ± 1407,82	0,00 ± 0,00	24,25 ± 6,47
	RDDM	73,45 ± 0,09	14840,33 ± 4312,36	32637,85 ± 2188,18	0,00 ± 0,00	23,42 ± 19,39
	SEED	56,31 ± 0,25	5349,87 ± 2684,95	287,01 ± 9,47	0,00 ± 0,00	1,04 ± 0,36
	SEQDRIFT2	70,98 ± 0,24	15610,00 ± 2283,45	5481,79 ± 1563,92	0,00 ± 0,00	0,38 ± 0,11
	STEPD	69,39 ± 0,50	2607,78 ± 950,42	994,89 ± 121,49	0,00 ± 0,00	1,10 ± 0,59
	WSTD	72,45 ± 0,15	9921,00 ± 2071,95	10524,95 ± 1973,32	0,00 ± 0,00	1,46 ± 0,47
	PEDD	73,14 ± 0,12	14835,12 ± 2973,89	30864,34 ± 3000,76	0,03 ± 0,03	5,49 ± 4,79
	RQADD	72,99 ± 0,11	7044,70 ± 2167,98	38735,36 ± 1029,97	0,00 ± 0,00	11,02 ± 3,41
SEDD	73,29 ± 0,09	8241,74 ± 2294,06	35203,55 ± 1274,19	0,00 ± 0,00	9,14 ± 3,70	

Fonte: Elaborada pelo autor (2021).

mostra que o ADWIN também obteve a menor taxa de detecções perdidas. Nessas configurações, apenas o EDDM não foi competitivo.

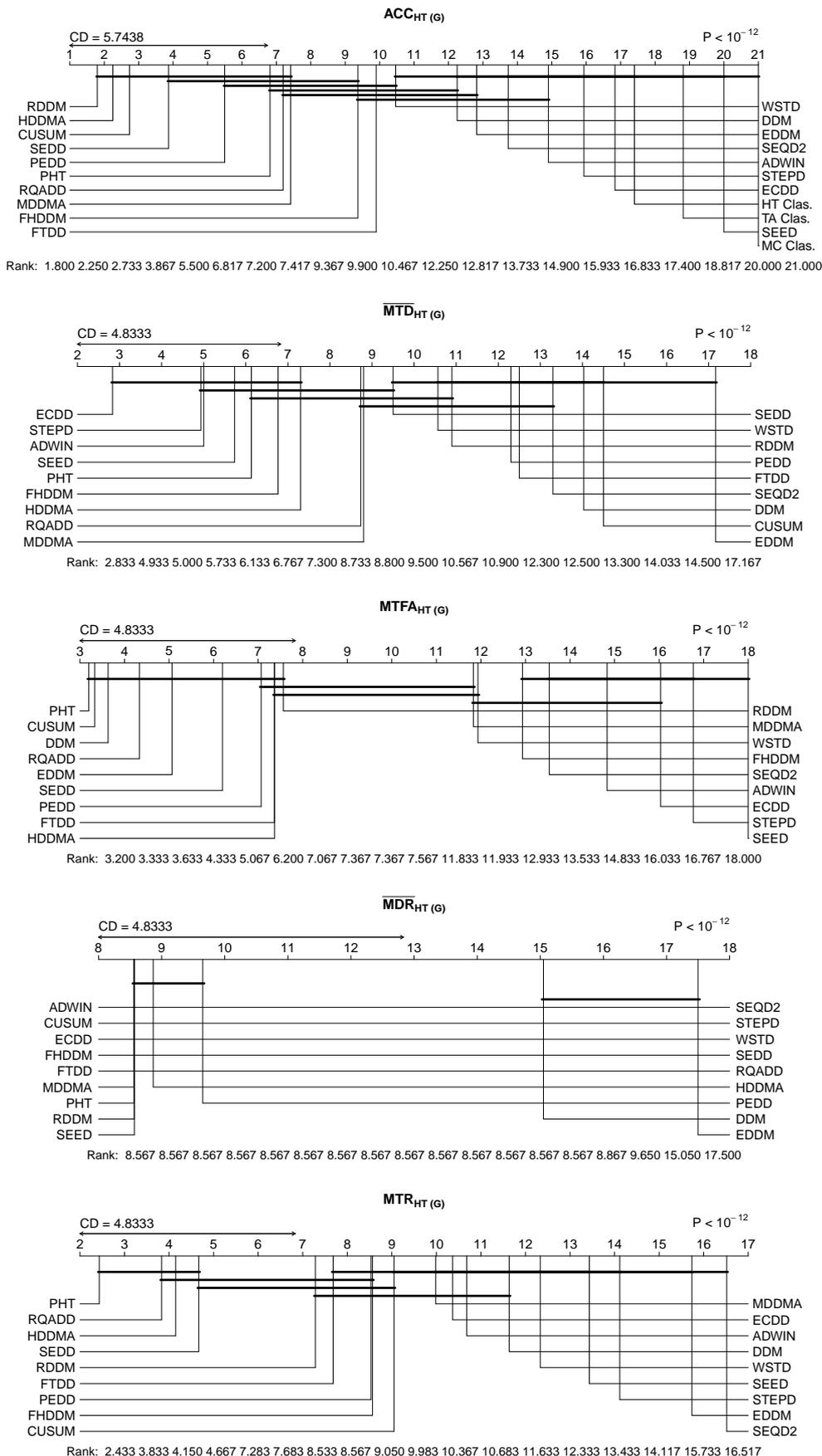
No tocante ao MDR nas bases com 5% de ruído ($LED_{(5\%)}$) e mudanças abruptas usando o NB, o ADWIN obteve a menor taxa de detecções perdidas como aponta o gráfico de

Figura 30 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 31 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

CD $\overline{MDR}_{NB(A)}$ (Figura 50 do Apêndice A). Nessa configuração, o teste F_F não identificou nenhuma diferença entre os detectores, em que $P = 0,9657$. No caso da configuração com mudanças graduais, o gráfico de CD $\overline{MRD}_{NB(G)}$ (Figura 51 do Apêndice A) mostra que o ADWIN também obteve o melhor resultado com o MDR. Nessa configuração, apenas os detectores HDDMA, SEDD, MDDMA, CUSUM, e WSTD não foram competitivos.

Nas bases com 10% de ruído ($LED_{(10\%)}$) e mudanças abruptas usando HT, o ADWIN apresentou a menor taxa de detecções perdidas, como apresentados no gráfico de CD $\overline{MDR}_{HT(A)}$ (Figura 30). Com relação às bases com mudanças graduais, o ADWIN também obteve a menor taxa de detecções perdidas como apresentados no gráfico de CD $\overline{MDR}_{HT(G)}$ (Figura 31). Nessas configurações, apenas o DDM e o EDDM não foram competitivos.

Nas bases com 10% de ruído ($LED_{(10\%)}$) e mudanças abruptas usando NB, o ADWIN também obteve a menor taxa de detecções perdidas, mas nenhum dos detectores avaliados apresentaram diferenças estatísticas significativas, como indicado no gráfico de CD $\overline{MDR}_{NB(A)}$ (Figura 52 do Apêndice A). Nessa configuração, o Teste não identificou diferenças estatísticas entre os detectores apesar de $P = 0,0367$, sugerindo que o nível de significância $\alpha = 5\%$ não foi apropriado para indicar diferenças nesses resultados.

Com relação às bases com mudanças graduais, o ADWIN obteve a menor taxa de detecções perdidas como apresentados nos gráficos de CD $\overline{MDR}_{NB(G)}$ (Figura 53 do Apêndice A), em que apenas o PEDD e o SEDD foram competitivos entre três detectores propostos. Nesse caso, apenas os detectores FTDD, CUSUM, e RQADD não foram competitivos.

Com relação ao MTR com 5% de ruído ($LED_{(5\%)}$) usando o HT, o MDDMA obteve o melhor desempenho nessa métrica com mudanças abruptas, não sendo estatisticamente diferente do SEDD e do RQADD como mostrado no gráfico de CD $MTR_{HT(A)}$ (Figura 28). Nessa configuração os detectores CUSUM e HDDMA também foram competitivos em relação ao MDDMA. Nas mudanças graduais, o gráfico de CD $MTR_{HT(G)}$ (Figura 29) mostra que o PHT obteve o melhor resultado com essa métrica. Nessa configuração, apenas o PEDD e o RQADD foram competitivos entre os detectores propostos. Entre os demais detectores, o FHDDM, HDDMA, e o MDDMA, também foram competitivos nessa configuração em relação ao PHT.

No tocante às bases com 5% de ruído ($LED_{(5\%)}$) usando NB, o MDDMA foi o melhor entre todos os detectores com relação ao MTR, em que o SEDD e o RQADD foram os únicos competitivos entre os detectores propostos como apresentado no gráfico de CD $\overline{MTR}_{NB(A)}$ (Figura 50 do Apêndice A). Nessa configuração, o MDDMA também não apresentou dife-

renças estatísticas significativas em relação aos detectores CUSUM e HDDMA. No caso da configuração com mudanças graduais, o gráfico de CD $MTR_{NB(G)}$ (Figura 51 do Apêndice A) mostra que o PHT obteve o melhor resultado com o MTR. Nessa configuração apenas o RQADD entre os detectores propostos foi competitivo. Nesse caso, os detectores DDM, FHDDM, e HDDMA também foram competitivos.

Nas bases com 10% de ruído ($LED_{(10\%)}$) usando o HT, o CUSUM obteve o melhor MTR com mudanças abruptas, onde apenas o SEDD e o RQADD dos detectores três propostos foram competitivos como mostrado nos gráfico de CD $MTR_{HT(A)}$ (Figura 30). Nessa configuração, os detectores HDDMA, MDDMA, e RDDM também foram competitivos em relação ao CUSUM. Na configuração com mudança gradual, o gráfico de CD $MTR_{HT(G)}$ (Figura 31) indica que o PHT obteve o melhor MTR, não apresentando diferença estatística significativa em relação ao HDDMA, com o HT. Nesse caso, apenas o SEDD e o RQADD dos detectores propostos foram competitivos.

Nas configurações com 10% de ruído ($LED_{(10\%)}$) usando NB, o SEDD foi o melhor entre todos os detectores com relação ao MTR. Além do SEDD, entre os três detectores propostos, o RQADD também foi competitivo como apresentado no gráfico de CD $MTR_{NB(A)}$ (Figura 52 do Apêndice A). Nesse caso, o SEDD não apresentou diferenças estatísticas significativas em relação aos detectores CUSUM, HDDMA, MDDMA, PHT, e RDDM. No caso da configuração com mudanças graduais, o gráfico de CD $MTR_{NB(G)}$ (Figura 53 do Apêndice A) mostra que o PHT obteve o melhor resultado com o MTR. Nesse caso, apenas o SEDD e o RQADD forma competitivos entre os detectores propostos. Entre os demais detectores, o DDM, e o HDDMA também foram competitivos nessa configuração.

5.3.1.3 Resultados e análises considerando as configurações com 2 e 5 classes

Os resultados experimentais com 2 classes ($RBF_{(2)}$) mostram que o *Classificador Temporalmente Aumentado* obteve acurácia superior em relação ao *Classificador de Classe Majoritária*, e aos classificadores HT e NB sem detectores. Entretanto, não apresentou acurácia superior à maioria dos detectores, considerando tanto mudanças abruptas quanto graduais, como apresentado na Tabela 17 e 30 do Apêndice A.

A Tabela 18 e a Tabela 31 do Apêndice A mostram que esse classificador também foi melhor, em termos de acurácia, nas bases com 5 classes ($RBF_{(5)}$) em relação ao *Classificador de Classe Majoritária*, e a ambos os classificadores HT e NB sem detectores, mas também não

apresentou acurácia superior à maioria dos detectores, considerando tanto mudanças abruptas quanto graduais.

A acurácia dos classificadores e detectores usando o HT nas bases com 2 classes ($RBF_{(2)}$), considerando mudanças abruptas, é confirmada no gráfico de CD $ACC_{HT(A)}$ (Figura 32). Nessa configuração, o FTDD obteve a melhor acurácia, não apresentando diferença estatística em relação ao WSTD, onde apenas o SEDD foi competitivo entre os detectores propostos. No gráfico de CD $ACC_{HT(G)}$ (Figura 33), referente ao teste estatístico obtido da configuração com mudanças graduais, o FTDD também obteve vantagem na acurácia em relação aos demais detectores usando o HT. Do mesmo modo, apenas o SEDD foi competitivo entre os detectores propostos. Nessas configurações, os detectores CUSUM, DDM, FHDDM, HDDMA, PHT, e SEQDR2, também foram competitivos em ambas as bases com mudanças abruptas e graduais.

Ainda na configuração com a base de dados $RBF_{(2)}$ considerando o NB, o MDDMA fica no topo do ranqueamento com mudanças abruptas, como indicado no gráfico de CD $ACC_{NB(A)}$ da Figura 54 do Apêndice A. Nessa configuração, apenas o SEDD e o RQADD foram competitivos entre os detectores propostos.

Na configuração com mudanças graduais considerando o NB, o MDDMA também fica no topo do ranqueamento no gráfico de CD $ACC_{NB(G)}$ da Figura 55 do Apêndice A. Nesse caso, apenas o RQADD foi competitivo entre os detectores propostos. Nessas configurações, os detectores FHDDM e RDDM foram competitivos em ambas as bases abruptas e graduais, o CUSUM foi competitivo nas bases com mudanças abruptas, e o ADWIN foi competitivo nas bases graduais.

Com relação aos resultados experimentais com 5 classes ($RBF_{(5)}$), a Tabela 18 e a Tabela 18 do Apêndice A mostram que o *Classificador Temporalmente Aumentado* também obteve desempenho superior em relação ao *Classificador de Classe Majoritária* e aos classificadores HT e NB usados isoladamente, considerando tanto mudanças abruptas quanto graduais. Entretanto, os detectores apresentaram uma acurácia superior ao *Classificador Temporalmente Aumentado* nesses resultados.

No cenário com 5 classes ($RBF_{(5)}$), o gráfico de CD $ACC_{HT(A)}$ na Figura 34 mostra que o FTDD obteve a melhor acurácia usando o HT, não apresentando diferença estatística significativa em relação ao PHT. No caso da mudança gradual, o PHT apresentou a melhor acurácia, não apresentando diferença estatística significativa em relação ao FTDD, como mostrado no gráfico de CD $ACC_{HT(A)}$ da Figura 35.

Nessas configurações, apenas o SEDD foi competitivo entre os detectores propostos em

ambas as bases com mudanças abruptas e graduais. Entre os demais detectores, o CUSUM, DDM, HDDMA, RDDM, e SEQDR2 também foram competitivos em ambas as bases.

Em relação ao classificador NB, o FHDDM obteve a melhor acurácia em mudanças abruptas, não apresentando diferenças estatísticas significativas em relação aos detectores FTDD e WSTD, como apresentado no gráfico de CD $ACC_{NB(A)}$ da Figura 34 do Apêndice A. No caso de mudanças graduais usando esse mesmo classificador, o gráfico de CD $ACC_{NB(G)}$ da Figura 35 do Apêndice A mostra que o FHDDM também obteve a melhor acurácia. Nessas configurações, os detectores CUSUM, HDDMA, MDDMA, RDDM, e STEPDP também foram competitivos em ambas as bases com mudanças abruptas e graduais. Vale salientar que nenhum dos três detectores propostos (PEDD, RQADD, e SEDD) foram competitivos.

Em relação ao MTD usando o HT com 2 classes ($RBF_{(2)}$) e mudanças abruptas, o ECDD apresentou o menor MTD como mostrado nos gráficos de CD $\overline{MTD}_{HT(A)}$ (Figura 32), em que nenhum dos três detectores propostos foi competitivo com o ECDD. Nesse caso, o RDDM e o ADWIN não apresentaram diferenças estatísticas significativas em relação ao ECDD. Na configuração com mudanças graduais, o ECDD obteve o melhor MTD, em que apenas o RQADD dos três detectores propostos foi competitivo, como mostrado no gráfico de CD $\overline{MTD}_{HT(G)}$ (Figura 33). Nessa configuração, o PHT foi competitivo em relação ao ECDD considerando as bases graduais, e os detectores SEED e STEPDP foram competitivos considerando ambas as bases abruptas e graduais.

Em relação ao MTD usando o NB com 2 classes ($RBF_{(2)}$) e mudanças abruptas, o ECDD apresentou o menor MTD como mostrado nos gráficos de CD $\overline{MTD}_{NB(A)}$ (Figura 54 do Apêndice A). Nessa configuração, apenas o SEDD dos três detectores propostos não foi estatisticamente diferente do ECDD. Entre os demais detectores, o ADWIN, MDDMA, e SEED também foram competitivos nessa configuração.

Analisando a configuração com mudanças graduais, o ECDD também obteve o melhor MTD. Dos detectores propostos, somente o RQADD não apresentou diferenças estatísticas significativas em relação ao ECDD, como mostrado no gráfico de CD $\overline{MTD}_{NB(G)}$ (Figura 55 do Apêndice A). Em ambas as bases com mudanças abruptas e graduais, o detector STEPDP não apresentou diferenças estatísticas significativas em relação ao ECDD.

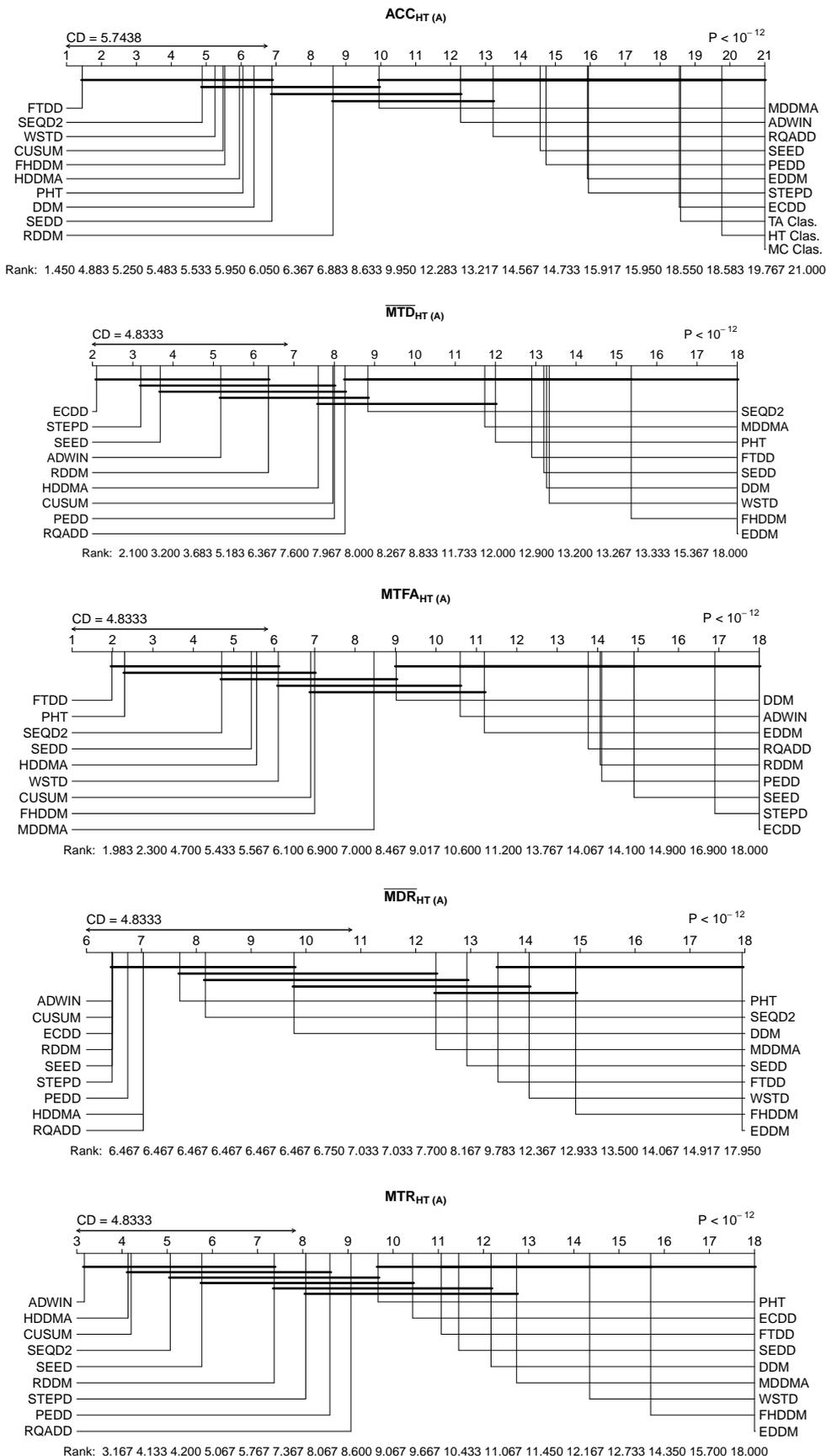
Nas bases com 5 classes ($RBF_{(5)}$) e mudanças abruptas usando o HT, o STEPDP apresentou o melhor desempenho como mostrado nos gráficos de CD $\overline{MTD}_{HT(A)}$ (Figura 34). Nessa configuração, os detectores FTDD, SEED, e WSTD foram competitivos em relação ao STEPDP. Nas bases com mudança graduais, o SEED obteve o melhor desempenho com o

Tabela 17 – Resultados obtidos usando HT, considerando a base artificial RBF com duas classes, com um intervalo de confiança de 95%.

$RBF_{(2)}$	$Classificador/Detector_{(HT)}$	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Hoeffding Tree	81,88 ± 0,11					
	Majority Class	70,01 ± 0,03					
	Temporally Augmented Classifier	82,11 ± 0,11					
	ADWIN	83,45 ± 0,06	604,80 ± 864,10	23618,14 ± 2218,00	0,00 ± 0,00	145,75 ± 25,55	
	CUSUM	84,31 ± 0,08	294,68 ± 26,03	30328,22 ± 2699,32	0,00 ± 0,00	107,86 ± 12,92	
	DDM	84,20 ± 0,13	7451,06 ± 2785,85	27186,46 ± 4197,66	0,11 ± 0,06	10,60 ± 4,40	
	ECDD	82,21 ± 0,04	82,28 ± 33,58	473,80 ± 5,91	0,00 ± 0,00	20,67 ± 8,08	
	EDDM	82,92 ± 0,10	30314,20 ± 154,00	21799,57 ± 1383,65	0,50 ± 0,00	0,36 ± 0,02	
	FHDDM	84,32 ± 0,12	12880,92 ± 20,54	32261,51 ± 678,60	0,25 ± 0,00	1,88 ± 0,04	
	FTDD	84,79 ± 0,09	11167,60 ± 1588,19	43404,92 ± 1532,16	0,21 ± 0,04	37,89 ± 43,60	
	HDDMA	84,26 ± 0,09	1317,61 ± 1243,19	33499,39 ± 2122,73	0,02 ± 0,02	145,57 ± 46,68	
	MDDMA	83,88 ± 0,09	9122,47 ± 2158,99	30015,92 ± 1276,43	0,18 ± 0,04	43,93 ± 29,80	
	PHT	84,29 ± 0,08	3868,22 ± 2050,07	42924,14 ± 2442,11	0,04 ± 0,04	21,50 ± 4,11	
	RDDM	84,00 ± 0,09	251,34 ± 61,55	10425,60 ± 2295,93	0,00 ± 0,00	56,61 ± 16,66	
	SEED	83,13 ± 0,06	966,13 ± 1210,12	8450,14 ± 1068,04	0,00 ± 0,00	90,46 ± 20,27	
	SEQDRIFT2	84,37 ± 0,12	2856,67 ± 1944,09	35340,44 ± 3861,74	0,05 ± 0,04	114,76 ± 31,20	
	STEPD	82,95 ± 0,04	175,96 ± 137,41	1999,29 ± 44,55	0,00 ± 0,00	46,17 ± 12,79	
	WSTD	84,35 ± 0,11	12367,79 ± 769,22	32356,12 ± 1000,09	0,22 ± 0,03	2,38 ± 0,85	
	Gradual	PEDD	83,11 ± 0,04	1285,72 ± 1180,93	10663,54 ± 1318,43	0,01 ± 0,02	37,29 ± 9,91
		RQADD	83,30 ± 0,16	1367,23 ± 1215,56	13321,53 ± 3384,94	0,02 ± 0,02	39,15 ± 12,47
SEDD		84,17 ± 0,09	9907,08 ± 2040,11	33356,67 ± 1008,76	0,19 ± 0,04	53,11 ± 41,99	
Hoeffding Tree		81,85 ± 0,09					
Majority Class		70,00 ± 0,03					
Temporally Augmented Classifier		82,03 ± 0,09					
ADWIN		83,27 ± 0,10	7948,27 ± 2479,26	22216,67 ± 2302,51	0,02 ± 0,03	5,32 ± 1,79	
CUSUM		83,96 ± 0,07	16339,92 ± 2253,65	35151,62 ± 1007,02	0,23 ± 0,02	2,65 ± 1,56	
DDM		84,27 ± 0,12	11086,51 ± 2909,19	32051,42 ± 3304,17	0,17 ± 0,05	8,45 ± 4,58	
ECDD		81,92 ± 0,04	460,28 ± 192,03	479,46 ± 6,30	0,00 ± 0,00	1,56 ± 0,26	
EDDM		82,89 ± 0,10	30170,44 ± 215,54	23169,17 ± 1671,21	0,50 ± 0,00	0,38 ± 0,03	
FHDDM		84,03 ± 0,12	18852,29 ± 2570,62	32346,41 ± 1479,23	0,34 ± 0,05	1,32 ± 0,23	
FTDD		84,37 ± 0,15	27393,27 ± 3423,69	44413,52 ± 1823,31	0,37 ± 0,06	1,23 ± 0,27	
HDDMA		83,98 ± 0,09	16079,82 ± 3250,97	37412,79 ± 2044,01	0,16 ± 0,05	3,57 ± 1,47	
MDDMA		83,66 ± 0,05	20248,29 ± 2409,46	31519,45 ± 43,69	0,13 ± 0,05	1,56 ± 0,26	
PHT		84,26 ± 0,05	3169,15 ± 1709,20	46522,09 ± 1946,45	0,02 ± 0,03	27,86 ± 4,74	
RDDM		83,76 ± 0,08	10413,37 ± 2560,15	14731,11 ± 1509,18	0,09 ± 0,05	6,13 ± 4,31	
SEED		82,78 ± 0,06	3136,00 ± 1293,96	8947,71 ± 1339,10	0,00 ± 0,00	4,59 ± 1,14	
SEQDRIFT2		84,11 ± 0,11	21541,67 ± 3674,48	34910,84 ± 2858,11	0,08 ± 0,04	2,19 ± 0,83	
STEPD		82,63 ± 0,05	1881,72 ± 462,74	2101,51 ± 66,45	0,00 ± 0,00	1,67 ± 0,44	
WSTD	83,64 ± 0,18	25503,58 ± 2600,01	34523,95 ± 1491,12	0,38 ± 0,05	0,95 ± 0,17		
PEDD	82,77 ± 0,04	8820,54 ± 2081,07	12157,55 ± 943,75	0,06 ± 0,04	2,26 ± 0,82		
RQADD	83,14 ± 0,19	3883,49 ± 1991,48	8382,29 ± 2345,96	0,02 ± 0,03	13,12 ± 6,39		
SEDD	83,94 ± 0,09	26298,82 ± 2762,61	34783,46 ± 1093,55	0,26 ± 0,05	1,22 ± 0,41		

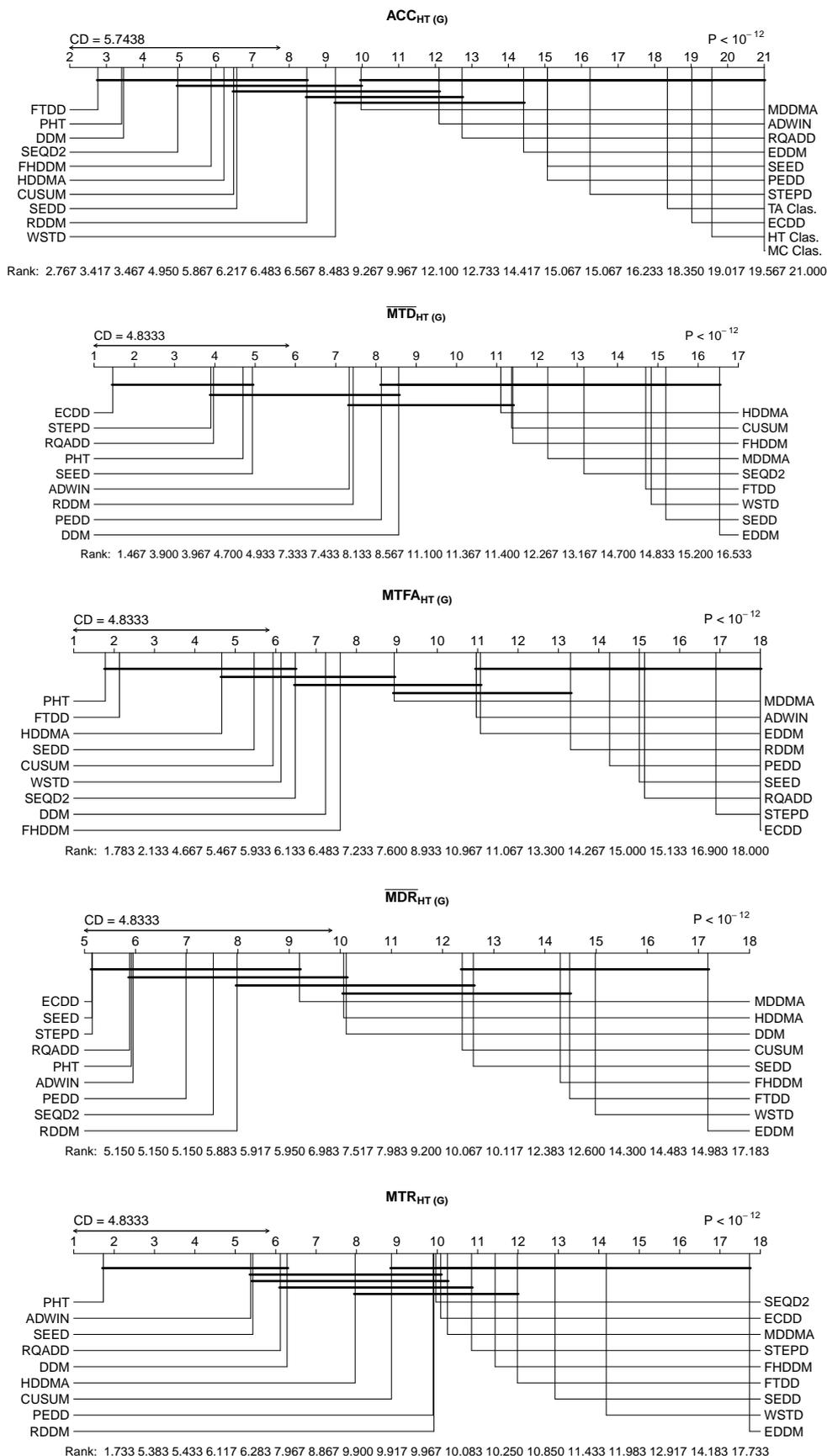
Fonte: Elaborada pelo autor (2021).

Figura 32 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com duas classes, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 33 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com duas classes, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

HT, não apresentando diferença estatística significativa em relação aos detectores ADWIN e STEPDM, como mostrado no gráfico de CD $\overline{MTD}_{HT(G)}$ (Figura 35). Os detectores MDDMA e FHDDM também foram competitivos em ambas as bases com mudanças abruptas e graduais. Entretanto, nenhum dos três detectores propostos foi competitivo nessas configurações,

Nas bases com 5 classes ($RBF_{(5)}$) e mudanças abruptas usando o NB, o STEPDM apresentou o melhor desempenho como indicado no gráfico de CD $\overline{MTD}_{NB(A)}$ (Figura 56 do Apêndice A), não apresentando diferença estatística significativa em relação aos detectores ADWIN, FHDDM, FTDD, MDDMA, SEED, e WSTD.

Nas bases com mudança graduais, o STEPDM também obteve o melhor desempenho com o NB, não apresentando diferença estatística significativa em relação aos detectores ADWIN, FHDDM, MDDMA, SEED, WSTD, como mostrado no gráfico de CD $\overline{MTD}_{NB(G)}$ (Figura 57 do Apêndice A). Nesse caso, os detectores ADWIN, FHDDM, e MDDMA também foram competitivos em ambas as bases com mudanças abruptas e graduais. Nessas configurações, nenhum dos três detectores propostos foi competitivo em relação ao STEPDM.

Analisando o MTFA nas bases com 2 classes ($RBF_{(2)}$) e mudanças abruptas, o FTDD apresentou o maior MTFA nas avaliações com o HT, não apresentando diferença estatística significativa em relação aos detectores PHT e FHDDM, como mostrado nos gráficos de CD $MTFA_{HT(A)}$ (Figura 32). Nessa configuração, apenas o SEED e o dos três detectores propostos não foram estatisticamente diferentes do FTDD.

Nas mudanças graduais, o gráfico de CD $MTFA_{HT(G)}$ (Figura 33) mostra que o PHT obteve o maior MTFA, não apresentando diferença estatística significativa em relação ao FTDD. Novamente, apenas o SEED foi competitivo entre os três detectores propostos. Em ambas as configurações, os detectores PHT, FHDDM, CUSUM, HDDMA, SEQDR2, e WSTD também foram competitivos.

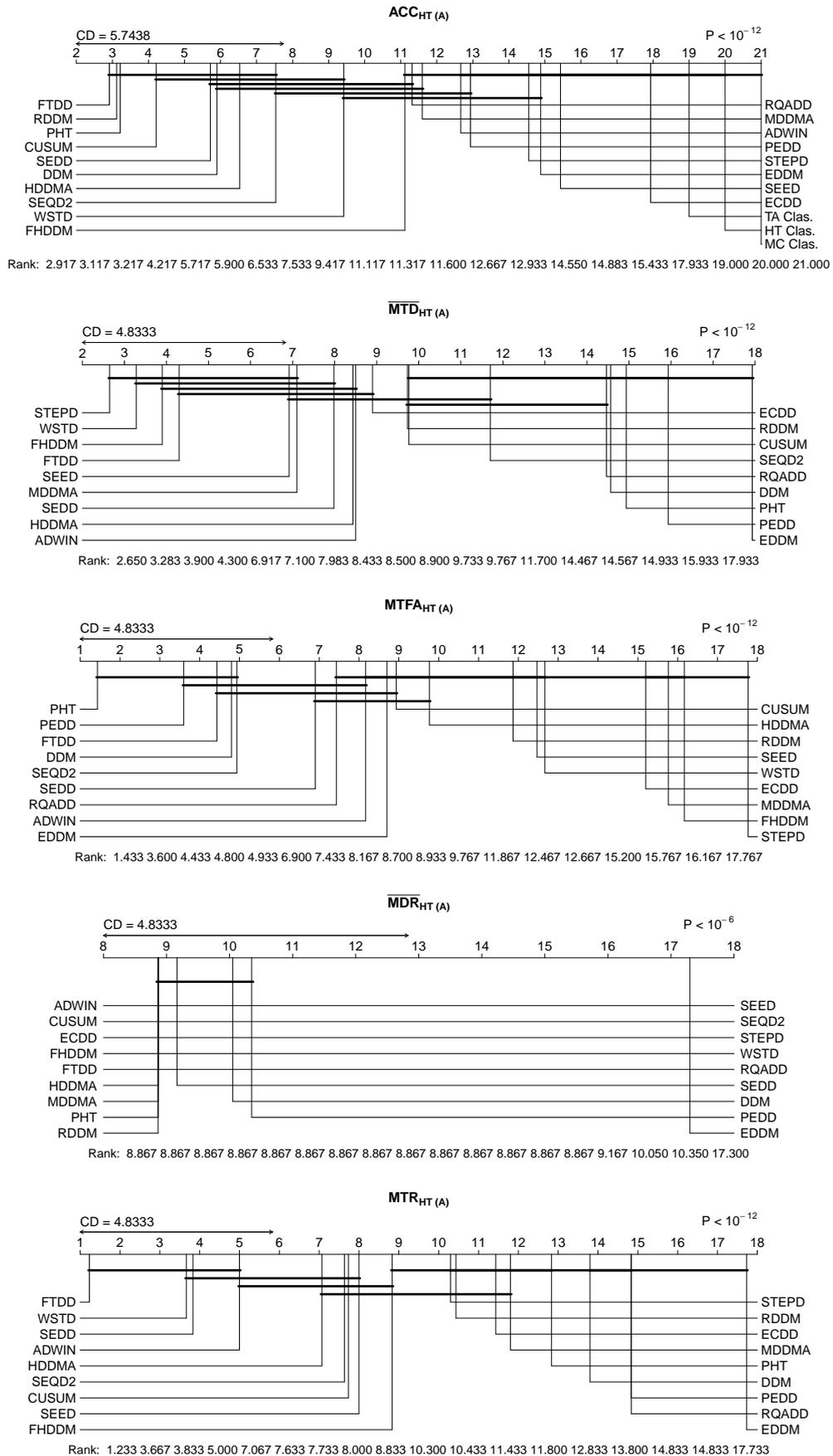
Analisando o MTFA nas bases com 2 classes ($RBF_{(2)}$) e mudanças abruptas, o PHT apresentou o maior MTFA nas avaliações com o NB, em que apenas o SEED foi competitivo entre os três detectores propostos, como indicado no gráfico de CD $MTFA_{NB(A)}$ (Figura 54 do Apêndice A). Nas mudanças graduais, o gráfico de CD $MTFA_{NB(G)}$ (Figura 55 do Apêndice A) mostra que o FTDD obteve o maior MTFA. Novamente, apenas o SEED foi competitivo entre os três detectores propostos. Nessas configurações, os detectores ADWIN, HDDMA, SEQDR2, e WSTD foram competitivos em ambas as bases com mudanças abrupta e graduais. Além desses, o FTDD também foi competitivo nas bases com mudanças abruptas, e o PHT também foi competitivo nas bases com mudanças graduais.

Tabela 18 – Resultados obtidos usando HT, considerando a base artificial RBF com cinco classes, com um intervalo de confiança de 95%.

$RBF_{(5)}$	Classificador/Detector _(HT)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Hoeffding Tree	59,36 ± 0,30					
	Majority Class	32,98 ± 0,02					
	Temporally Augmented Classifier	61,02 ± 0,25					
	ADWIN	64,93 ± 0,07	650,67 ± 905,72	18611,24 ± 1141,51	0,00 ± 0,00	243,23 ± 37,99	
	CUSUM	66,26 ± 0,18	115,09 ± 19,50	16335,42 ± 3019,07	0,00 ± 0,00	149,61 ± 28,57	
	DDM	66,04 ± 0,30	3298,27 ± 1867,04	28500,93 ± 3970,17	0,03 ± 0,03	31,29 ± 9,65	
	ECDD	63,75 ± 0,10	2967,50 ± 1634,29	3775,75 ± 973,78	0,00 ± 0,00	198,34 ± 161,43	
	EDDM	64,56 ± 0,20	21039,88 ± 2244,34	17204,65 ± 2802,64	0,33 ± 0,06	0,50 ± 0,07	
	FHDDM	65,06 ± 0,10	103,07 ± 66,83	2575,85 ± 210,94	0,00 ± 0,00	125,61 ± 28,09	
	FTDD	66,49 ± 0,12	27,95 ± 5,00	29491,26 ± 3594,13	0,00 ± 0,00	1174,06 ± 183,11	
	HDDMA	65,94 ± 0,14	109,65 ± 32,35	14125,39 ± 2347,24	0,00 ± 0,00	223,84 ± 91,88	
	MDDMA	65,02 ± 0,09	127,22 ± 82,53	2665,95 ± 140,33	0,00 ± 0,00	67,31 ± 15,20	
	PHT	66,40 ± 0,09	1300,39 ± 236,90	41379,85 ± 2184,22	0,00 ± 0,00	37,17 ± 5,20	
	RDDM	66,41 ± 0,13	119,78 ± 12,73	9378,45 ± 2202,99	0,00 ± 0,00	90,68 ± 31,48	
	SEED	64,66 ± 0,06	699,47 ± 944,21	6338,71 ± 420,85	0,00 ± 0,00	133,19 ± 23,42	
	SEQDRIFT2	65,66 ± 0,19	200,00 ± 0,00	27405,90 ± 3206,82	0,00 ± 0,00	137,03 ± 16,03	
	STEPD	64,76 ± 0,08	28,46 ± 19,34	1503,07 ± 47,05	0,00 ± 0,00	80,45 ± 9,28	
	WSTD	65,38 ± 0,18	37,20 ± 36,82	6598,68 ± 1381,07	0,00 ± 0,00	334,23 ± 70,06	
	PEDD	64,83 ± 0,18	4483,05 ± 2053,32	30996,66 ± 1237,06	0,04 ± 0,04	16,68 ± 4,19	
	RQADD	65,20 ± 0,23	1425,33 ± 602,13	20712,73 ± 4709,79	0,00 ± 0,00	18,41 ± 4,37	
	SEDD	66,02 ± 0,13	499,27 ± 870,38	20879,27 ± 1398,37	0,01 ± 0,02	277,75 ± 31,27	
	Gradual	Hoeffding Tree	59,28 ± 0,29				
		Majority Class	32,98 ± 0,02				
		Temporally Augmented Classifier	60,93 ± 0,24				
		ADWIN	64,37 ± 0,08	3320,53 ± 1518,24	19636,42 ± 1595,10	0,00 ± 0,00	10,40 ± 2,05
		CUSUM	65,73 ± 0,17	28774,98 ± 3921,63	36842,29 ± 2773,70	0,12 ± 0,05	1,39 ± 0,47
		DDM	65,85 ± 0,33	12637,18 ± 3600,30	41122,57 ± 3202,17	0,14 ± 0,05	35,21 ± 22,52
ECDD		62,99 ± 0,09	17599,96 ± 1151,87	7972,93 ± 1624,62	0,03 ± 0,04	0,42 ± 0,07	
EDDM		64,43 ± 0,25	21419,29 ± 2234,83	17556,46 ± 2828,04	0,34 ± 0,06	0,51 ± 0,09	
FHDDM		64,45 ± 0,10	1562,24 ± 229,76	2528,46 ± 194,31	0,00 ± 0,00	1,85 ± 0,28	
FTDD		65,84 ± 0,17	14291,02 ± 2648,71	33876,10 ± 2299,36	0,11 ± 0,05	4,77 ± 3,23	
HDDMA		65,26 ± 0,17	8595,85 ± 2011,61	18200,34 ± 2300,09	0,00 ± 0,00	3,04 ± 0,79	
MDDMA		64,43 ± 0,10	2023,05 ± 355,61	2743,89 ± 170,63	0,00 ± 0,00	1,81 ± 0,47	
PHT		66,08 ± 0,12	17224,39 ± 2738,54	47727,25 ± 2307,40	0,23 ± 0,02	3,26 ± 1,98	
RDDM		65,87 ± 0,13	21274,22 ± 3521,63	22945,51 ± 2749,09	0,07 ± 0,04	1,84 ± 1,45	
SEED		64,07 ± 0,08	1884,80 ± 1193,96	6246,56 ± 750,23	0,00 ± 0,00	6,19 ± 1,15	
SEQDRIFT2		65,02 ± 0,18	16970,00 ± 3680,25	28278,46 ± 2097,43	0,02 ± 0,02	2,55 ± 0,95	
STEPD		64,16 ± 0,07	1411,49 ± 257,18	1558,91 ± 47,38	0,00 ± 0,00	1,37 ± 0,27	
WSTD		64,52 ± 0,19	4320,03 ± 734,26	6875,28 ± 1088,90	0,00 ± 0,00	1,99 ± 0,46	
PEDD		63,55 ± 0,24	34170,31 ± 2496,97	34840,40 ± 903,14	0,49 ± 0,04	0,57 ± 0,12	
RQADD		64,53 ± 0,21	19346,41 ± 4805,72	26065,15 ± 5211,33	0,19 ± 0,07	1,81 ± 0,70	
SEDD		65,19 ± 0,17	17208,37 ± 2565,19	24696,15 ± 1596,65	0,02 ± 0,03	1,84 ± 0,52	

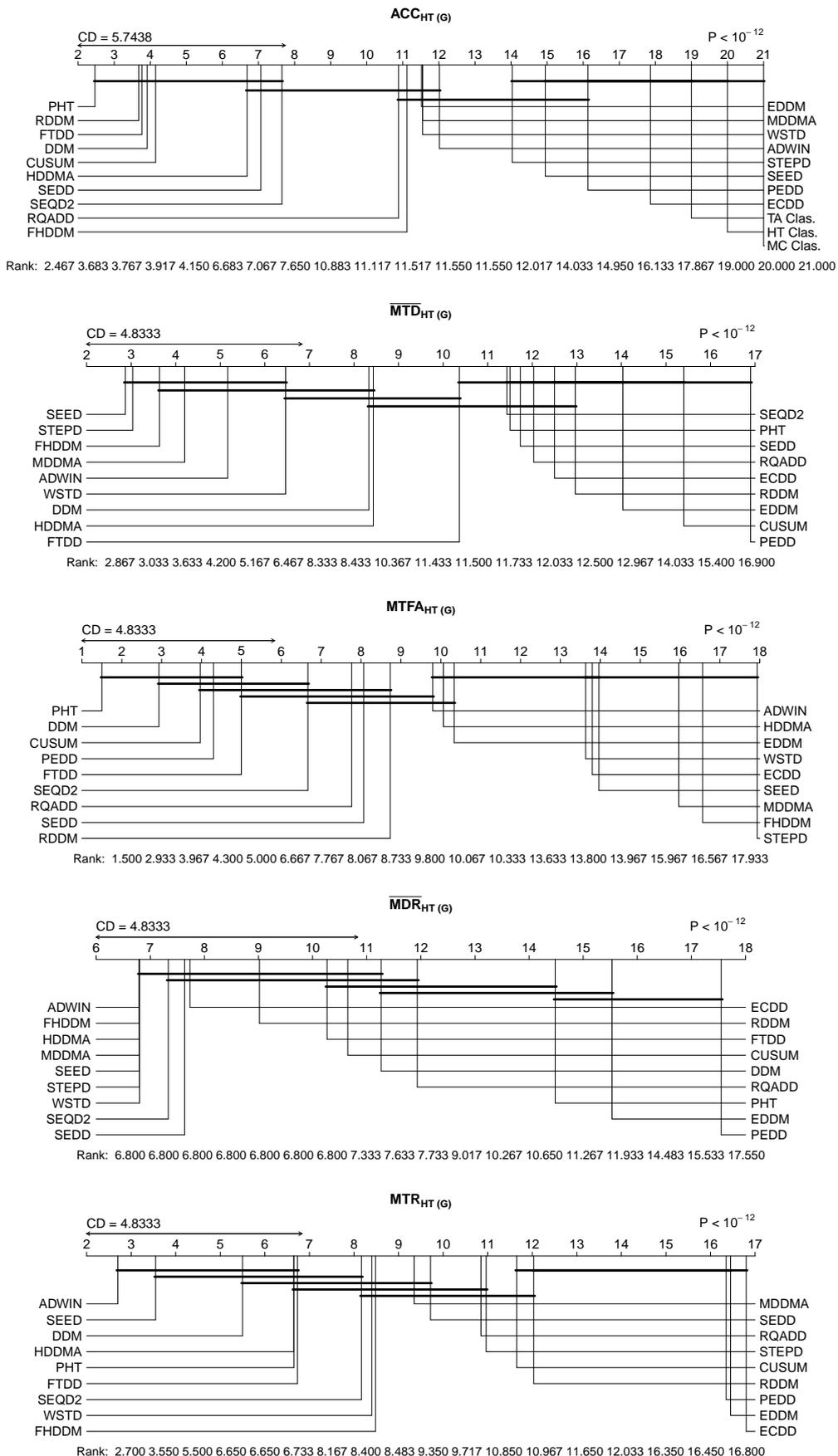
Fonte: Elaborada pelo autor (2021).

Figura 34 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 35 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Nas bases com 5 classes ($RBF_{(5)}$) e mudanças abruptas usando HT, o PHT apresentou o melhor desempenho como mostrado no gráfico de CD $MTFA_{HT(A)}$ (Figura 34), não apresentando diferença estatística significativa em relação aos detectores DDM, FTDD, e SEQDR2. Nessa configuração, apenas o PEDD foi competitivo entre os detectores propostos. Nas bases com mudanças graduais, o gráfico de CD $MTFA_{HT(G)}$ da Figura 35 indica que o PHT também obteve o melhor desempenho nessa configuração. Novamente, apenas o PEDD foi competitivo entre os detectores propostos, e o PHT também não apresentou diferenças estatísticas significativas em relação aos detectores CUSUM, DDM, FTDD.

Nas bases com 5 classes ($RBF_{(5)}$) e mudanças abruptas usando o NB, o PHT também apresentou o melhor desempenho como indica o gráfico de CD $MTFA_{NB(A)}$ (Figura 34 do Apêndice A), onde apenas o PEDD foi competitivo entre os três detectores propostos. Na configuração com mudança gradual, o PHT apresentou novamente o melhor desempenho como mostrado no gráfico de CD $MTFA_{NB(G)}$ (Figura 35 do Apêndice A). Novamente, apenas o PEDD foi competitivo entre os três detectores propostos. Nessas configurações, os detectores ADWIN, DDM, FTDD, e SEQDR2 foram competitivos em relação ao PHT considerando ambas as mudanças abruptas e graduais.

Em relação ao MDR nas bases com 2 classes ($RBF_{(2)}$) e mudanças abruptas usando o HT, o ADWIN obteve a menor taxa de detecções perdidas como aponta o gráfico de CD $\overline{MDR}_{HT(A)}$ (Figura 32). Nessa configuração, apenas o RQADD e o PEDD entre os três detectores propostos não apresentaram diferenças estatísticas significativas em relação ao ADWIN. Somente os detectores MDDMA, SEDD, FTDD, WSTD, FHDDM, e o EDDM não foram competitivos. Nas mudanças graduais, o gráfico de CD $\overline{MDR}_{HT(G)}$ (Figura 33) mostra que o ECDD obteve a menor taxa de detecções perdidas. Novamente, o RQADD e o PEDD entre os três detectores propostos não apresentaram diferenças estatísticas significativas em relação ao ECDD. Entre os demais detectores, o ADWIN, DDE, MDDMA, PHT, RDDM, SEED, SEQDR2, e STEPD foram competitivos.

No tocante ao MDR nas bases com 2 classes ($RBF_{(2)}$) e mudanças abruptas usando o NB, o ADWIN também obteve a menor taxa de detecções perdidas como aponta o gráfico de CD $\overline{MDR}_{NB(A)}$ (Figura 54 do Apêndice A). Nessa configuração, os três detectores propostos também não apresentam diferenças estatísticas significativas em relação ao ADWIN. Nessa configuração, apenas o WSTD e o FHDDM não foram competitivos em relação ao ADWIN. No caso da configuração com mudanças graduais, o gráfico de CD $\overline{MRD}_{NB(G)}$ (Figura 55 do Apêndice A) mostra que o ECDD também obteve o melhor resultado com o MDR. Nessa

caso, apenas o PEDD e o RQADD foram competitivos entre os três detectores propostos. O ECDD também não apresentou diferença estatística significativa em relação aos detectores ADWIN, DDE, DDM, EDDM, MDDMA, RDDM, SEED, e STEPD.

Nas bases com 5 classes ($RBF_{(5)}$) e mudanças abruptas usando HT, o ADWIN também apresentou a menor taxa de detecções perdidas, e nenhum dos detectores propostos apresentaram diferenças estatísticas significativas, como apresentados no gráfico de CD $\overline{MDR}_{HT(A)}$ (Figura 34). Nesse caso, apenas o EDDM não foi competitivo. Com relação às bases com mudanças graduais, o ADWIN também obteve a menor taxa de detecções perdidas como apresentados no gráfico de CD $\overline{MDR}_{HT(G)}$ (Figura 35), em que apenas o SEDD foi competitivo entre os detectores propostos. Nessa configuração, apenas os detectores PHT, EDDM, e PEDD não foram competitivos.

Nas bases com 5 classes ($RBF_{(5)}$) e mudanças abruptas usando NB, o ADWIN também obteve a menor taxa de detecções perdidas, em que nenhum dos detectores propostos apresentaram diferenças estatísticas significativas, como indicado no gráfico de CD $\overline{MDR}_{NB(A)}$ (Figura 34 do Apêndice A). Nesse caso, apenas o ECDD não foi competitivo. Considerando as bases com mudanças graduais, o ADWIN obteve novamente a menor taxa de detecções perdidas como apresentados nos gráficos de CD $\overline{MDR}_{NB(G)}$ (Figura 35 do Apêndice A), em que apenas o RQADD e o SEDD foram competitivos entre três detectores propostos. Nessa configuração, os detectores DDM, PHT, ECDD, e PEDD não foram competitivos.

Com relação ao MTR com 2 classes ($RBF_{(2)}$) usando o HT, o ADWIN obteve o melhor desempenho nessa métrica com mudanças abruptas, como mostrado no gráfico de CD $MTR_{HT(A)}$ (Figura 32). Nessa configuração, nenhum dos três detectores propostos foram competitivos com o ADWIN. Entretanto, os detectores CUSUM, HDDMA, RDDM, SEED, e SEQDR2 não apresentaram diferenças estatísticas significativas em relação ao ADWIN. Nas mudanças graduais, o gráfico de CD $MTR_{HT(G)}$ (Figura 33) mostra que o PHT obteve o melhor resultado com essa métrica. Nessa configuração, apenas o RQADD foi competitivo entre os três detectores propostos. Nessa configuração, os detectores ADWIN, DDM, e SEED foram competitivos em relação ao PHT.

No tocante às bases com 2 classes ($RBF_{(2)}$) usando NB, o ADWIN foi o melhor entre todos os detectores com relação ao MTR, em que o SEDD foi o único competitivo entre os detectores propostos, como indicado no gráfico de CD $\overline{MTR}_{NB(A)}$ (Figura 54 do Apêndice A). Nesse caso, os detectores CUSUM e SEQDR2 também foram competitivos em relação ao ADWIN. No caso da configuração com mudanças graduais, o gráfico de CD $MTR_{NB(G)}$

(Figura 55 do Apêndice A) mostra que o RQADD obteve o melhor resultado com o MTR, não apresentaram diferenças estatísticas significativas em relação aos detectores ADWIN, DDM, e SEED. Nessa configuração, o SEDD também foi competitivo. Os detectores FTDD, HDDMA, e MDDMA foram competitivos em ambas as configurações com mudanças abruptas e graduais, usando o NB.

Nas bases com 5 classes ($RBF_{(5)}$) usando o HT, o FTDD obteve o melhor MTR com mudanças abruptas, onde apenas o SEDD dos detectores propostos foram competitivos como mostrado no gráfico de CD $MTR_{HT(A)}$ (Figura 34). Entre os demais detectores, o ADWIN e o WSTD não apresentaram diferenças estatísticas significativas em relação ao FTDD. Na configuração com mudança gradual, o gráfico de CD $MTR_{HT(G)}$ (Figura 35) indica que o ADWIN obteve o melhor MTR com o HT. Nesse caso, nenhum dos detectores propostos foram competitivos. Nessa configuração os detectores DDM, FTDD, HDDMA, PHT, e SEED foram competitivos em relação ao ADWIN.

Com 5 classes ($RBF_{(5)}$) usando NB, o FTDD também foi o melhor entre todos os detectores com relação ao MTR, não apresentando diferença estatística significativa em relação aos detectores ADWIN e WSTD (gráfico de CD $MTR_{NB(A)}$ da Figura 34 do Apêndice A). O gráfico de CD $MTR_{NB(G)}$ (Figura 35 do Apêndice A) mostra que o ADWIN obteve o melhor resultado com o MTR, não apresentando diferença estatística significativa em relação aos detectores HDDMA e SEED. Entre os três detectores propostos, apenas o SEDD foi competitivo, e somente com mudanças abruptas.

5.3.2 Avaliação do uso de memória e tempo de processamento

De modo similar às análises da Seção 5.3, a Tabela 19 apresenta os resultados do uso da memória em bytes e o tempo médio de execução em segundos, onde os melhores resultados estão destacados em **negrito**. Esses resultados agregam todos os cenários das 6 bases artificiais testados com os classificadores HT e NB, respectivamente. Como mostrado na tabela, o uso dos detectores com o classificador HT apresenta um custo computacional significativamente mais elevado quando comparado com o NB.

A Figura 36 apresenta os gráficos de CD relacionados à avaliação geral em termos do uso da memória (MEM^{-1}) e tempo médio de execução ($TIME^{-1}$), também considerando os resultados com os classificadores HT (os gráficos MEM_{HT}^{-1} e $TIME_{HT}^{-1}$), e NB (os gráficos MEM_{NB}^{-1} e $TIME_{NB}^{-1}$), respectivamente. Semelhante às variáveis \overline{MTD} e \overline{MDR} definidas

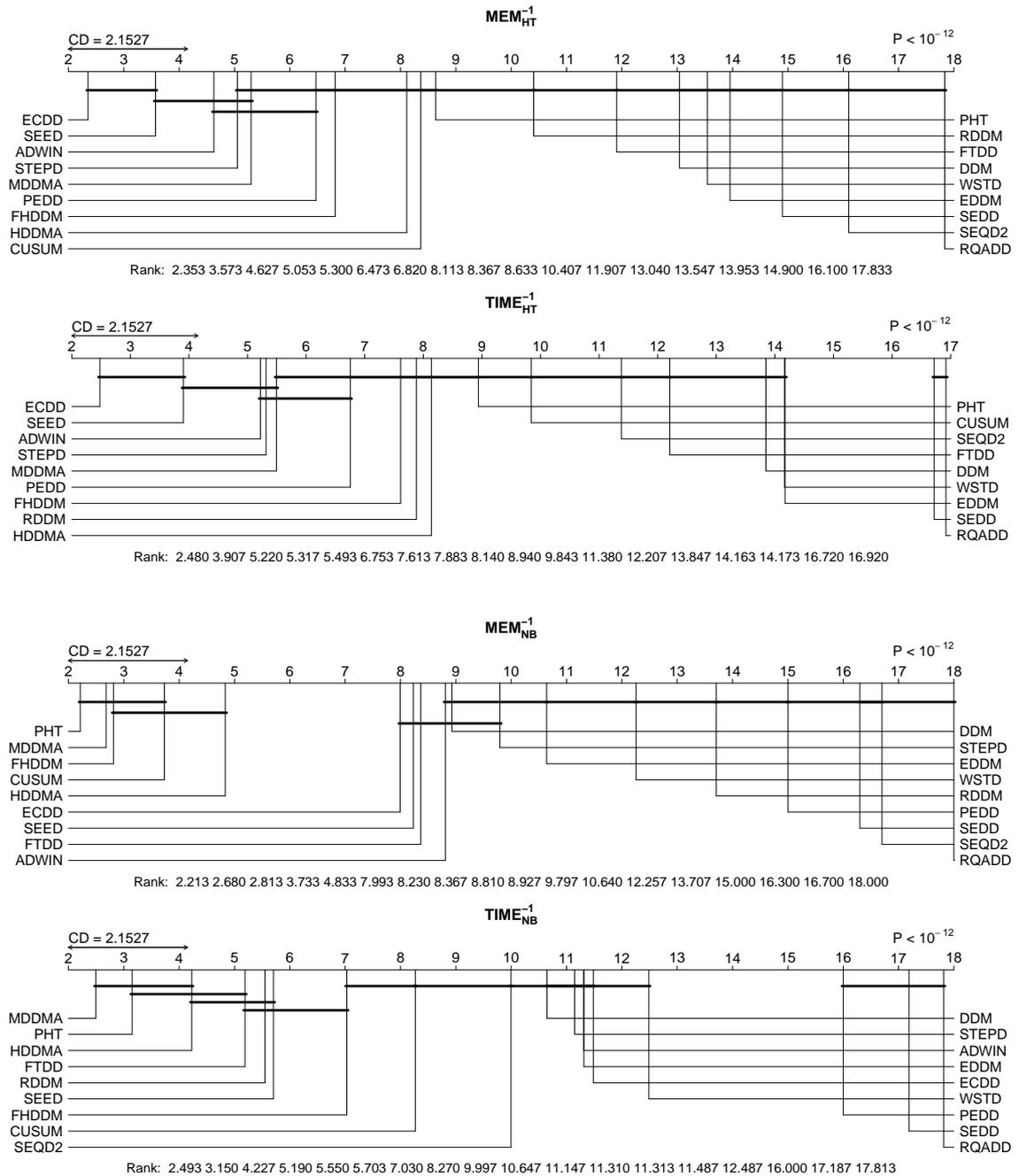
Tabela 19 – Resultados do uso de memória em bytes e tempo médio de execução em segundos obtidos usando os classificadores HT e NB, considerando as 6 bases artificiais obtidas dos geradores, com um intervalo de confiança de 95%.

<i>Classificador</i>	<i>Detector</i>	Memória (bytes/segundos)	Tempo (segundos)
HT	ADWIN	2908,17 ± 414,06	126,28 ± 9,45
	CUSUM	20138,98 ± 2060,11	451,43 ± 21,81
	DDM	21510,43 ± 2421,27	465,04 ± 25,39
	ECDD	610,08 ± 26,65	53,90 ± 1,03
	EDDM	21376,07 ± 2942,78	439,40 ± 29,52
	FHDDM	18644,21 ± 2194,80	428,80 ± 26,94
	FTDD	22632,83 ± 2147,38	478,65 ± 23,30
	HDDMA	19066,76 ± 2616,64	425,03 ± 30,41
	MDDMA	18399,20 ± 1798,17	416,78 ± 20,90
	PHT	20341,47 ± 2231,01	442,71 ± 23,93
	RDDM	19306,08 ± 2375,26	418,70 ± 26,20
	SEED	3406,50 ± 429,83	130,19 ± 8,40
	SEQD2	33817,22 ± 4451,45	330,03 ± 31,33
	STEPD	7440,05 ± 1923,76	223,56 ± 30,28
	WSTD	37874,76 ± 11536,34	600,59 ± 79,90
	PEDD	11389,06 ± 1588,40	305,57 ± 25,42
	RQADD	128830,94 ± 6660,77	688,49 ± 27,16
	SEDD	26530,23 ± 2311,95	608,23 ± 22,49
NB	ADWIN	127,37 ± 2,03	27,79 ± 0,31
	CUSUM	110,51 ± 1,88	26,75 ± 0,41
	DDM	123,47 ± 2,71	26,84 ± 0,39
	ECDD	128,11 ± 1,97	28,23 ± 0,45
	EDDM	130,72 ± 2,94	27,13 ± 0,44
	FHDDM	107,78 ± 2,28	26,40 ± 0,39
	FTDD	123,52 ± 2,12	25,25 ± 0,38
	HDDMA	115,39 ± 2,48	25,79 ± 0,32
	MDDMA	105,98 ± 1,80	24,39 ± 0,31
	PHT	105,57 ± 2,45	24,92 ± 0,42
	RDDM	173,52 ± 3,66	25,58 ± 0,42
	SEED	126,74 ± 2,31	25,88 ± 0,39
	SEQD2	1662,36 ± 47,46	28,00 ± 0,42
	STEPD	125,38 ± 2,54	26,24 ± 0,46
	WSTD	137,31 ± 2,06	27,68 ± 0,40
	PEDD	388,89 ± 6,56	56,94 ± 0,94
	RQADD	41001,81 ± 637,64	272,03 ± 4,21
	SEDD	1481,71 ± 23,52	213,73 ± 3,03

Fonte: Elaborada pelo autor (2021).

na Seção 5.3 para atender a ordem do ranqueamento dos testes estatísticos, as variáveis $MEM^{-1} = 1/MEM$ e $TIME^{-1} = 1/TIME$ são definidas como as medidas inversas ao uso de memória (MEM) e ao tempo médio de execução ($TIME$), respectivamente. Essas variáveis também foram utilizadas para manter a correspondência em relação a ordem do ranqueamento dos testes estatísticos, onde valores mais elevados significam melhores resultados.

Figura 36 – Comparação dos resultados para o consumo de memória (*MEM*) e tempo de processamento (*TIME*) usando o teste post-hoc de Nemenyi sobre as 6 bases artificiais, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Os resultados com o classificador HT mostram que o ECDD foi o melhor método tanto em uso de memória quanto em tempo de execução, sendo estatisticamente superior a todos os outros métodos, mas não apresentando diferença estatística significativa em relação ao SEED. Por outro lado, ao considerar o classificador NB, os resultados mostram que o PHT foi superior aos demais em termos de uso de memória, não apresentando diferença estatística significativa em comparação com os detectores MDDMA, FHDDM e CUSUM. Em relação ao tempo de execução, o MDDMA obteve o melhor resultado, mas não apresentou diferença estatística significativa em comparação com o PHT, e HDDMA.

A avaliação desses resultados mostra que os detectores PEDD, RQADD, e SEDD, obtiveram as piores posições nos ranqueamentos, tanto em uso de memória quanto em tempo de execução, com ambos os classificadores HT e NB. De fato, esses detectores são computacionalmente mais caros em relação aos demais. Ambos os detectores PEDD e SEDD tem complexidade assintótica $\mathcal{O}(n \log n)$, relacionada à ordenação para a identificação dos padrões simbólicos e à Fast Fourier Transform (traduzido para o português, Transformada Rápida de Fourier) (FFT), respectivamente.

Finalmente, é óbvio que o RQADD é computacionalmente mais caro por causa de sua complexidade assintótica $\mathcal{O}(n^2)$, relacionada ao cálculo das distâncias entre os estados do *espaço de fase*. Além da complexidade, vale ressaltar que de acordo com a Tabela 11 a maioria dos detectores utilizam uma janela pequena de observações (cerca de 30), o que reduz significativamente o tempo de execução e o consumo de memória, diferentemente dos detectores propostos, como descrito na Seção 5.2.3.

5.4 RESULTADOS E ANÁLISE DO EXPERIMENTO COM BASES REAIS ROTULADAS

Esta seção apresenta a análise dos resultados obtidos com os detectores propostos utilizando as 20 bases reais descritas na Seção 5.1.2.1. Vale destacar que nesse contexto não há garantia de que os dados sejam i.i.d., ou saber previamente quando uma *mudança de conceito* irá ocorrer. Além disso, algumas bases apresentam dependência temporal, ou algum grau de desbalanceamento entre as classes. Nestas condições, faz-se necessário uma análise mais criteriosa com a utilização do *Classificador de Classe Majoritária*, e o *Classificador Temporalmente Aumentado*, como recomendado em Žliobaitė et al. (2015).

De modo similar às análises anteriores, a Tabela 20 apresenta os resultados da acurácia e do ARL de todos os detectores testados com os classificadores HT e NB. Esses resultados agregam

todas as 20 bases reais, em que os melhores resultados também estão destacados em **negrito**. Como mostrado na tabela, o *Classificador Temporalmente Aumentado* apresenta a melhor acurácia, o que sugere que uma quantidade significativa das bases apresentam dependência temporal. Os resultados da Tabela 20 representam o resumo das acurácias obtidas com os classificadores HT e NB, apresentadas nas Tabelas 32 e 33, e das medidas ARL, apresentadas nas Tabelas 34 e 35, no Apêndice B.

A Figura 37 abrange os resultados dos testes de Nemenyi e Friedman aplicados aos detectores usando os classificadores HT e NB. Devido a impossibilidade da identificação do ponto de *mudança de conceito*, apenas a acurácia (ACC) e o ARL são avaliados nos gráficos de CD. O intervalo de confiança (95%) e o nível de significância α (5%) são os mesmos usados nos experimentos anteriores. Como $P < 10^6$, apresentado no canto superior direito dos gráficos da Figura 37, a hipótese nula pode ser rejeitada com segurança, de modo que o teste *post-hoc* de Nemenyi pode ser aplicado para identificar a diferença entre os métodos avaliados.

Como mostrado nos gráficos ACC_{HT} e ACC_{NB} da Figura 37, o *Classificador Temporalmente Aumentado* (TA Class.) obteve os melhores resultados em ambas as avaliações relacionadas a acurácia. Esse resultado sugere que a maioria das bases reais avaliadas apresentam dependência temporal, o que favorece o desempenho desse classificador, como apresentado nas Tabelas 32 e 33 do Apêndice B. Por outro lado, o *Classificador de Classe Majoritária* (MC Class.) apresentou o pior desempenho, com um ranqueamento abaixo dos classificadores HT e NB, o que sugere que a maioria das bases não apresentam um desbalanceamento significativo das classes a ponto de enviesar a classificação. Além disso, a maioria dos detectores apresentaram desempenho superior com respeito à acurácia em comparação aos classificadores HT e NB, usados isoladamente. Esses resultados sugerem que as bases apresentam mudanças que quando consideradas favorecem o aumento da acurácia. Entretanto, devido a dependência temporal, essa melhora não está relacionada à generalização da hipótese indutiva dos classificadores HT e NB. Como explicado em Žliobaitė et al. (2015), a sinalização de muitos alarmes falsos (menor ARL) favorece o aumento da acurácia em bases com dependência temporal.

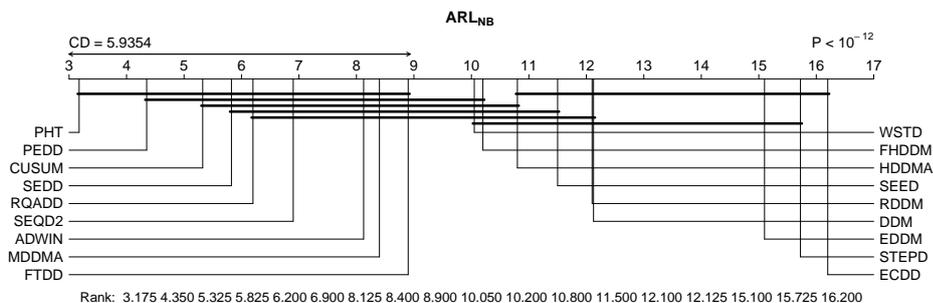
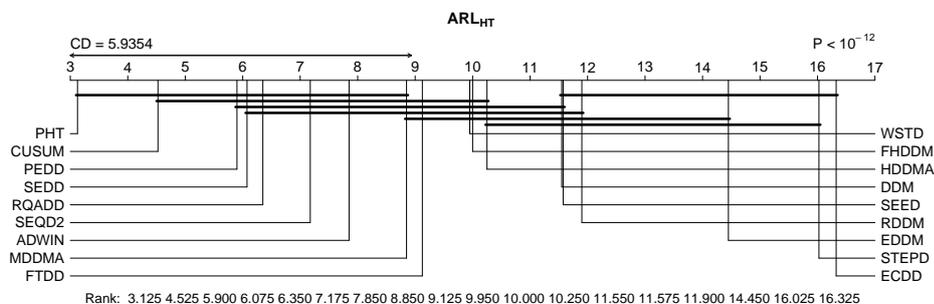
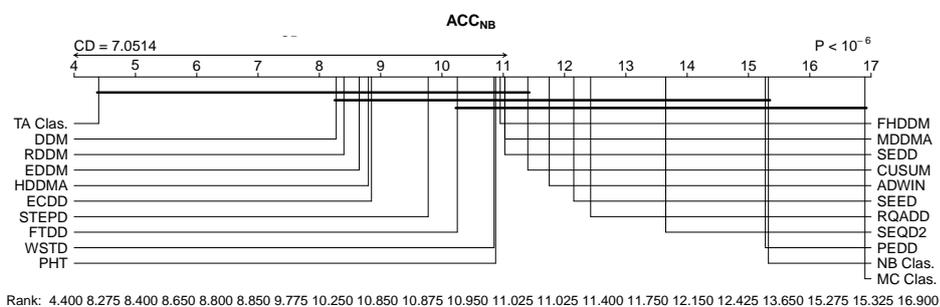
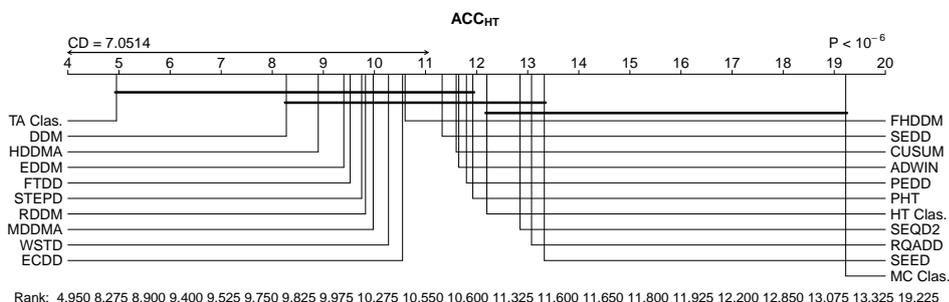
Com relação aos detectores, o gráfico de CD ACC_{HT} na Figura 37 sugere que os detectores DDM, HDDMA, EDDM, FTDD, RDDM, MDDMA, STEPDP, WSTD, ECDD, FHDDM, SEDD, CUSUM, ADWIN, PEDD e PHT não apresentaram diferença estatística significativa com relação à acurácia, usando o HT. De modo semelhante, o gráfico de CD ACC_{NB} sugere que os detectores DDM, RDDM, ECDD, EDDM, HDDMA, STEPDP, FTDD, PHT, WSTD, SEDD, FHDDM, MDDMA, e CUSUM, também não apresentaram diferença estatística signifi-

Tabela 20 – Resultados obtidos usando HT e NB, considerando as 20 bases reais, com um intervalo de confiança de 95%.

Bases reais rotuladas	<i>Classificador/Detector</i>	Acurácia(%)	MTFA
HT	Hoeffding Tree	73,22 ± 9,16	
	Majority Class	55,09 ± 14,98	
	Naive Bayes	62,91 ± 11,51	
	Temporally Augmented Classifier	82,53 ± 8,02	
	ADWIN	77,24 ± 8,41	7487,02 ± 8446,03
	CUSUM	77,60 ± 7,37	9217,01 ± 8293,26
	DDM	79,16 ± 7,82	3796,91 ± 2925,12
	ECDD	78,18 ± 8,44	931,44 ± 1414,55
	EDDM	78,90 ± 7,65	1528,69 ± 1674,38
	FHDDM	78,66 ± 7,73	6518,38 ± 8297,56
	FTDD	78,85 ± 7,61	8186,68 ± 8439,86
	HDDMA	79,00 ± 7,83	8160,94 ± 8302,02
	MDDMA	78,35 ± 7,70	6610,82 ± 8285,04
	PHT	77,69 ± 7,41	12062,33 ± 8249,14
	RDDM	78,82 ± 7,89	3175,22 ± 2689,78
	SEED	75,31 ± 10,13	3383,82 ± 4861,47
	SEQDRIFT2	75,90 ± 8,19	5966,10 ± 7772,78
	STEPD	76,06 ± 10,63	461,00 ± 285,81
	WSTD	78,54 ± 7,84	7072,19 ± 8323,23
	NB	PEDD	77,32 ± 7,38
RQADD		74,97 ± 8,49	13441,78 ± 16481,05
SEDD		77,76 ± 7,59	8850,31 ± 8362,40
ADWIN		73,05 ± 9,38	1716,92 ± 2117,49
CUSUM		75,03 ± 7,18	6006,48 ± 7595,62
DDM		77,92 ± 7,93	5698,31 ± 7663,06
ECDD		77,56 ± 8,22	920,08 ± 1416,02
EDDM		77,32 ± 7,38	706,44 ± 1005,86
FHDDM		76,80 ± 7,27	4503,13 ± 7595,84
FTDD		74,72 ± 9,14	6143,88 ± 7846,74
HDDMA		78,13 ± 7,58	6271,94 ± 7796,41
MDDMA		76,38 ± 7,32	4822,79 ± 7581,89
PHT		75,07 ± 7,32	9126,09 ± 7794,05
RDDM		77,28 ± 7,55	1735,65 ± 1656,15
SEED		74,11 ± 9,51	1200,22 ± 1157,39
SEQDRIFT2		72,88 ± 7,71	1865,79 ± 2252,07
STEPD		72,14 ± 11,82	324,29 ± 229,43
WSTD		74,30 ± 9,46	4757,58 ± 7589,00
PEDD		71,52 ± 8,71	3366,44 ± 2310,03
RQADD		68,61 ± 9,53	7235,11 ± 7870,29
SEDD	77,15 ± 7,54	6500,80 ± 7695,08	

Fonte: Elaborada pelo autor (2021).

Figura 37 – Comparação dos classificadores e detectores com HT e NB usando o teste *post-hoc* de Nemenyi sobre as bases reais rotuladas, considerando um intervalo de confiança de 95%, e o nível de significância $\alpha = 0,05$.



Fonte: Elaborada pelo autor (2021).

cativa com relação à acurácia, considerando o NB. Entretanto, todos os detectores apresentaram um desempenho significativamente inferior ao *Classificador Temporalmente Aumentado* (TA Class.), apesar do teste *post-hoc* de Nemenyi não identificar diferenças estatísticas significativas entre esse classificador e todas as configurações de detectores testadas.

Os gráficos de CD ARL_{HT} e ARL_{NB} agregam a avaliação da média da distância entre alarmes, avaliada por meio da medida ARL. Os gráficos de CD ARL_{HT} e ARL_{NB} mostram que os detectores PHT, CUSUM, PEDD, SEDD, RQADD, SEQD2, ADWIN, MDDMA, e FTDD, não apresentam diferenças estatísticas significativas em relação à sinalização de alarmes falsos, considerando ambos os classificadores HT e NB. Vale salientar que o PHT obteve o melhor ranqueamento em ambas as avaliações do ARL, como apresentado nos gráficos de CD ARL_{HT} e ARL_{NB} , usando o HT e o NB, respectivamente.

É importante observar que os resultados dos testes estatísticos apresentados na Figura 37 corroboram com os argumentos teóricos explanados em Žliobaitė et al. (2015), os quais atestam que a sinalização de muitos alarmes falsos favorece a acurácia da classificação em bases com dependência temporal, independentemente da qualidade do detector de *mudança de conceito*. Desse modo, pode-se observar que os detectores que obtiveram as melhores posições no ranqueamento do gráfico de CD ACC_{HT} , são praticamente os mesmos que ficaram com as piores posições no ranqueamento do gráfico de CD ARL_{HT} , ou seja, sinalizaram mais alarmes falsos. O mesmo pode ser observado entre os gráficos de CD ACC_{NB} e ARL_{NB} .

Entre os detectores propostos, o SEDD obteve o melhor desempenho em termos de acurácia e ARL. Vale lembrar que o critério de qualidade adotado está relacionado com a barra de Critical Differences (traduzido para o português, Diferença Crítica) (CD) ligada ao topo do ranque. Por esse critério, são bem avaliados os métodos ligados à barra conectada ao *Classificador Temporalmente Aumentado* (TA Class.), no caso dos gráficos de CD ACC_{HT} e ACC_{NB} , e aqueles ligados à barra de CD conectada ao PHT, no caso dos gráficos de CD ARL_{HT} e ARL_{NB} . Assim, pode-se afirmar que o SEDD apresentou um resultado equilibrado entre uma boa acurácia e um número reduzido de alarmes, sendo competitivo em relação aos melhores detectores deste experimento.

Os detectores PEDD e RQADD foram bastante competitivos em termos do ARL, ficando bem colocados no ranque dos gráficos de CD ARL_{HT} e ARL_{NB} . Entretanto, somente o PEDD obteve resultado satisfatório em termos de acurácia, e apenas com o classificador HT. A justificativa desse desempenho pode ser o número reduzido de alarmes (ARL elevado), como apresentado nas Tabelas 34 e 35 do Apêndice B.

5.5 RESULTADOS E ANÁLISE DO EXPERIMENTO COM O SISTEMA DE LORENZ

Esta seção apresenta a análise dos resultados utilizando os fluxos de dados gerados com o Sistema Lorenz Transiente descritos na Seção 5.1.1.2. Neste experimento, o Symbolic Labeling Adapter (SLA) foi usado para prover os rótulos por meio dos padrões simbólicos obtidos dos estados do espaço de fase com o objetivo de identificar *mudanças de conceitos*. Vale salientar que não foi investigado o sentido semântico desta rotulação com respeito aos padrões ordinais obtidos dos estados do espaço de fase.

Na comparação dos resultados, as ferramentas de análise Entropia de Permutação (PE), Análise de Quantificação de Recorrência (RQA) e Entropia Espectral (SE) foram utilizadas como referenciais para avaliar a qualidade da detecção das *mudanças de conceito* dos demais detectores, sendo denominadas como: PEDD*, referente à Entropia de Permutação (PE); RQADD* relacionado ao *determinismo percentual* RQA_{DET} da Análise de Quantificação de Recorrência (RQA); e o SEDD*, correspondendo à Entropia Espectral (SE). Esses novos detectores foram configurados com as mesmas parametrizações da Seção 5.2.3, com exceção do parâmetro δ , otimizado para os fluxos de dados usados neste experimento em termos da métrica MTR. O objetivo é avaliar se os detectores acoplados ao HT e NB são competitivos em comparação com as ferramentas da NLTA: PEDD* (Entropia de Permutação (PE)), RQADD* (Análise de Quantificação de Recorrência (RQA)), e SEDD* (Entropia Espectral (SE)) otimizados para este experimento. Nesta avaliação, além dos detectores acoplados ao HT e NB, também foram utilizados o *Classificador de Classe Majoritária*, o *Classificador Temporalmente Aumentado*, e os classificadores HT e NB sem detectores.

Os detectores PEDD*, RQADD*, e SEDD* processam apenas o primeiro componente $x_{i+km\tau}$ (uma observação a cada $m\tau$ observações da série temporal) dos vetores $\mathbf{v}_{(i+km\tau)}^{m,\tau}$, em cada instância $(\mathbf{v}_{(i+km\tau)}^{m,\tau}, \boldsymbol{\pi}_{(i+(k+1)m\tau)}^m)$, onde $k = 0, 1, 2, \dots$, descritas na Seção 4.3. Neste experimento, o parâmetro (τ) relacionado aos retardos temporais foi estimado com a Average Mutual Information Function (traduzido para o português, Função de Informação Mútua Média) (AMIF). Considerando N como o número de observações, m a dimensão de reconstrução do espaço de fase, e τ a defasagem temporal, $N/\tau m$ define a quantidade de instâncias obtidas com o SLA. Logo, o valor dos parâmetros τ e m são inversamente proporcionais ao total de instâncias obtidas. Desse modo, o parâmetro τ foi definido como um quarto da média dos primeiros mínimos locais obtidos como resultados da aplicação da AMIF aos 30 fluxos de dados descritos na Seção 5.1.1.2. Assim, as reconstruções dos espaços

de fase desses 30 fluxos de dados foram definidas nas dimensões $m = 3$, $m = 4$, e $m = 5$, com o retardo temporal $\tau = 3$. Note que o parâmetro m está diretamente relacionado com a quantidade $m!$ de padrões ordinais (ou símbolos) usados como os rótulos das instâncias geradas pelo SLA. Nesse caso, $m!$ corresponde a problemas com 6, 24 e 120 classes, respectivamente.

As Tabelas 21, 22 e 23 apresentam os resultados obtidos usando o *Classificador de Classe Majoritária*, o *Classificador Temporalmente Aumentado*, os classificadores HT e NB sem e com os detectores descritos na Seção 5.2, em que os melhores resultados são mostrados em **negrito**. Nessas tabelas, o *Classificador Temporalmente Aumentado* (TA Clas.) apresentou o melhor desempenho em relação à acurácia, indicando dependência temporal entre os rótulos das instâncias obtidas com o SLA. Isso ocorre devido ao acoplamento das variáveis do sistema, podendo ser reduzida com o aumento dos parâmetros m e τ . Entretanto, à medida que o parâmetro τ aumenta, a relação entre o atributo de entrada $v_{(i+km\tau)}^{m,\tau}$ e o atributo alvo $\pi_{(i+(k+1)m\tau)}^m$ diminui.

No caso do parâmetro m , o que ocorre é o aumento da granularidade dos rótulos definidos por $m!$ símbolos, como constatado nos resultados da acurácia ao longo das Tabelas 21, 22 e 23, referentes às dimensões $m = 3$, $m = 4$, e $m = 5$, respectivamente. Esses resultados apontam uma queda no desempenho geral em termos da acurácia devido à redução do número de rótulos repetidos consecutivamente como resultado do aumento da dimensão m .

As Figuras 38 a 39 apresentam os gráficos de CD obtidos dos testes estatísticos relacionados aos resultados da Tabela 21, considerando $m = 3$ em ambos os classificadores HT e NB, respectivamente. De modo semelhante aos demais experimentos, os testes referem-se às comparações com base na acurácia (ACC), MTD, MTFA, MDR e MTR. Os gráficos de CD obtidos dos testes estatísticos relacionados aos resultados com $m = 4$, apresentados na Tabela 22, são agregados nas Figuras 40 e 41. Por fim, as Figuras 42 e 43 apresentam os gráficos de CD obtidos dos testes estatísticos relacionados aos resultados resumidos na Tabela 23, referentes à dimensão $m = 5$.

Os gráficos de CD ACC_{HT} e ACC_{NB} respectivos às Figuras 38 e 39 mostram que o *Classificador Temporalmente Aumentado* (TA Clas.) obteve a melhor acurácia, tanto em comparação aos detectores usando HT quanto usando NB, com o parâmetro $m = 3$. Esse classificador também ficou no topo do ranqueamento quando avaliado nos gráficos de CD ACC_{HT} e ACC_{NB} das Figuras 40 a 41 referentes à reconstrução do espaço de fase com o parâmetro $m = 4$, bem como nos gráficos de CD ACC_{HT} e ACC_{NB} das Figuras 42 a 43 com $m = 5$. As Figuras 38 e 39 mostram que apenas o RQADD e o PEDD foram competitivos entre os

detectores propostos com relação à acurácia com o NB e $m = 3$. Nessa parametrização, os detectores DDM, CUSUM, PHT, EDDM, e FHDDM foram competitivos usando HT, inclusive esse próprio classificador sem uso de detectores também foi competitivo como mostra o gráfico de CD ACC_{HT} da Figura 38. Com o NB, foram competitivos os detectores ADWIN, SEED, SEQD2, e STEPD, como mostra o (gráfico de CD ACC_{NB} da Figura 39).

Em relação à configuração com o parâmetro $m = 4$, somente os detectores ADWIN e EDDM usando HT foram competitivos em relação ao *Classificador Temporalmente Aumentado*, como apresentado no gráfico de CD ACC_{HT} (Figura 40). No gráfico de CD ACC_{NB} (Figura 41) referente ao classificador NB, foram competitivos os detectores ADWIN, SEED, SEQD2, STEPD, e FTDD. Apenas o PEDD foi competitivo entre os três detectores propostos com relação à acurácia, e somente com o NB. Com ambos os classificadores HT e NB e $m = 5$, os detectores ADWIN, STEPD, SEQD2, e SEED, foram competitivos com relação à acurácia, como apresentado no gráfico de CD ACC_{HT} (Figura 42) e no gráfico de CD ACC_{NB} (Figura 43). Nesse caso, nenhum dos detectores propostos foi competitivo com relação à acurácia.

Os gráficos de CD \overline{MTD}_{HT} (Figura 38) e \overline{MTD}_{NB} (Figura 39) mostram que o detector RQADD* apresentou o melhor resultado para a métrica MTD com o parâmetro $m = 3$, usando ambos os classificadores HT e NB, em que somente os detectores PEDD*, ADWIN, SEED, SEQD2, STEPD, e o PEDD, entre os detectores propostos, foram competitivos. Nessa configuração, o SEDD* também foi competitivo com NB.

Em relação à configuração com o parâmetro $m = 4$, os detectores competitivos foram: SEDD*, PEDD*, ADWIN, e SEED, como apresentado no gráfico de CD \overline{MTD}_{HT} (Figura 40) e no gráfico de CD \overline{MTD}_{NB} (Figura 41) referentes aos classificadores HT e NB, respectivamente. Nesse caso, o RQADD* ficou no topo do ranqueamento. O SEQD2 foi competitivo usando o HT. Entre os detectores propostos somente o PEDD foi competitivo, com ambos os detectores HT e NB. Na configuração com o parâmetro $m = 5$, o RQADD* obteve o melhor ranqueamento, e os detectores PEDD*, ADWIN e SEED foram competitivos com relação ao MTD, como indicado no gráfico de CD \overline{MTD}_{HT} (Figura 42) e no gráfico de CD \overline{MTD}_{NB} (Figura 43) referentes aos classificadores HT e NB, respectivamente. Vale salientar que nenhum dos três detectores propostos foram competitivos em relação ao MTD nessa parametrização.

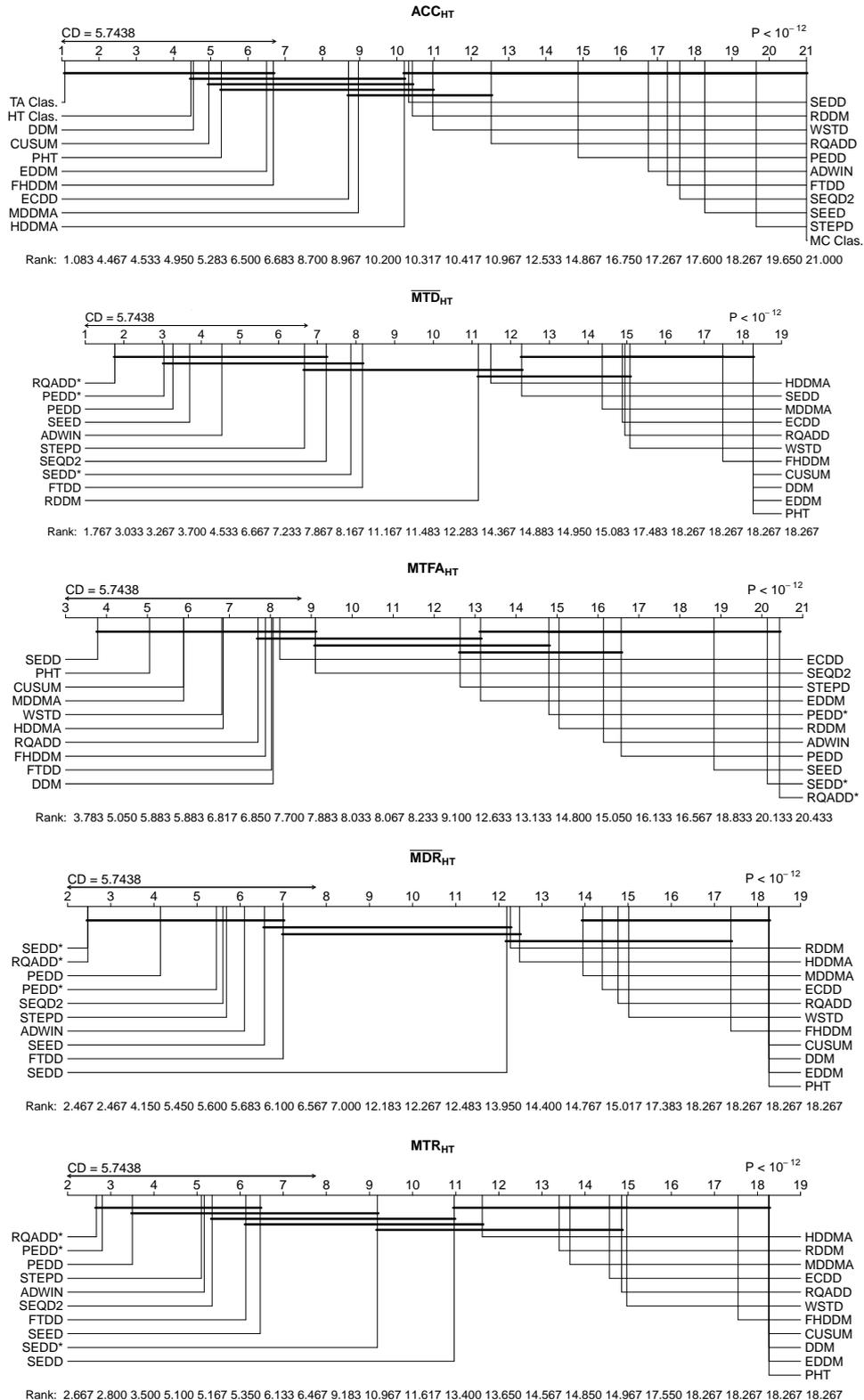
Os resultados dos testes estatísticos para a medida MTFA usando ambos o HT e NB são apresentados nos gráficos de CD $MTFA_{HT}$ e $MTFA_{NB}$ das Figuras 38 e 39 referentes à parametrização $m = 3$, das Figuras 40 e 41 correspondente a $m = 4$, e das Figuras 42 e 43 referentes a $m = 5$. Esses gráficos mostram que o SEDD obteve o melhor ranqueamento

Tabela 21 – Resultados obtidos usando HT e NB a partir das rotulações simbólicas da base artificial do Sistema Lorenz Transiente, considerando a dimensão de imersão $m = 3$ e a defasagens $\tau = 3$, com um intervalo de confiança de 95%.

$Lorenz_{(d=3,\tau=3)}$	Classificador/Detector _(HT)	Acurácia(%)	MTD	MTFA	MDR	MTR	
HT	HT Clas.	83,35 ± 0,45					
	MC Clas.	47,09 ± 0,06					
	NB Clas.	62,83 ± 0,03					
	TA Clas.	86,43 ± 0,28					
	PEDD*		5739,21 ± 935,53	17562,74 ± 504,53	0,17 ± 0,04	3,17 ± 0,64	
	RQADD*		3701,56 ± 523,14	9727,86 ± 1038,64	0,00 ± 0,00	2,78 ± 0,27	
	SEDD*		11642,27 ± 636,37	11163,85 ± 393,12	0,00 ± 0,00	0,97 ± 0,05	
	ADWIN	67,66 ± 0,04	7673,35 ± 1101,31	15919,70 ± 684,45	0,20 ± 0,04	2,07 ± 0,48	
	CUSUM	82,94 ± 0,51	23333,25 ± 0,00	24101,23 ± 1379,55	1,00 ± 0,00	0,00 ± 0,00	
	DDM	83,21 ± 0,48	23333,25 ± 0,00	22692,48 ± 1509,38	1,00 ± 0,00	0,00 ± 0,00	
	ECDD	80,55 ± 0,98	20142,11 ± 976,70	23155,43 ± 1746,29	0,81 ± 0,06	0,23 ± 0,06	
	EDDM	82,27 ± 0,56	23333,25 ± 0,00	18841,50 ± 130,78	1,00 ± 0,00	0,00 ± 0,00	
	FHDDM	81,54 ± 1,25	22846,03 ± 503,26	22838,78 ± 1386,51	0,96 ± 0,04	0,04 ± 0,04	
	FTDD	67,52 ± 0,14	12101,38 ± 660,75	22643,09 ± 409,69	0,25 ± 0,02	1,43 ± 0,09	
	HDDMA	79,40 ± 0,69	16886,26 ± 979,69	23720,68 ± 1439,55	0,70 ± 0,05	0,47 ± 0,15	
	MDDMA	79,90 ± 1,17	20307,00 ± 791,58	24355,48 ± 1302,07	0,78 ± 0,05	0,28 ± 0,06	
	PHT	83,00 ± 0,51	23333,25 ± 0,00	24613,36 ± 1293,38	1,00 ± 0,00	0,00 ± 0,00	
	RDDM	79,52 ± 0,55	17146,88 ± 769,94	16974,55 ± 1894,03	0,68 ± 0,05	0,32 ± 0,07	
	SEED	67,45 ± 0,07	7261,40 ± 999,29	13006,85 ± 487,60	0,22 ± 0,03	1,69 ± 0,35	
	SEQD2	67,52 ± 0,09	11090,80 ± 1205,06	21894,38 ± 504,42	0,18 ± 0,05	1,91 ± 0,39	
	STEPD	67,08 ± 0,12	10191,30 ± 882,49	19320,26 ± 966,06	0,18 ± 0,04	1,66 ± 0,21	
	WSTD	76,15 ± 2,33	20455,98 ± 1364,61	23588,91 ± 927,55	0,80 ± 0,09	0,27 ± 0,12	
	PEDD	69,24 ± 0,52	6098,49 ± 898,40	16030,31 ± 372,35	0,09 ± 0,05	3,33 ± 1,02	
	RQADD	75,63 ± 1,23	20280,65 ± 1539,18	22634,21 ± 343,88	0,81 ± 0,07	0,27 ± 0,12	
	SEDD	78,56 ± 1,32	17559,93 ± 1599,91	26218,54 ± 491,24	0,64 ± 0,08	0,64 ± 0,20	
	NB	ADWIN	67,65 ± 0,03	6280,35 ± 1159,62	16046,15 ± 532,09	0,16 ± 0,05	3,79 ± 1,37
		CUSUM	62,83 ± 0,03	23333,25 ± 0,00	24324,08 ± 1359,09	1,00 ± 0,00	0,00 ± 0,00
		DDM	62,82 ± 0,03	23333,25 ± 0,00	22422,26 ± 1507,86	1,00 ± 0,00	0,00 ± 0,00
ECDD		62,82 ± 0,03	23333,25 ± 0,00	24273,08 ± 1388,57	1,00 ± 0,00	0,00 ± 0,00	
EDDM		62,87 ± 0,03	23333,25 ± 0,00	18741,62 ± 52,93	1,00 ± 0,00	0,00 ± 0,00	
FHDDM		64,19 ± 0,59	21897,08 ± 627,92	23230,90 ± 1492,28	0,89 ± 0,05	0,12 ± 0,05	
FTDD		65,26 ± 0,27	19600,76 ± 295,62	25869,94 ± 907,69	0,75 ± 0,00	0,33 ± 0,01	
HDDMA		64,24 ± 0,09	20462,72 ± 274,19	23499,11 ± 1473,46	0,58 ± 0,04	0,48 ± 0,06	
MDDMA		64,39 ± 0,46	20824,17 ± 596,71	24811,34 ± 1277,22	0,68 ± 0,04	0,39 ± 0,07	
PHT		62,83 ± 0,03	23333,25 ± 0,00	24858,02 ± 1245,14	1,00 ± 0,00	0,00 ± 0,00	
RDDM		64,12 ± 0,04	20908,62 ± 66,28	21674,52 ± 1443,12	0,74 ± 0,02	0,27 ± 0,02	
SEED		67,41 ± 0,06	6815,72 ± 1112,96	13472,65 ± 528,31	0,31 ± 0,05	1,85 ± 0,52	
SEQD2		66,93 ± 0,08	13291,22 ± 1347,40	24811,36 ± 1052,08	0,57 ± 0,05	0,94 ± 0,21	
STEPD		66,32 ± 0,25	12686,34 ± 1199,40	18993,24 ± 1713,34	0,52 ± 0,04	0,75 ± 0,11	
WSTD		62,99 ± 0,07	22917,40 ± 158,21	25339,65 ± 1126,89	0,88 ± 0,05	0,14 ± 0,05	
PEDD		66,25 ± 0,23	8650,47 ± 895,74	16375,87 ± 890,13	0,28 ± 0,03	1,50 ± 0,22	
RQADD		66,21 ± 0,20	15071,23 ± 1130,75	24069,73 ± 316,55	0,38 ± 0,06	1,04 ± 0,15	
SEDD	64,25 ± 0,51	20349,27 ± 1159,00	26666,60 ± 0,00	0,68 ± 0,04	0,46 ± 0,10		

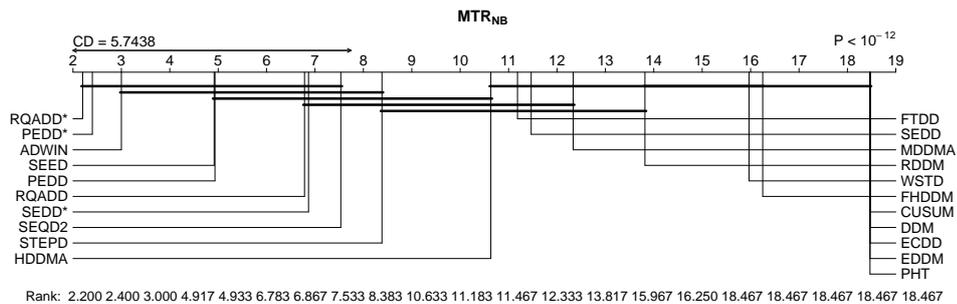
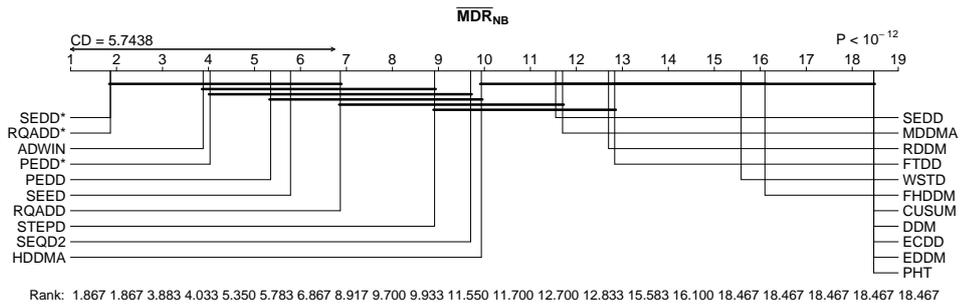
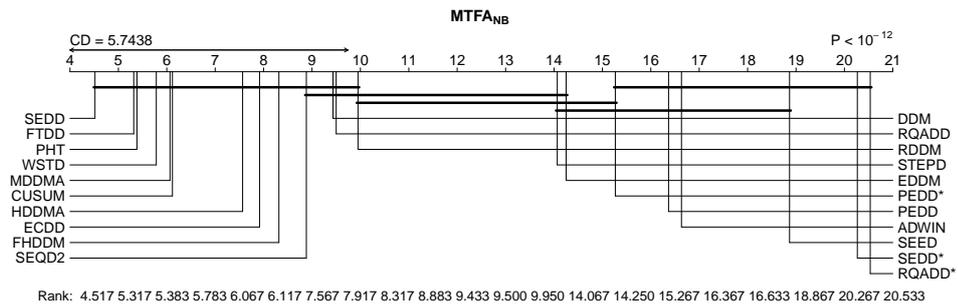
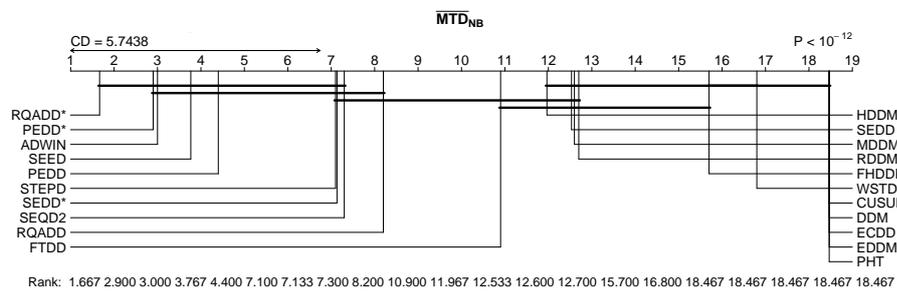
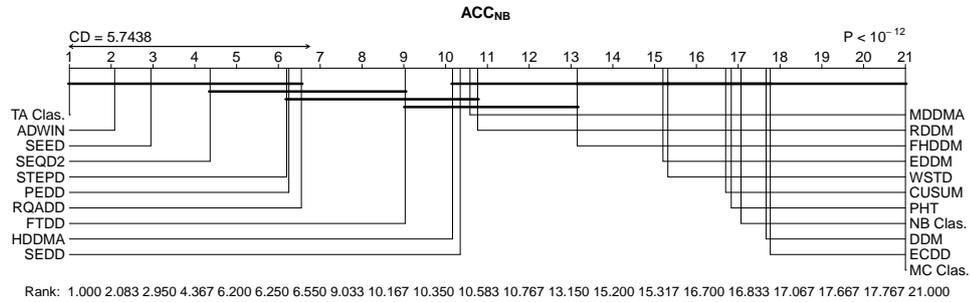
Fonte: Elaborada pelo autor (2021).

Figura 38 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 3$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 39 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 3$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

com $m = 3$, e o PHT ficou no topo do ranqueamento com $m = 4$ e $m = 5$. Esses gráficos mostram que os detectores DDM, ECDD, FHDDM, FTDD, HDDMA, PHT, WSTD, e o SEDD entre os detectores propostos, foram competitivos nas parametrizações com $m = 3$, $m = 4$, e $m = 5$. Os detectores MDDMA, CUSUM, SEQD2, e o RQADD, entre os detectores propostos, foram competitivos somente com $m = 3$, usando ambos os classificadores HT e NB. Nessa parametrização, o RDDM também foi competitivo com o NB.

Os detectores PEDD*, MDDMA, CUSUM, e SEQD2 foram competitivos com ambos os classificadores HT e NB, considerando $m = 4$. Entre os detectores propostos, apenas o PEDD usando NB foi competitivo. Somente os detectores RDDM, RQADD e o PEDD, entre os detectores propostos, foram competitivos usando ambos os classificadores HT e NB. Os detectores CUSUM e MDDMA foram competitivos usando NB.

Com relação ao MDR, o SEDD* obteve o melhor ranqueamento em todas as parametrizações ($m = 3$, $m = 4$, e $m = 5$), e os detectores ADWIN, PEDD*, RQADD*, e SEED foram competitivos na parametrização com $m = 3$ usando ambos os classificadores HT e NB, apresentados nos gráficos de CD MDR_{HT} e MDR_{NB} das Figuras 38 e 39, respectivamente. Com apenas o HT, os detectores FTDD, SEQD2, e STEPDP também foram competitivos. Entre os detectores propostos, somente o PEDD e o RQADD foram competitivos, o primeiro usando ambos os classificadores HT e NB, e o segundo usando o NB. Considerando a parametrização $m = 4$, os detectores PEDD*, RQADD*, SEDD*, ADWIN, e SEED foram competitivos usando ambos os classificadores HT e NB, como apontam os gráficos de CD \overline{MDR}_{HT} e \overline{MDR}_{NB} (Figuras 40 e 41). Nessa parametrização, apenas o RQADD e o PEDD foram competitivos entre os detectores propostos. Nesse caso o RQADD com ambos os classificadores, e o PEDD somente com o HT. O SEQD2 também foi competitivo com o HT nessa parametrização. Com a parametrização $m = 5$, usando ambos os classificadores HT e NB, somente os detectores PEDD*, RQADD*, ADWIN, MDDMA, SEED, e SEQD2 foram competitivos, como mostram os gráficos de CD \overline{MDR}_{HT} e \overline{MDR}_{NB} (Figuras 42 e 43). Nessa parametrização, nenhum dos detectores propostos foi competitivo.

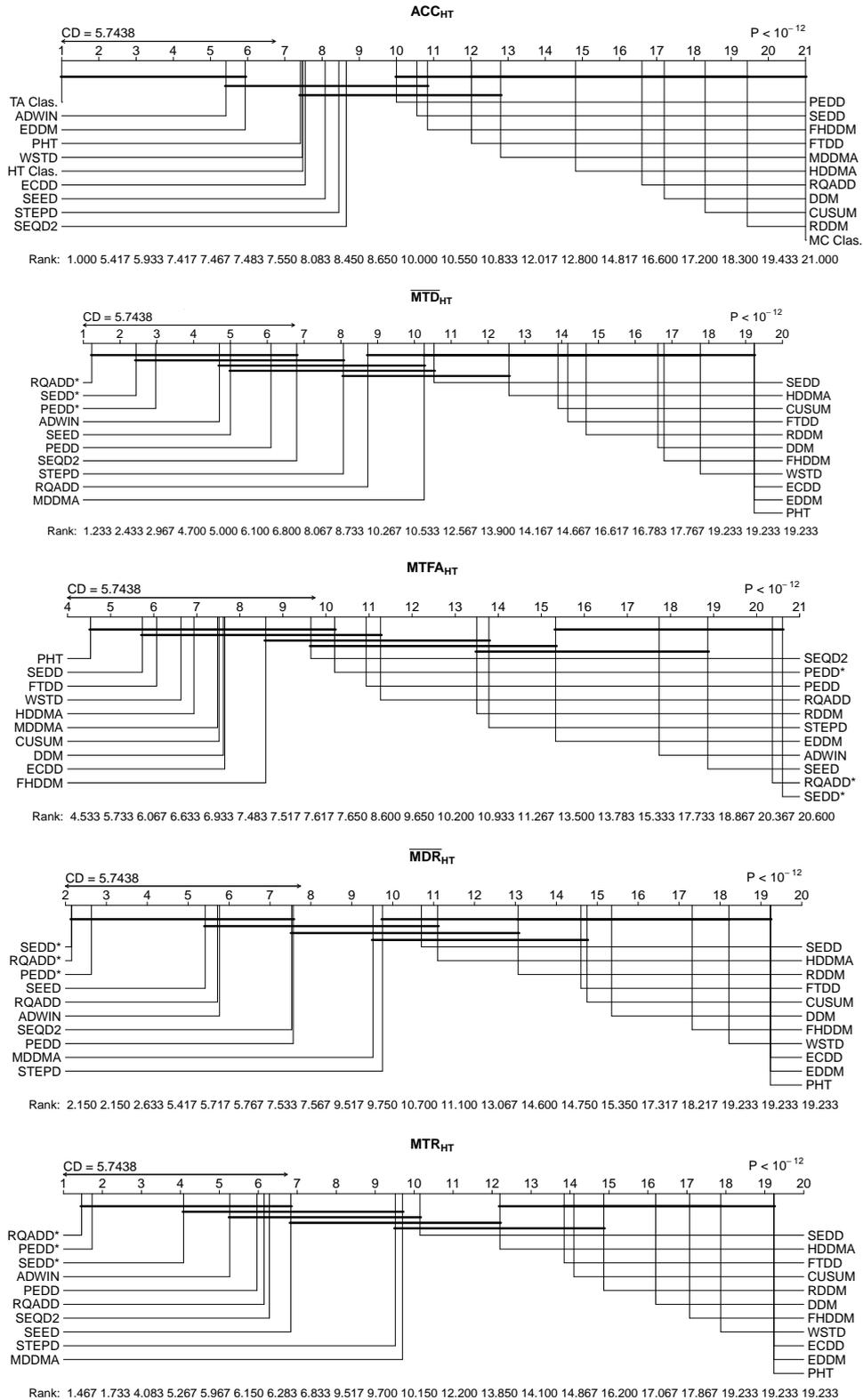
Em relação ao MTR, o RQADD* obteve o melhor ranqueamento em todas as parametrizações ($m = 3$, $m = 4$, e $m = 5$). Os gráficos de CD \overline{MDR}_{HT} e \overline{MDR}_{NB} (Figuras 38 e 39) referentes à parametrização $m = 3$ mostram que os detectores PEDD*, ADWIN, SEED, e SEQD2 foram competitivos com ambos os classificadores HT e NB. Os detectores FTDD e STEPDP foram competitivos somente com o HT, e o SEDD* com NB. Entre os detectores propostos, somente o PEDD foi competitivo com HT, e o RQADD usando NB. De modo seme-

Tabela 22 – Resultados obtidos usando HT e NB a partir das rotulações simbólicas da base artificial do Sistema Lorenz Transiente, considerando a dimensão de imersão $m = 4$ e a defasagens $\tau = 3$, com um intervalo de confiança de 95%.

$Lorenz_{(d=4,\tau=3)}$	Classificador/Detector _(HT)	Acurácia(%)	MTD	MTFA	MDR	MTR	
HT	HT Clas.	63,94 ± 0,33					
	MC Clas.	41,20 ± 0,06					
	NB Clas.	60,85 ± 0,05					
	TA Clas.	72,93 ± 0,56					
	PEDD*		3917,78 ± 213,55	17570,77 ± 499,04	0,03 ± 0,03	4,45 ± 0,32	
	RQADD*		1624,63 ± 281,31	7390,73 ± 931,84	0,00 ± 0,00	4,88 ± 0,51	
	SEDD*		3358,66 ± 557,40	6198,00 ± 454,47	0,00 ± 0,00	2,15 ± 0,31	
	ADWIN	64,25 ± 0,12	6026,40 ± 746,81	12134,98 ± 273,87	0,25 ± 0,00	1,73 ± 0,27	
	CUSUM	60,66 ± 0,07	15487,38 ± 140,38	18320,13 ± 986,93	0,75 ± 0,00	0,30 ± 0,02	
	DDM	61,07 ± 0,45	16946,21 ± 104,43	18010,73 ± 1068,45	0,78 ± 0,03	0,23 ± 0,04	
	ECDD	63,94 ± 0,33	17500,00 ± 0,00	18404,80 ± 1004,67	1,00 ± 0,00	0,00 ± 0,00	
	EDDM	64,14 ± 0,30	17500,00 ± 0,00	14097,01 ± 87,41	1,00 ± 0,00	0,00 ± 0,00	
	FHDDM	62,93 ± 0,63	16427,69 ± 521,74	17537,73 ± 1084,51	0,89 ± 0,05	0,12 ± 0,05	
	FTDD	62,89 ± 0,38	15539,08 ± 219,63	18805,37 ± 907,41	0,74 ± 0,02	0,31 ± 0,02	
	HDDMA	62,52 ± 0,36	14032,73 ± 582,79	18460,06 ± 954,30	0,55 ± 0,04	0,62 ± 0,08	
	MDDMA	63,02 ± 0,05	11753,86 ± 539,61	18383,79 ± 1010,83	0,47 ± 0,03	0,87 ± 0,13	
	PHT	63,97 ± 0,34	17500,00 ± 0,00	19811,94 ± 384,63	1,00 ± 0,00	0,00 ± 0,00	
	RDDM	60,61 ± 0,06	15657,01 ± 366,20	15607,23 ± 1019,66	0,66 ± 0,05	0,35 ± 0,06	
	SEED	63,91 ± 0,20	6820,10 ± 476,33	10461,25 ± 288,76	0,22 ± 0,03	1,35 ± 0,38	
	SEQD2	63,63 ± 0,21	8345,58 ± 651,62	17392,70 ± 766,22	0,35 ± 0,06	1,49 ± 0,27	
	STEPD	63,65 ± 0,19	9638,56 ± 975,07	15197,43 ± 1051,78	0,48 ± 0,05	0,93 ± 0,19	
	WSTD	63,85 ± 0,40	16622,33 ± 684,86	18607,49 ± 958,65	0,94 ± 0,05	0,09 ± 0,08	
	PEDD	63,34 ± 0,05	7677,35 ± 473,26	16863,58 ± 520,44	0,35 ± 0,06	1,54 ± 0,28	
	RQADD	61,18 ± 0,09	9970,57 ± 481,93	17030,25 ± 395,10	0,24 ± 0,05	1,33 ± 0,13	
	SEDD	63,27 ± 0,18	11933,96 ± 455,00	19511,44 ± 424,32	0,53 ± 0,04	0,78 ± 0,07	
	NB	ADWIN	64,20 ± 0,12	6224,27 ± 787,53	12321,77 ± 339,58	0,25 ± 0,00	1,71 ± 0,27
		CUSUM	60,66 ± 0,06	15487,99 ± 140,46	18341,35 ± 970,86	0,75 ± 0,00	0,30 ± 0,02
		DDM	60,67 ± 0,07	16883,48 ± 92,54	18015,06 ± 1066,17	0,77 ± 0,02	0,25 ± 0,03
ECDD		60,85 ± 0,05	17500,00 ± 0,00	18404,77 ± 1004,69	1,00 ± 0,00	0,00 ± 0,00	
EDDM		60,87 ± 0,05	17500,00 ± 0,00	14068,29 ± 52,80	1,00 ± 0,00	0,00 ± 0,00	
FHDDM		60,84 ± 0,10	16700,78 ± 513,91	17537,73 ± 1084,51	0,92 ± 0,04	0,09 ± 0,05	
FTDD		62,99 ± 0,43	14808,09 ± 581,65	18509,60 ± 1001,81	0,76 ± 0,02	0,31 ± 0,03	
HDDMA		62,45 ± 0,38	14159,79 ± 610,93	18369,37 ± 959,64	0,56 ± 0,04	0,60 ± 0,08	
MDDMA		63,03 ± 0,05	11753,86 ± 539,61	18379,45 ± 1013,52	0,47 ± 0,03	0,87 ± 0,13	
PHT		60,85 ± 0,05	17500,00 ± 0,00	19243,39 ± 732,80	1,00 ± 0,00	0,00 ± 0,00	
RDDM		60,62 ± 0,06	15626,54 ± 350,61	14855,36 ± 1035,18	0,67 ± 0,04	0,32 ± 0,06	
SEED		63,86 ± 0,21	6753,43 ± 567,40	10420,82 ± 288,04	0,22 ± 0,03	1,38 ± 0,38	
SEQD2		63,59 ± 0,19	8593,92 ± 566,46	17297,29 ± 854,90	0,35 ± 0,06	1,39 ± 0,22	
STEPD		63,47 ± 0,17	9662,06 ± 981,29	14981,86 ± 1006,16	0,48 ± 0,05	0,90 ± 0,19	
WSTD		61,18 ± 0,33	16760,76 ± 647,00	18607,41 ± 958,70	0,95 ± 0,05	0,07 ± 0,07	
PEDD		63,35 ± 0,05	7677,35 ± 473,26	16863,58 ± 520,44	0,35 ± 0,06	1,54 ± 0,28	
RQADD		61,17 ± 0,08	10094,72 ± 668,66	16857,18 ± 391,59	0,25 ± 0,05	1,30 ± 0,13	
SEDD		63,07 ± 0,17	11932,11 ± 454,11	19511,44 ± 424,32	0,53 ± 0,04	0,78 ± 0,07	

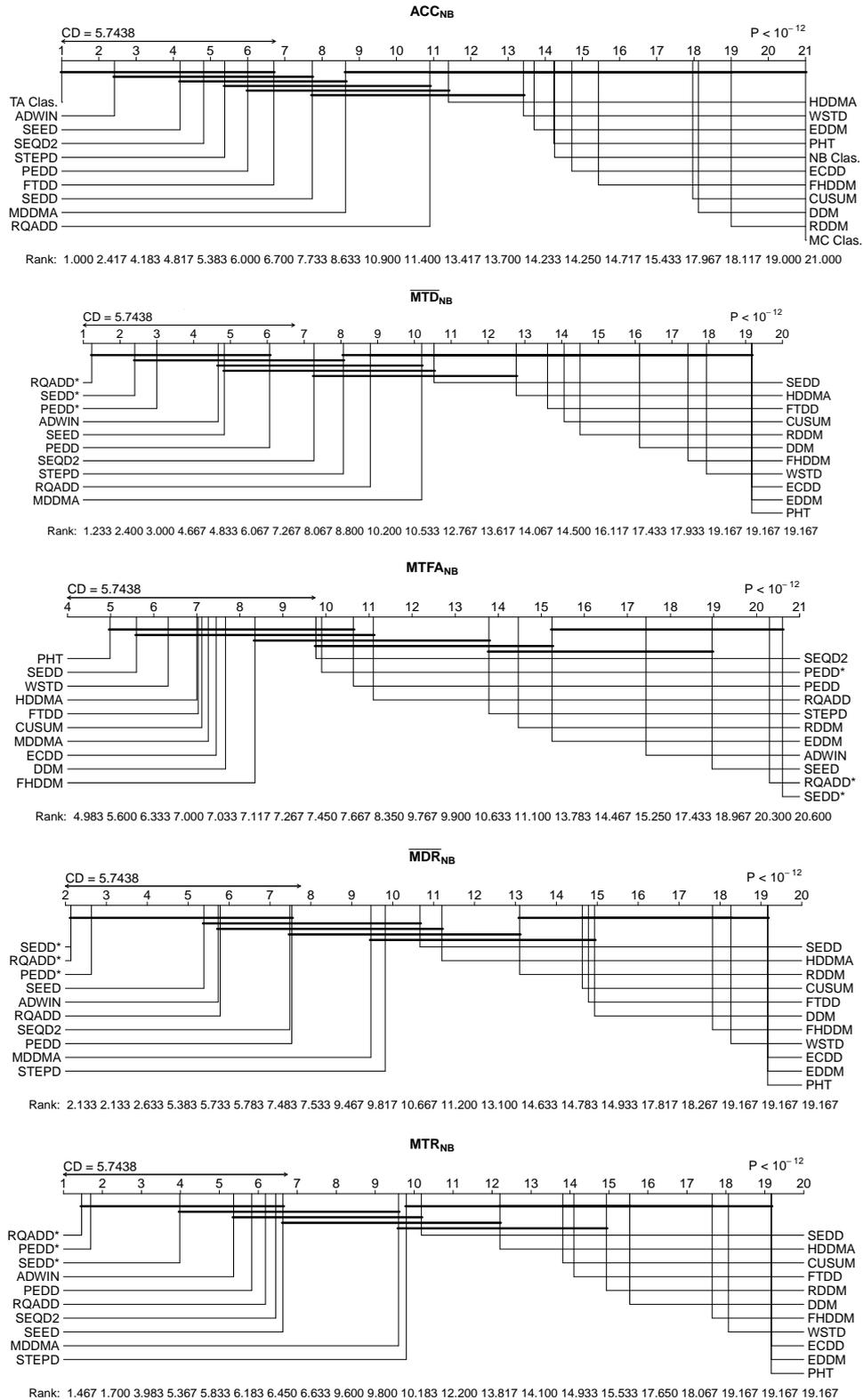
Fonte: Elaborada pelo autor (2021).

Figura 40 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 4$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 41 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 4$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.



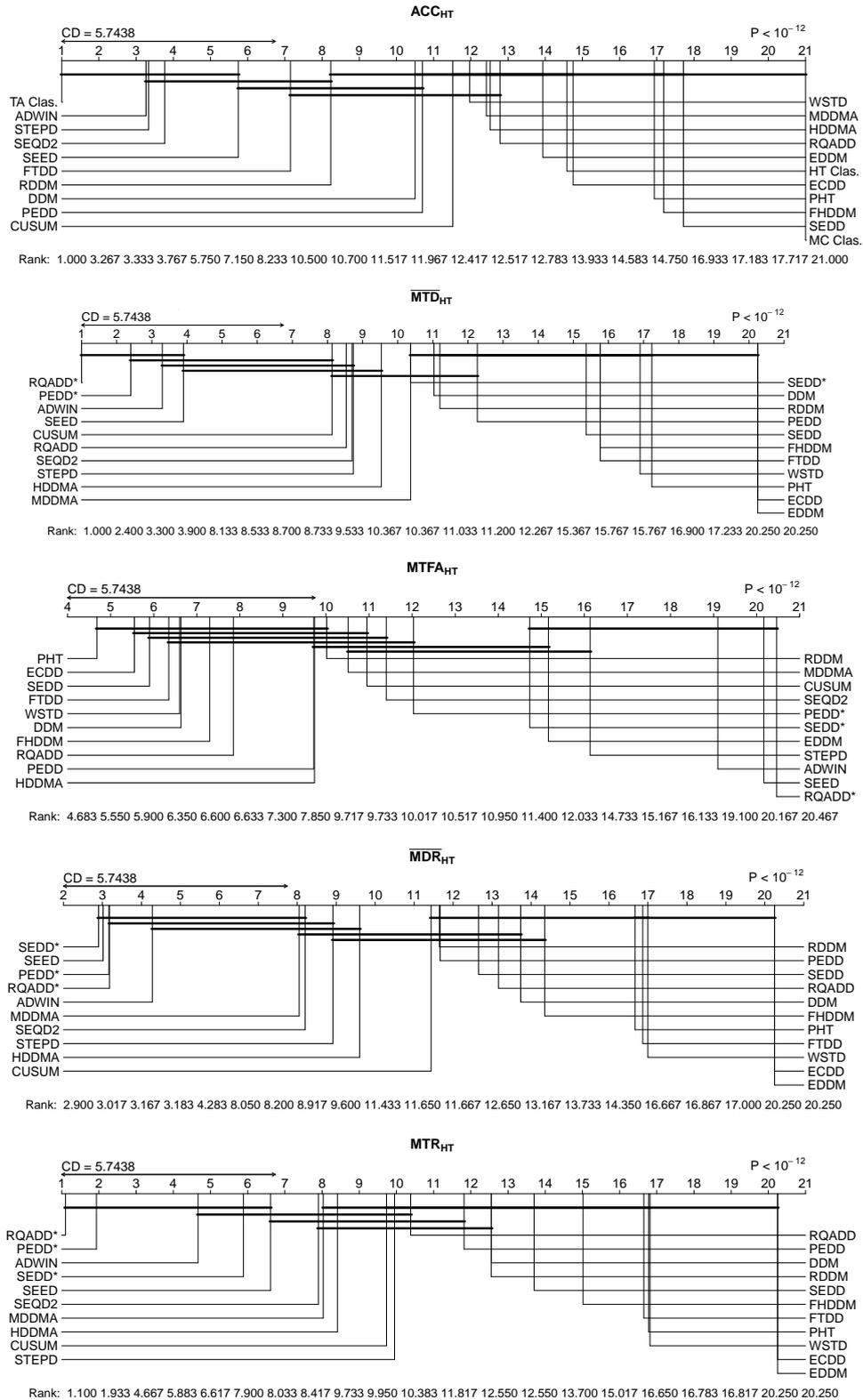
Fonte: Elaborada pelo autor (2021).

Tabela 23 – Resultados obtidos usando HT e NB a partir das rotulações simbólicas da base artificial do Sistema Lorenz Transiente, considerando a dimensão de imersão $m = 5$ e a defasagens $\tau = 3$, com um intervalo de confiança de 95%.

$Lorenz_{(d=5,\tau=3)}$	Classificador/Detector(HT)	Acurácia(%)	MTD	MTFA	MDR	MTR
HT	HT Clas.	55,24 ± 0,06				
	MC Clas.	35,36 ± 0,07				
	NB Clas.	55,25 ± 0,06				
	TA Clas.	60,82 ± 0,12				
	PEDD*		4931,48 ± 185,26	13621,85 ± 558,76	0,02 ± 0,02	2,75 ± 0,16
	RQADD*		1358,07 ± 329,49	6897,60 ± 639,20	0,02 ± 0,02	6,38 ± 1,19
	SEDD*		9747,54 ± 398,34	11650,04 ± 536,48	0,00 ± 0,00	1,21 ± 0,07
	ADWIN	57,46 ± 0,20	5796,22 ± 558,27	8672,74 ± 321,20	0,08 ± 0,04	1,45 ± 0,14
	CUSUM	55,57 ± 0,11	9021,33 ± 209,30	14358,11 ± 434,41	0,47 ± 0,03	0,86 ± 0,08
	DDM	55,52 ± 0,30	10231,63 ± 739,06	14825,81 ± 762,60	0,60 ± 0,07	0,63 ± 0,12
	ECDD	55,23 ± 0,06	14000,00 ± 0,00	15362,33 ± 617,44	1,00 ± 0,00	0,00 ± 0,00
	EDDM	55,28 ± 0,07	14000,00 ± 0,00	11869,05 ± 615,49	1,00 ± 0,00	0,00 ± 0,00
	FHDDM	53,54 ± 0,47	12418,71 ± 621,60	14793,16 ± 708,08	0,61 ± 0,06	0,50 ± 0,11
	FTDD	56,49 ± 0,49	12526,98 ± 436,23	15127,15 ± 702,20	0,73 ± 0,02	0,33 ± 0,05
	HDDMA	55,01 ± 0,48	9796,07 ± 642,47	14949,57 ± 339,19	0,38 ± 0,06	1,00 ± 0,14
	MDDMA	54,86 ± 0,57	10057,15 ± 627,28	14445,26 ± 557,52	0,29 ± 0,06	1,07 ± 0,15
	PHT	54,49 ± 0,24	13153,94 ± 348,05	15562,21 ± 498,83	0,72 ± 0,03	0,33 ± 0,04
	RDDM	55,82 ± 0,10	10115,42 ± 235,83	13660,02 ± 932,68	0,48 ± 0,03	0,71 ± 0,06
	SEED	56,41 ± 0,17	6504,08 ± 340,77	7442,42 ± 375,21	0,01 ± 0,02	1,15 ± 0,08
	SEQD2	57,28 ± 0,31	8907,85 ± 607,66	13500,82 ± 718,83	0,30 ± 0,06	1,09 ± 0,13
STEPD	57,49 ± 0,34	8983,65 ± 658,10	11527,50 ± 814,54	0,34 ± 0,06	0,88 ± 0,13	
WSTD	55,49 ± 0,44	12688,40 ± 617,83	15097,66 ± 636,78	0,78 ± 0,07	0,31 ± 0,11	
PEDD	55,50 ± 0,25	10763,42 ± 678,17	14660,88 ± 348,95	0,48 ± 0,07	0,75 ± 0,13	
RQADD	55,21 ± 0,42	8875,58 ± 859,57	15409,67 ± 286,17	0,55 ± 0,06	0,92 ± 0,25	
SEDD	53,76 ± 0,27	12314,40 ± 263,17	15669,93 ± 199,29	0,52 ± 0,03	0,61 ± 0,04	
NB	ADWIN	57,49 ± 0,19	5973,28 ± 545,46	8799,56 ± 297,79	0,07 ± 0,04	1,45 ± 0,14
	CUSUM	55,58 ± 0,11	9015,87 ± 238,06	14424,15 ± 442,37	0,47 ± 0,03	0,87 ± 0,09
	DDM	55,60 ± 0,26	9949,77 ± 676,16	14711,57 ± 813,97	0,57 ± 0,07	0,67 ± 0,11
	ECDD	55,24 ± 0,06	14000,00 ± 0,00	15362,33 ± 617,44	1,00 ± 0,00	0,00 ± 0,00
	EDDM	55,28 ± 0,06	14000,00 ± 0,00	11729,85 ± 541,05	1,00 ± 0,00	0,00 ± 0,00
	FHDDM	53,55 ± 0,47	12418,71 ± 621,60	14793,16 ± 708,08	0,61 ± 0,06	0,50 ± 0,11
	FTDD	56,49 ± 0,49	12526,98 ± 436,23	14968,28 ± 747,39	0,73 ± 0,02	0,33 ± 0,05
	HDDMA	55,10 ± 0,51	9663,58 ± 684,38	14779,90 ± 358,98	0,38 ± 0,07	1,02 ± 0,15
	MDDMA	54,86 ± 0,57	10057,15 ± 627,28	14445,22 ± 557,53	0,29 ± 0,06	1,07 ± 0,15
	PHT	54,55 ± 0,25	13038,77 ± 415,76	15429,60 ± 553,37	0,72 ± 0,03	0,34 ± 0,05
	RDDM	55,84 ± 0,11	10107,86 ± 237,64	13777,03 ± 846,18	0,48 ± 0,03	0,72 ± 0,06
	SEED	56,43 ± 0,17	6512,08 ± 339,56	7471,82 ± 386,91	0,01 ± 0,02	1,16 ± 0,08
	SEQD2	57,29 ± 0,31	8907,85 ± 607,66	13500,82 ± 718,83	0,30 ± 0,06	1,09 ± 0,13
	STEPD	57,50 ± 0,34	8983,65 ± 658,10	11528,28 ± 814,26	0,34 ± 0,06	0,88 ± 0,13
	WSTD	55,50 ± 0,44	12688,40 ± 617,83	15257,21 ± 577,22	0,78 ± 0,07	0,31 ± 0,11
	PEDD	55,50 ± 0,25	10763,42 ± 678,17	14660,88 ± 348,95	0,48 ± 0,07	0,75 ± 0,13
RQADD	55,19 ± 0,43	9077,31 ± 971,64	15391,30 ± 296,32	0,58 ± 0,06	0,87 ± 0,26	
SEDD	53,82 ± 0,29	12372,69 ± 272,51	15652,63 ± 198,30	0,53 ± 0,03	0,59 ± 0,05	

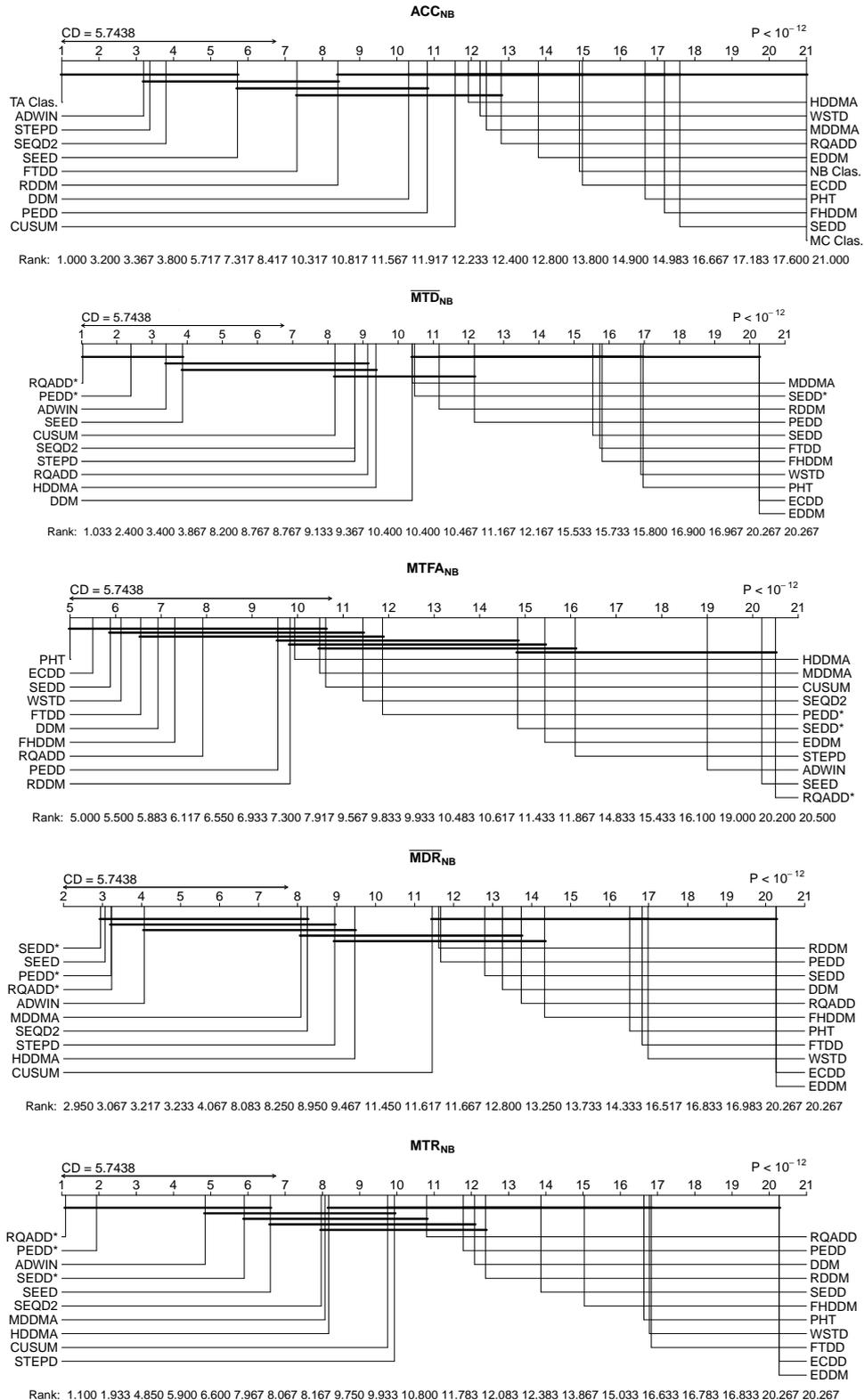
Fonte: Elaborada pelo autor (2021).

Figura 42 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 5$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 43 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi nas bases de dados com rotulações simbólicas com a dimensão de imersão $m = 5$ e a defasagens $\tau = 3$ geradas a partir do Sistema de Lorenz Transiente, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

lhante, os gráficos de CD \overline{MDR}_{HT} e \overline{MDR}_{NB} (Figuras 40 e 41) referentes à parametrização $m = 4$ mostram que os detectores PEDD*, SEDD*, ADWIN, SEED, SEQD2, bem como os detectores propostos PEDD e RQADD, foram competitivos com ambos os classificadores HT e NB. Com a parametrização $m = 5$, os gráficos de CD \overline{MDR}_{HT} e \overline{MDR}_{NB} (Figuras 42 e 43) mostram que os detectores PEDD*, SEDD*, ADWIN, e SEED foram competitivos com ambos os classificadores HT e NB. Nessa caso, nenhum dos detectores foi competitivo.

Esses resultados sugerem que as mudanças na dinâmica desses fluxos de dados podem ser identificadas por meio dos erros das predições das instâncias fornecida pelo SLA, ao considerar a rotulação simbólica dos estados do espaço de fase, os quais também apontam que os detectores acoplados aos classificados HT e NB apresentam desempenho competitivo em termos da detecção de mudanças na dinâmica do *fluxo de dados* em comparação ao estado da arte (Entropia de Permutação (PE), Análise de Quantificação de Recorrência (RQA), e Entropia Espectral (SE)) usados na avaliação direta dos dados brutos. Assim, o SLA pode ser usado para expandir a aplicação de *modelos adaptativos supervisionados* a contextos com *fluxos de dados* não rotulados. Vale salientar que a indisponibilidade de dados rotulados é um fator limitante para aplicações desses modelos no mundo real (DYER; CAPO; POLIKAR, 2014; VALLIM; MELLO, 2014; COSTA; RIOS; MELLO, 2016; SETHI; KANTARDZIC, 2017).

5.6 RESULTADOS E ANÁLISE DO EXPERIMENTO COM BASES DE EEG

Esta seção apresenta a análise dos resultados utilizando as 26 bases de EEG descritas na Seção 5.1.2.1. Semelhante ao experimento com as bases geradas com o Sistema de Lorenz Transiente, o SLA também foi usado para prover os rótulos com o objetivo de identificar *mudanças de conceitos*. Novamente, além dos detectores acoplados ao HT e NB, também foram utilizados o *Classificador de Classe Majoritária*, o *Classificador Temporalmente Aumentado*, e os classificadores HT e NB sem detectores. Também foram utilizados os detectores PEDD*, RQADD*, e SEDD*, sem os classificadores HT e NB, e com as mesmas parametrizações do experimento com as bases do Sistema de Lorenz Transiente. O objetivo é avaliar se os detectores acoplados ao HT e NB são competitivos em comparação à Entropia de Permutação (PE), à Análise de Quantificação de Recorrência (RQA), e à Entropia Espectral (SE).

Vale salientar que neste experimento o parâmetro dos retardos temporais foi definido como $\tau = 1$, ou seja, não foi estimado com a AMIF. Experimentos preliminares mostraram que nessa base, a estimação forneceu amostras muito pequenas, pois considerando N como o número

de observações, m a dimensão de reconstrução do espaço de fase, e τ a defasagem temporal, $N/\tau m$ define a quantidade de instâncias obtidas com o SLA. Isso é devido à AMIF fornecer os primeiros mínimos locais com defasagens elevadas para essas bases de EEG. Assim, a reconstrução do espaço de fase dos sinais de EEG foram definidos com a dimensão $m = 4$ e o retardo temporal $\tau = 1$. Para a avaliação dos detectores PEDD*, RQADD*, e SEDD*, também foram utilizados apenas o primeiro componente $x_{i+km\tau}$ dos vetores $\mathbf{v}_{(i+km\tau)}^{m,\tau}$, em cada instância $(\mathbf{v}_{(i+km\tau)}^{m,\tau}, \boldsymbol{\pi}_{(i+(k+1)m\tau)}^m)$, onde $k = 0, 1, 2, \dots$.

A Tabela 24 apresenta os resultados obtidos usando os classificadores isoladamente, e com os detectores acoplados ao HT e NB. Nesta tabela, os melhores resultados também foram destacados em **negrito**. Note que o *Classificador Temporalmente Aumentado* obteve o melhor resultado em termos de acurácia entre todos os classificadores e detectores, indicando uma dependência temporal significativa entre os rótulos produzidos com o SLA. Além disso, com essa parametrização os detectores não apresentaram uma acurácia significativa devido a um fraco determinismo entre os estados dos espaços de fase. Devido a dependência temporal, os resultados apontam que a maioria dos detectores com os menores valores para a métrica MTFA alcançaram as melhores acurácias. Uma exceção foi o EDDM com o classificador HT. As Figuras 44 e 45 abrangem os resultados dos testes de Nemenyi e Friedman aplicados aos detectores usando os classificadores HT e NB, respectivamente, bem como os demais classificadores.

Os gráficos de CD ACC_{HT} (Figura 44) e ACC_{NB} (Figura 45) mostram que o *Classificador Temporalmente Aumentado* (TA Clas.) obteve a melhor acurácia, tanto em comparação aos detectores usando HT quanto usando NB. Os detectores EDDM, FHDDM, HDDMA, E STEPDP foram competitivos com ambos os classificadores HT e NB, enquanto ECDD foi competitivo com o HT, e o RDDM com NB. Entretanto, nenhum dos três detectores propostos foram competitivos com relação à acurácia. É importante destacar que, assim como a rotulação dos fluxos de dados do Sistema de Lorenz Transiente, não há um sentido semântico desta rotulação com respeito aos padrões ordinais obtidos dos sinais de EEG, a não ser o próprio padrão simbólico em si. De qualquer modo, os erros das predições são utilizados com um meio indireto para identificar mudanças globais entre as relações dos estados do espaço de fase.

Os gráficos de CD \overline{MTD}_{HT} (Figura 44) e \overline{MTD}_{NB} (Figura 45) mostram que o STEPDP obteve o melhor resultado para a métrica MTD, usando ambos os classificadores HT e NB. Esse detector não foi estatisticamente diferente dos detectores EDDM, FHDDM, FTDD, HDDMA, MDDMA, RDDM, e WSTD, com ambos os classificadores HT e NB. Além disso, os resul-

tados com a Entropia Espectral (SE), representada pelo detector SEDD*, também foram competitivos com HT e NB.

Os resultados dos testes estatísticos para a medida MTFA apresentados nos gráficos de CD $MTFA_{HT}$ (Figura 44) e $MTFA_{NB}$ (Figura 45) indicam que o RQADD* usando o HT e o RQADD usando o NB obtiveram o menor número de alarmes falsos. Esses testes também mostram que os detectores DDM,PHT, e CUSUM, usando ambos os classificadores HT e NB, não apresentaram diferenças estatísticas significativas em relação ao RQADD* e ao RQADD. Com esses classificadores caso, a Entropia de Permutação (PE) representada por PEDD* também foi competitiva, bem como os detectores propostos PEDD e SEDD. O detector SEQD2 também foi competitivo, mas apenas com o HT.

Em relação ao MDR, o STEPDP apresentou a menor taxa de detecções perdidas como aponta os gráficos de CD \overline{MDR}_{HT} (Figura 44) e \overline{MDR}_{NB} (Figura 45). Nessa métrica, apenas o Drift Detection Method (DDM) e o RQADD, entre os três métodos propostos, não foram competitivos usando ambos os classificadores HT e NB.

Com relação ao MTR, o PEDD obteve o melhor desempenho com essa métrica em ambos os classificadores HT e NB, como apresentado nos gráficos de CD MTR_{HT} (Figura 44) e MTR_{NB} (Figura 45). Nesse caso, os detectores MDDMA, CUSUM, SEQD2, RDDM, FHDDM, ADWIN, STEPDP, WSTD, HDDMA não apresentaram diferenças estatísticas significativas em relação ao PEDD, e apenas o SEDD foi competitivo entre os três detectores propostos. A Entropia Espectral (SE) (SEDD*) e a Entropia de Permutação (PE) (PEDD*) também não apresentaram diferenças estatísticas significativas em relação ao PEDD. Além desses detectores, o SEED foi competitivo com o HT, e o EDDM foi competitivo com o NB.

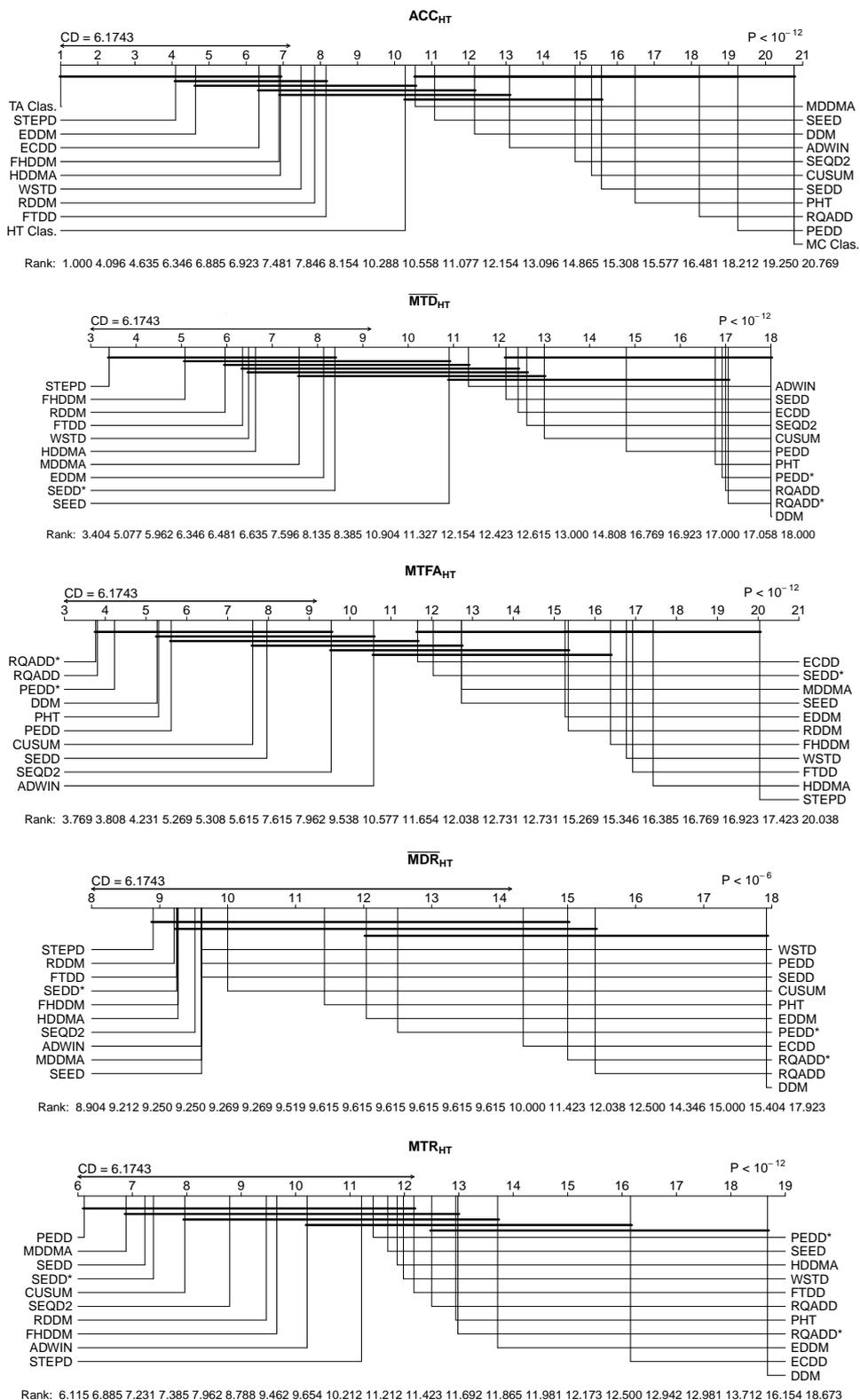
Ao considerar a rotulação simbólica usando padrões ordinais dos estados obtidos por meio do Symbolic Labeling Adapter (SLA), os detectores acoplados aos classificadores HT e NB apresentam melhor desempenho em termos da detecção de mudanças na dinâmica do *fluxo de dados*. Neste experimento, em termos da métrica MTR, os detectores propostos PEDD e SEDD foram mais competitivos que às respectivas versões sem classificadores PEDD* e SEDD*, referentes à Entropia de Permutação (PE) e à Entropia Espectral (SE). Vale salientar que esses resultados também favorecem a hipótese de que essa pseudo-rotulação pode ser utilizada para identificar *mudanças de conceito* em *fluxos de dados* univariados, contornando limitações relacionadas à rotulação (DYER; CAPO; POLIKAR, 2014; SETHI; KANTARDZIC, 2017).

Tabela 24 – Resultados obtidos usando HT e NB, considerando as 26 bases de EEG, com um intervalo de confiança de 95%.

EEG	Classificador/Detector	Acurácia(%)	MTD	MTFA	MDR	MTR	
	Hoeffding Tree	56,79 ± 8,98					
	Majority Class	10,30 ± 2,13					
	Naive Bayes	44,40 ± 10,94					
	Temporally Augmented Classifier	72,68 ± 9,49					
	PEDD*		7008,22 ± 4806,06	9666,07 ± 6582,42	0,12 ± 0,08	1,37 ± 0,26	
	RQADD*		42109,25 ± 33944,34	49409,45 ± 36462,67	0,34 ± 0,18	1,04 ± 0,40	
	SEDD*		684,92 ± 581,79	915,63 ± 626,85	0,01 ± 0,01	1,80 ± 0,31	
HT	ADWIN	54,92 ± 10,49	3006,61 ± 3554,91	775,04 ± 61,34	0,01 ± 0,02	1,97 ± 0,97	
	CUSUM	53,44 ± 10,37	1115,92 ± 410,52	1682,95 ± 501,68	0,04 ± 0,05	1,92 ± 0,57	
	DDM	55,07 ± 9,61	112032,89 ± 63441,52	84830,02 ± 47545,80	0,65 ± 0,17	0,39 ± 0,30	
	ECDD	58,20 ± 9,89	71951,61 ± 58499,64	63691,99 ± 53877,16	0,42 ± 0,23	0,68 ± 0,45	
	EDDM	59,75 ± 9,89	38854,41 ± 43067,38	30293,09 ± 37218,37	0,26 ± 0,19	2,01 ± 1,39	
	FHDDM	58,20 ± 10,51	257,70 ± 113,34	283,71 ± 21,79	0,01 ± 0,01	1,62 ± 0,50	
	FTDD	57,76 ± 10,70	304,21 ± 137,03	265,76 ± 15,92	0,01 ± 0,01	1,43 ± 0,42	
	HDDMA	58,12 ± 10,92	487,41 ± 423,54	261,72 ± 37,36	0,01 ± 0,01	1,34 ± 0,39	
	MDDMA	57,10 ± 10,65	303,40 ± 112,36	464,01 ± 32,64	0,01 ± 0,02	2,03 ± 0,57	
	PHT	50,72 ± 9,75	6982,43 ± 3801,13	6559,19 ± 3694,43	0,07 ± 0,08	1,44 ± 0,63	
	RDDM	58,02 ± 10,84	368,66 ± 214,42	388,25 ± 151,51	0,01 ± 0,03	1,37 ± 0,22	
	SEED	56,86 ± 10,59	29109,64 ± 51011,38	446,89 ± 30,43	0,01 ± 0,02	1,39 ± 0,61	
	SEQD2	53,80 ± 10,47	714,76 ± 165,17	974,61 ± 124,78	0,01 ± 0,02	1,60 ± 0,30	
	STEPD	59,12 ± 10,57	1748,60 ± 3296,10	163,43 ± 6,46	0,00 ± 0,00	1,38 ± 0,33	
	WSTD	58,15 ± 10,67	504,54 ± 369,59	271,35 ± 16,85	0,01 ± 0,02	1,34 ± 0,48	
		PEDD	47,55 ± 9,40	1725,15 ± 318,15	3076,36 ± 393,48	0,01 ± 0,02	1,94 ± 0,34
		RQADD	45,67 ± 10,46	18989,85 ± 6862,16	27854,15 ± 11880,19	0,26 ± 0,11	1,60 ± 0,84
	SEDD	53,27 ± 10,57	997,23 ± 373,80	1314,12 ± 274,65	0,01 ± 0,02	2,24 ± 0,87	
NB	ADWIN	54,62 ± 10,78	2318,69 ± 3445,56	784,37 ± 73,17	0,01 ± 0,02	2,23 ± 0,93	
	CUSUM	53,27 ± 10,55	1286,33 ± 515,97	1592,46 ± 445,31	0,05 ± 0,06	1,73 ± 0,58	
	DDM	49,78 ± 11,70	89766,29 ± 59124,59	60115,19 ± 43291,00	0,50 ± 0,18	0,56 ± 0,64	
	ECDD	50,79 ± 13,60	75732,21 ± 58870,27	68717,17 ± 53542,84	0,45 ± 0,22	0,88 ± 0,95	
	EDDM	57,83 ± 11,86	19691,61 ± 23251,56	4993,22 ± 5202,34	0,13 ± 0,15	1,75 ± 0,85	
	FHDDM	58,01 ± 10,95	247,59 ± 83,01	296,42 ± 29,62	0,01 ± 0,01	1,60 ± 0,50	
	FTDD	57,61 ± 11,15	326,13 ± 138,84	279,95 ± 21,79	0,01 ± 0,02	1,32 ± 0,43	
	HDDMA	58,03 ± 11,29	267,23 ± 134,98	272,10 ± 38,77	0,01 ± 0,01	1,51 ± 0,40	
	MDDMA	57,04 ± 10,89	327,97 ± 112,70	478,79 ± 49,06	0,01 ± 0,02	1,96 ± 0,57	
	PHT	51,40 ± 10,35	13818,47 ± 8656,07	5461,93 ± 3067,20	0,15 ± 0,10	1,03 ± 0,49	
	RDDM	58,00 ± 11,15	395,09 ± 192,11	376,51 ± 139,86	0,01 ± 0,02	1,25 ± 0,22	
	SEED	56,83 ± 10,84	58235,30 ± 69817,17	451,84 ± 31,40	0,01 ± 0,02	1,00 ± 0,58	
	SEQD2	53,55 ± 10,74	691,07 ± 194,00	929,57 ± 80,69	0,01 ± 0,01	1,67 ± 0,33	
	STEPD	59,08 ± 11,08	135,51 ± 23,34	170,71 ± 7,59	0,00 ± 0,00	1,43 ± 0,26	
	WSTD	57,97 ± 11,11	355,18 ± 232,90	284,56 ± 20,61	0,01 ± 0,02	1,53 ± 0,51	
		PEDD	47,09 ± 9,70	1691,65 ± 345,17	3035,95 ± 420,48	0,01 ± 0,02	1,97 ± 0,32
		RQADD	43,75 ± 11,03	69879,22 ± 52467,12	72678,93 ± 49307,76	0,43 ± 0,17	1,07 ± 0,60
	SEDD	53,01 ± 10,82	1062,38 ± 428,56	1368,67 ± 323,12	0,01 ± 0,02	2,27 ± 0,88	

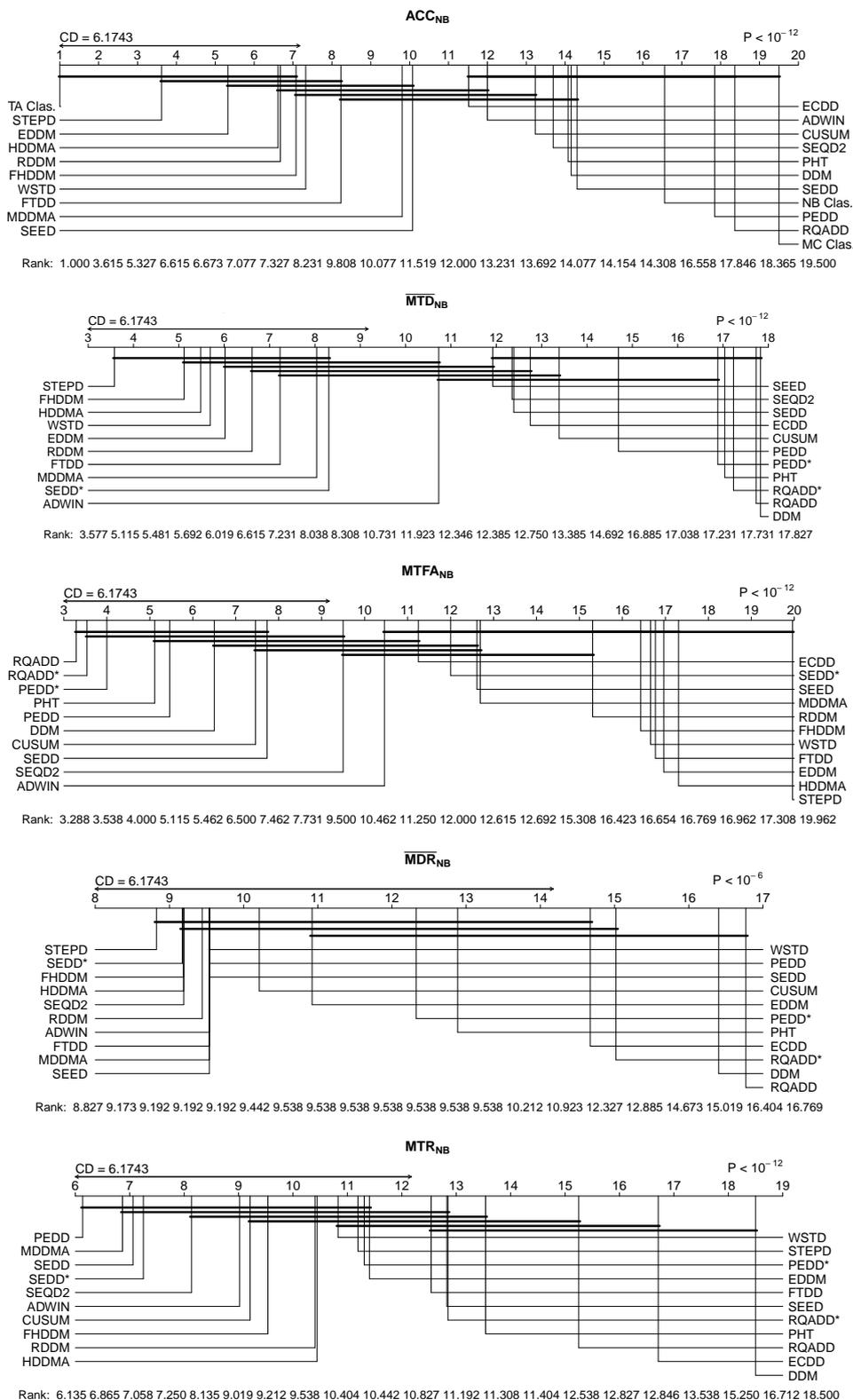
Fonte: Elaborada pelo autor (2021).

Figura 44 – Comparação dos classificadores e detectores com HT usando o teste post-hoc de Nemenyi sobre todas as bases de dados com rotulações simbólicas geradas a partir dos sinais de EEG, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 45 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi sobre todas as bases de dados com rotulações simbólicas geradas a partir dos sinais de EEG, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

5.7 RESUMO DO CAPÍTULO

Neste capítulo, foram apresentados os resultados e as suas respectivas análises referentes aos experimentos usados na avaliação dos métodos Spectral Entropy Drift Detector (SEDD), Permutation Entropy Drift Detector (PEDD) e Recurrence Quantification Analysis Drift Detector (RQADD), e da abordagem de rotulação simbólica Symbolic Labeling Adapter (SLA).

Embora os detectores propostos não tenham apresentados resultados superiores ao estado da arte, as análises sugerem que esses métodos constituem formas alternativas e competitivas, não restritos ao pressuposto de observações i.i.d., para avaliar *mudanças de conceito* no contexto do aprendizado adaptativo supervisionado.

A análise dos resultados também apontam que as mudanças na dinâmica desses *fluxos de dados* não rotulados podem ser identificadas por meio dos erros das previsões das instâncias fornecida pelo SLA, ao considerar a rotulação simbólica dos estados do espaço de fase. Esses resultados mostraram que os detectores acoplados aos classificados HT e NB apresentaram desempenho competitivo em termos da detecção de mudanças na dinâmica do *fluxo de dados* em comparação ao estado da arte (Entropia de Permutação (PE), Análise de Quantificação de Recorrência (RQA), e Entropia Espectral (SE)) usados na avaliação direta dos dados brutos.

Em resumo, os diversos experimentos computacionais usando bases de dados reais e artificiais mostram que os detectores propostos foram competitivos em diversos cenários, e que a abordagem de rotulação simbólica é uma ferramenta viável para generalizar o uso de modelos adaptativos supervisionados a domínios de aplicação com *fluxos de dados* não rotulados.

6 CONSIDERAÇÕES FINAIS

Nesta tese, técnicas da Nonlinear Time Series Analysis (traduzido para o português, Análise de Séries Temporais Não Lineares) (NLTA) foram utilizadas com o objetivo de prover métodos não restritos ao pressuposto de observações i.i.d., e mais apropriados para lidar com *fluxos de dados* sujeitos à dependência temporal. Também foi proposta uma abordagem para contornar o problema da rotulagem com o intuito de expandir a aplicação de modelos adaptativos supervisionados a domínios onde *fluxos de dados* não são rotulados, visando a detecção de *mudança de conceito*.

Com base em abordagens não-supervisionadas propostas por Vallim e Mello (2014), Costa, Rios e Mello (2016), Costa et al. (2017) para a detecção de mudanças em *fluxos de dados* não rotulados, três métodos foram propostos para detectar *mudanças de conceito* no contexto do aprendizado supervisionado: Permutation Entropy Drift Detector (PEDD) com base na Entropia de Permutação (PE), a qual considera padrões simbólicos para identificar mudanças; Recurrence Quantification Analysis Drift Detector (RQADD), fundamenta no *determinismo percentual* RQA_{DET} da Análise de Quantificação de Recorrência (RQA), a qual tem por base a avaliação de recorrências de estados em um *espaço de fase*; e Spectral Entropy Drift Detector (SEDD), o qual tem por base a análise do Espectro de Potência (PS). Como alternativa ao problema da rotulação, a qual limita a aplicação de modelos adaptativos supervisionados para propósito de detecção de *mudanças de conceito* em muitos cenários do mundo real, foi proposto o Symbolic Labeling Adapter (SLA), o qual considera o mapeamento proposto em Pagliosa e Mello (2017) para viabilizar uma rotulação automática considerando padrões simbólicos obtidos dos estados de um espaço de fase.

Para avaliar a eficácia das abordagens propostas, diversos cenários com *mudanças de conceito* abruptas e graduais foram considerados com o objetivo de verificar a resiliência dos detectores em diferentes contextos, como o desbalanceamento das classes, ruídos, número de classes, bem como a influência da dependência temporal no desempenho dos modelos adaptativos. As abordagens foram confrontadas com métodos fundamentados na *análise sequencial*, no *controle estatístico de processos* e na *avaliação de distribuições entre janelas*, com base na acurácia e nas seguintes métricas usadas para avaliar a qualidade das detecções: Missed Detection Rate (traduzido para o português, Taxa de Detecções Perdidas) (MDR), Mean Time for Detection (traduzido para o português, Tempo Médio de Detecção) (MTD), Mean

Time Between False Alarms (traduzido para o português, Tempo Médio Entre Alarmes Falsos) (MTFA) e Mean Time Ratio (traduzido para o português, Razão de Tempo Médio) (MTR). em diversas medidas de qualidade da detecção de *mudança de conceito* (BASSEVILLE; NIKIFOROV, 1993; BIFET, 2017). A metodologia de investigação contemplou dois grupos de experimentos com a finalidade de avaliar as hipóteses de pesquisa relacionadas à eficácia do uso de técnicas derivadas da NLTA em métodos adaptativos explícitos, e à rotulagem simbólica, como forma de viabilizar a aplicação desses métodos em *fluxos de dados* não rotulados com o objetivo de detectar *mudanças de conceito*.

O primeiro grupo de experimentos, com bases rotuladas, sugere que o SEDD apresentou resultados superiores a ambos os métodos RQADD e PEDD em termos de acurácia, menor número de alarmes falsos (MTFA elevado) e qualidade geral da detecção (MTR). Além disso, entre os três, foi o mais competitivo em relação ao estado da arte, principalmente em cenários com *mudanças de conceito* abruptas. O detector RQADD obteve resultados modestos, com bom desempenho em termos de qualidade geral (MTR) de detecção ao considerar mudanças graduais. O PEDD apresentou o pior resultado entre os três, e com o RQADD, não foram competitivos com o estado da arte na maioria dos cenários.

No segundo grupo experimentos com bases não rotuladas, o SLA foi utilizado para fornecer uma pseudo-rotulação simbólica e viabilizar a aplicação dos detectores acoplados aos classificadores HT e NB, o qual apresentou resultados promissores. Nesse caso, os resultados experimentais com a pseudo-rotulação simbólica apontam que o PEDD foi melhor em termos da detecção de mudanças, e que o SEDD apresentou alguns resultados competitivos em termos de acurácia. Entretanto, o RQADD obteve os piores resultados entre os detectores propostos. A Tabela 25 resume as principais conclusões em relação às questões de pesquisa levantadas na Seção 1.2.

Baseado nos resultados obtidos com os métodos propostos, esta tese fornece as seguintes contribuições para lidar com a *mudanças de conceito* no contexto do aprendizado adaptativo supervisionado: i) métodos apropriados para lidar com a dependência temporal do fluxo de erros ao invés de pressupor a independência das observações; e ii) prover a disponibilidade de pseudo-rótulos de modo a viabilizar a utilização de modelos adaptativos supervisionados em cenários com *fluxos de dados* não rotulados.

Em relação à principal diferença entre os métodos propostos nesta tese e os demais da literatura, vale destacar a perspectiva da identificação da *mudança de conceito* por meio da avaliação de alterações entre regimes dinâmicos distintos, em vez da avaliação da não

estacionaridade da distribuição dos dados, a qual é utilizada pela maioria dos detectores. Assim, os métodos propostos possibilitam investigar mudanças relacionadas ao mecanismo por trás do processo gerador dos dados.

Como ponto positivo, a aplicação desses métodos não presume que o fluxo de erro das predições seja um conjunto de observações estatisticamente independentes, sendo mais permissíveis em relação às restrições teóricas do estado da arte. Além disso, o SLA viabiliza a expansão de aplicações de modelos adaptativos supervisionados a cenários do mundo real, onde a maioria dos *fluxos de dados* não são rotulados. Essa limitação em relação à necessidade de dados rotulados levanta diversas críticas em relação à utilização de modelos adaptativos supervisionados (VALLIM; MELLO, 2014; DYER; CAPO; POLIKAR, 2014; SETHI; KANTARDZIC, 2017). Por outro lado, como principal fator negativo, os detectores propostos foram computacionalmente menos eficientes quando comparados aos detectores do estado da arte.

Para trabalhos futuros, existem alguns pontos interessantes ainda a serem explorados:

- Em relação a avaliação dos detectores, um ponto a ser melhorado é a utilização de métodos alternativos para a redução do consumo de recursos computacionais. Consequentemente, métodos que dependem da avaliação topológica dos estados, como a RQA, deve ser aplicados de forma parcimoniosa. Entretanto, outros métodos derivados da PE (CUESTA-FRAU, 2019; DING; LUO, 2019) podem ser explorados, e a obtenção das estatísticas dos símbolos otimizada com *hashing perfeito* (CORMEN et al., 2009). Além disso, outros métodos com base no PS podem ser investigados (GHADERPOUR; PAGIATAKIS; HASSAN, 2021).
- Com respeito à rotulação simbólica, uma questão interessante está relacionada ao sentido semântico dos padrões ordinais no domínio da aplicação. Outra questão está na viabilidade de estender essa abordagem com técnicas capazes de reconstruir espaços de fase a partir de *fluxo de dados* multivariados (DEYLE; SUGIHARA, 2011), bem como a investigação de outras formas de rotulação (LIND; MARCUS, 2021), com o objetivo de melhorar a identificação de *mudanças de conceito*.
- Ainda no tocante ao Symbolic Labeling Adapter (SLA), pretende-se investigar a utilização dessa abordagem para generalizar a aplicação de modelos adaptativos supervisionados a outros cenários do mundo real relacionados à análise de séries temporais.

Tabela 25 – Questões de pesquisa e as respectivas respostas obtidas por meio da análise dos resultados experimentais.

<p>QP 1. Os detectores fundamentados na NLTA apresentam desempenho superior ao estado da arte em termos de acurácia e detecção?</p> <p><i>O SEDD apresentou os melhores resultados entre os três detectores propostos, seguido do RQADD. O PEDD não foi competitivo na maioria das avaliações com respeito a acurácia, mas obteve bons resultados em termos da métrica MTR. Logo, não foram superiores ao desempenho dos melhores detectores do estado da arte em termos de acurácia e detecção. Entretanto esses detectores não apresentaram diferenças estatísticas significativas em relação ao SEDD na maioria dos experimentos com as bases artificiais.</i></p>
<p>QP 2. Os classificadores acoplados a detectores de <i>mudança de conceito</i> aplicados a fluxos de dados com dependência temporal apresentam diferença significativas em termos de acurácia em comparação a um classificador persistente (um classificador ingênuo que realiza a predição da instância atual com o rótulo da instância anterior (ŽLIOBAITĚ et al., 2015))?</p> <p><i>O Classificador Temporalmente Aumentado obteve a melhor acurácia em relação às bases reais rotuladas, as quais apresentam dependência temporal. Entretanto, esse classificador não apresentou diferença estatística significativa em relação a maioria dos detectores do estado da arte, e entre os detectores propostos, o SEDD obteve o melhor desempenho em termos de acurácia e ARL (menor número de alarmes). Os detectores PEDD e RQADD foram competitivos em termos do ARL. Vale salientar que, com a dependência temporal, o aumento da quantidade de alarmes favorece a acurácia (ŽLIOBAITĚ et al., 2015), ou seja, os detectores que mais disparam alarmes tendem a apresentar melhor desempenho em relação à acurácia.</i></p>
<p>QP 3. A rotulação de um fluxo de dados unidimensional por meio de padrões ordinais, definidos sobre os estados de um espaço de fase obtido a partir da aplicação do teorema de imersão de Takens (TAKENS, 1981), viabiliza a aplicação de modelos adaptativos supervisionados para fins de detecção da mudança de conceito de modo eficaz?</p> <p><i>Os resultados sugerem que ao considerar a pseudo-rotulação simbólica dos estados do espaço de fase, mudanças na dinâmica de fluxos de dados podem ser identificadas por meio dos erros das predições das instâncias fornecida pelo SLA. Logo, o SLA pode ser usado para expandir a aplicação de modelos adaptativos supervisionados a contextos com fluxos de dados não rotulados.</i></p>
<p>QP 4. Ao considerar a rotulação simbólica usando padrões ordinais dos estados, detectores acoplados a classificadores apresentam melhor desempenho em termos da detecção de mudanças na dinâmica do <i>fluxo de dados</i> em comparação ao estado da arte usados na avaliação direta dos dados brutos?</p> <p><i>As ferramentas da NLTA (Entropia de Permutação (PE), RQA, e Entropia Espectral (SE)), otimizadas nos experimentos para o propósito de detecção de mudanças de conceito, não apresentaram diferenças estatísticas significativas apenas a um número reduzido de detectores do estado da arte (principalmente o ADWIN e o SEED). Entre os detectores propostos, os testes estatísticos indicam que o PEDD foi o mais competitivo na maioria das avaliações, com ranqueamento superior a todos os detectores nas bases de EEG.</i></p>

Fonte: Elaborada pelo autor (2021).

6.1 PUBLICAÇÃO RELACIONADA À TESE

- **Using Spectral Entropy and Bernoulli Map to Handle Concept Drift** (Rohgi Toshio Meneses Chikushi, Roberto Souto Maior de Barros, Marilú Gomes Netto Monte Da Silva, Bruno Iran Ferreira Maciel) - Publicado em Abril de 2021 no periódico **Expert Systems with Applications** (CHIKUSHI et al., 2021) - Qualis A1.

REFERÊNCIAS

- ABARBANEL, H. D. I.; BROWN, R.; SIDOROWICH, J. J.; TSIMRING, L. S. The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.*, American Physical Society, v. 65, p. 1331–1392, Oct 1993.
- AGGARWAL, C. *Data Streams: Models and Algorithms*. [S.l.: s.n.], 2007. v. 31.
- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. N. Database mining: a performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, v. 5, n. 6, p. 914–925, Dec 1993. ISSN 1041-4347.
- ALBANO, A.; PASSAMANTE, A.; FARRELL, M. E. Using higher-order correlations to define an embedding window. *Physica D: Nonlinear Phenomena*, v. 54, n. 1, p. 85 – 97, 1991.
- ALLIGOOD, K. T.; SAUER, T. D.; YORKE, J. A. *Chaos: An Introduction to Dynamical Systems*. [S.l.]: Springer, 1997.
- AMIGÓ, J. *Permutation complexity in dynamical systems: ordinal patterns, permutation entropy and all that*. [S.l.]: Springer Science & Business Media, 2010.
- ANDERSON, R.; KOH, Y. S.; DOBBIE, G.; BIFET, A. Recurring concept meta-learning for evolving data streams. *Expert Systems with Applications*, v. 138, p. 112832, 2019. ISSN 0957-4174.
- BAENA-GARCÍA, M.; CAMPO-ÁVILA, J. D.; FIDALGO, R.; BIFET, A.; GAVALDÀ, R.; MORALES-BUENO, R. Early drift detection method. In: *Fourth International Workshop on Knowledge Discovery from Data Streams*. [S.l.: s.n.], 2006. p. 77–86.
- BANDT, C.; POMPE, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, v. 88, p. 174102, 2002.
- BARROS, R. S. M.; CABRAL, D. R. L.; JR, P. M. G.; SANTOS, S. G. T. C. RDDM: Reactive drift detection method. *Expert Systems with Applications*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 90, n. C, p. 344–355, dez. 2017. ISSN 0957-4174.
- BARROS, R. S. M.; HIDALGO, J. I. G.; CABRAL, D. R. L. Wilcoxon rank sum test drift detector. *Neurocomputing*, v. 275, p. 1954 – 1963, 2018. ISSN 0925-2312.
- BARROS, R. S. M.; SANTOS, S. G. T. C. A large-scale comparison of concept drift detectors. *Information Sciences*, v. 451-452, p. 348 – 370, 2018. ISSN 0020-0255.
- BARROS, R. S. M.; SANTOS, S. G. T. C. An overview and comprehensive comparison of ensembles for concept drift. *Information Fusion*, v. 52, p. 213 – 244, 2019. ISSN 1566-2535.
- BASSEVILLE, M.; NIKIFOROV, I. V. *Detection of Abrupt Changes: Theory and Application*. USA: Prentice-Hall, Inc., 1993.
- BERNSTEIN, S. *The theory of probabilities*. Moscow: Gastehizdat Publishing House, 1946.
- BIFET, A. Adaptive learning and mining for data streams and frequent patterns. *ACM SIGKDD Explorations Newsletter*, Association for Computing Machinery, New York, NY, USA, v. 11, n. 1, p. 55–56, 2009.

BIFET, A. Classifier concept drift detection and the illusion of progress. In: *Artificial Intelligence and Soft Computing*. Cham: Springer International Publishing, 2017. p. 715–725.

BIFET, A.; GAVALDÀ, R. Learning from time-changing data with adaptive windowing. In: *Proceedings of The 7th SIAM International Conference on Data Mining (SDM'07)*. Minneapolis, MN, USA: [s.n.], 2007. p. 443–448.

BIFET, A.; GAVALDÀ, R.; HOLMES, G.; PFAHRINGER, B. *Machine learning for data streams: with practical examples in MOA*. [S.l.]: MIT press, 2018.

BIFET, A.; HOLMES, G.; KIRKBY, R.; PFAHRINGER, B. MOA: massive online analysis. *Journal of Machine Learning Research*, v. 11, p. 1601–1604, 2010.

BIFET, A.; HOLMES, G.; PFAHRINGER, B.; KIRKBY, R.; GAVALDÀ, R. New ensemble methods for evolving data streams. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2009. (KDD '09), p. 139–148.

BIFET, A.; READ, J.; ŽLIOBAITĚ, I.; PFAHRINGER, B.; HOLMES, G. Pitfalls in benchmarking data stream classification and how to avoid them. In: *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part I*. [S.l.]: Springer, 2013. (Lecture Notes in Computer Science, v. 8188), p. 465–479.

BISHOP, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York: Springer-Verlag, 2006.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.

BRADLEY, E.; KANTZ, H. Nonlinear time-series analysis revisited. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, AIP Publishing LLC, v. 25, n. 9, p. 097610, 2015.

BROOMHEAD, D.; KING, G. P. Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, v. 20, n. 2, p. 217 – 236, 1986. ISSN 0167-2789.

CABRAL, D. R. L.; BARROS, R. S. M. Concept drift detection based on fisher's exact test. *Information Sciences*, v. 442-443, p. 220 – 234, 2018. ISSN 0020-0255.

CANDILLIER, L.; LEMAIRE, V. Design and analysis of the nomao challenge active learning in the real-world. In: . [S.l.: s.n.], 2013.

CAO, L. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, v. 110, n. 1, p. 43 – 50, 1997. ISSN 0167-2789.

CAO, Y.; TUNG, W.; GAO, J.; PROTOPOPESCU, V. A.; HIVELY, L. M. Detecting dynamical changes in time series using the permutation entropy. *Physical review E, APS*, v. 70, n. 4, p. 046217:1–8, 2004.

CASDAGLI, M. Recurrence plots revisited. *Physica D: Nonlinear Phenomena*, v. 108, n. 1, p. 12 – 44, 1997. ISSN 0167-2789.

CAVALCANTE, R.; MINKU, L.; OLIVEIRA, A. Fedd: Feature extraction for explicit concept drift detection in time series. In: . [S.l.: s.n.], 2016. p. 740–747.

- CHIKUSHI, R. T. M.; BARROS, R. S. M. de; SILVA, M. G. N. M. da; MACIEL, B. I. F. Using spectral entropy and bernoulli map to handle concept drift. *Expert Systems with Applications*, v. 167, p. 114114, 2021.
- COOLEY, J.; TUKEY, J. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, v. 19, p. 297–301, 1965.
- CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. *Introduction to Algorithms*. 3rd. ed. [S.l.]: The MIT Press, 2009.
- COSTA, F. G.; DUARTE, F. S. L. G.; VALLIM, R. M. M.; MELLO, R. F. Multidimensional surrogate stability to detect data stream concept drift. *Expert Systems with Applications*, v. 87, p. 15 – 29, 2017. ISSN 0957-4174.
- COSTA, F. G.; RIOS, R. A.; MELLO, R. F. Using dynamical systems tools to detect concept drift in data streams. *Expert Systems with Applications*, v. 60, p. 39 – 50, 2016. ISSN 0957-4174.
- COSTA, F. G. d. *Employing nonlinear time series analysis tools with stable clustering algorithms for detecting concept drift on data streams*. Tese (Doutorado) — Instituto de Ciências Matemáticas e de Computação, 2017.
- CUESTA–FRAU, D. Permutation entropy: Influence of amplitude information on time series classification performance. *Mathematical Biosciences and Engineering*, v. 16, n. mbe-16-06-342, p. 6842, 2019. ISSN 1551-0018.
- DEMSAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, JMLR.org, v. 7, p. 1–30, January 2006. ISSN 1532-4435. Disponível em: <<http://dl.acm.org/citation.cfm?id=1248547.1248548>>.
- DEVANEY, R. *A First Course In Chaotic Dynamical Systems: Theory And Experiment*. [S.l.]: Avalon Publishing, 1992. (Studies in Nonlinearity).
- DEYLE, E.; SUGIHARA, G. Generalized theorems for nonlinear state space reconstruction. *PloS one*, v. 6, p. e18295, 03 2011.
- DING, F.; LUO, C. The entropy-based time domain feature extraction for online concept drift detection. *Entropy*, v. 21, n. 12, 2019.
- DOMINGOS, P.; HULTEN, G. A general method for scaling up machine learning algorithms and its application to clustering. In: *Proceedings of the Eighteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. (ICML '01), p. 106–113.
- DOMINGOS, P.; PAZZANI, M. On the optimality of the simple bayesian classifier under zero-one loss. *Machine learning*, Springer, v. 29, n. 2, p. 103–130, 1997.
- DONNER, R. V.; SMALL, M.; DONGES, J. F.; MARWAN, N.; ZOU, Y.; XIANG, R.; KURTHS, J. Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcation and Chaos*, World Scientific, v. 21, n. 04, p. 1019–1046, 2011.

- DONNER, R. V.; ZOU, Y.; DONGES, J. F.; MARWAN, N.; KURTHS, J. Recurrence networks—a novel paradigm for nonlinear time series analysis. *New Journal of Physics*, IOP Publishing, v. 12, n. 3, p. 033025, mar 2010.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.
- DUONG, Q.-H.; RAMAMPIARO, H.; NØRVÅG, K. Applying temporal dependence to detect changes in streaming data. *Applied Intelligence*, p. 19, 12 2018.
- DYER, K. B.; CAPO, R.; POLIKAR, R. Compose: A semisupervised learning framework for initially labeled nonstationary streaming data. *IEEE Transactions on Neural Networks and Learning Systems*, v. 25, n. 1, p. 12–26, 2014.
- ECKMANN, J.-P.; KAMPHORST, S.; RUELLE, D. Recurrence plots of dynamical systems. *Europhysics Letters (epl)*, v. 4, p. 973–977, 11 1987.
- ELWELL, R.; POLIKAR, R. Incremental learning of concept drift in nonstationary environments. *Neural Networks, IEEE Transactions on*, v. 22, p. 1517 – 1531, 11 2011.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. F. d. *Inteligência artificial: uma abordagem de aprendizado de máquina*. [S.l.]: LTC, 2011.
- FERREIRA, L. B.; BARDDAL, J. P.; GOMES, H. M.; ENEMBRECK, F. An experimental perspective on sampling methods for imbalanced learning from financial databases. In: . [S.l.: s.n.], 2018.
- FISHER, R. A. On the interpretation of χ^2 from contingency tables, and the calculation of p . *Journal of the Royal Statistical Society*, [Wiley, Royal Statistical Society], v. 85, n. 1, p. 87–94, 1922. ISSN 09528385.
- FRASER, A. M.; SWINNEY, H. L. Independent coordinates for strange attractors from mutual information. *Physical review A, APS*, v. 33, n. 2, p. 1134, 1986.
- FREY, P. W.; SLATE, D. J. Letter recognition using holland-style adaptive classifiers. *Mach. Learn.*, Kluwer Academic Publishers, USA, v. 6, n. 2, p. 161–182, mar. 1991.
- FRÍAS-BLANCO, I.; CAMPO-ÁVILA, J. del; RAMOS-JIMÉNEZ, G.; MORALES-BUENO, R.; ORTIZ-DÍAZ, A.; CABALLERO-MOTA, Y. Online and non-parametric drift detection methods based on hoeffding's bounds. *IEEE Transactions on Knowledge and Data Engineering*, v. 27, n. 3, p. 810–823, 2015.
- GAMA, J. *Knowledge discovery from data streams*. [S.l.]: CRC Press, 2010.
- GAMA, J.; MEDAS, P.; CASTILLO, G.; RODRIGUES, P. Learning with drift detection. In: *Advances in Artificial Intelligence: SBIA 2004*. [S.l.]: Springer, 2004, (Lecture Notes in Computer Science, v. 3171). p. 66–112.
- GAMA, J.; ŽLIOBAITĚ, I.; BIFET, A.; PECHENIZKIY, M.; BOUCHACHIA, A. A survey on concept drift adaptation. *ACM Computing Surveys*, ACM, New York, NY, USA, v. 46, n. 4, p. 44:1–44:37, mar. 2014. ISSN 0360-0300.
- GHADERPOUR, E.; PAGIATAKIS, S. D.; HASSAN, Q. K. A survey on change detection and time series analysis with applications. *Applied Sciences*, v. 11, 2021.

- GOMES, H. M.; BARDDAL, J. P.; ENEMBRECK, F.; BIFET, A. A survey on ensemble learning for data stream classification. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 50, n. 2, mar. 2017.
- GRASSBERGER, P.; PROCACCIA, I. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, v. 9, n. 1, p. 189 – 208, 1983. ISSN 0167-2789.
- HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.I.]: Elsevier, 2011.
- HARRIES, M. B.; SAMMUT, C.; HORN, K. Extracting hidden context. *Mach. Learn.*, Kluwer Academic Publishers, USA, v. 32, n. 2, p. 101–126, ago. 1998. ISSN 0885-6125.
- HAUSSLER, D. *Probably approximately correct learning*. [S.I.]: University of California, Santa Cruz-CA, USA, Computer Research Laboratory, 1990.
- HIDALGO, J. I. G.; MACIEL, B. I. F.; BARROS, R. S. M. Experimenting with prequential variations for data stream learning evaluation. *Computational Intelligence*, Wiley, v. 35, n. 4, p. 670–692, 2019.
- HOEFFDING, W. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, Taylor & Francis Group, v. 58, n. 301, p. 13–30, 1963.
- HUANG, D. T. J.; KOH, Y. S.; DOBBIE, G.; PEARS, R. Detecting volatility shift in data streams. In: IEEE. *Data Mining (ICDM), 2014 IEEE International Conference on*. [S.I.], 2014. p. 863–868.
- HULTEN, G.; SPENCER, L.; DOMINGOS, P. Mining time-changing data streams. In: *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA: [s.n.], 2001. (KDD '01), p. 97–106.
- INOUYE, T.; SHINOSAKI, K.; SAKAMOTO, H.; TOI, S.; UKAI, S.; IYAMA, A.; KATSUDA, Y.; HIRANO, M. Quantification of EEG irregularity by use of the entropy of the power spectrum. *Electroencephalography and clinical Neurophysiology*, Elsevier, v. 79, n. 3, p. 204–210, 1991.
- JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. In: *Proceedings of The 11th Conference on Uncertainty in Artificial Intelligence (UAI'95)*. San Francisco, CA, USA: Morgan Kaufmann, 1995. p. 338–345.
- KANTZ, H. A robust method to estimate the maximal lyapunov exponent of a time series. *Physics Letters A*, v. 185, n. 1, p. 77 – 87, 1994. ISSN 0375-9601.
- KANTZ, H.; SCHREIBER, T. *Nonlinear time series analysis*. [S.I.]: Cambridge university press, 2004. v. 7.
- KATAKIS, I.; TSOUMAKAS, G.; VLAHAVAS, I. Tracking recurring contexts using ensemble classifiers: An application to email filtering. *Knowledge and Information Systems*, Springer-Verlag, Berlin, Heidelberg, v. 22, n. 3, p. 371–391, mar. 2010. ISSN 0219-1377.
- KEMBER, G.; FOWLER, A. A correlation function for choosing time delays in phase portrait reconstructions. *Physics Letters A*, v. 179, n. 2, p. 72 – 80, 1993. ISSN 0375-9601.
- KENNEL, M. B.; BROWN, R.; ABARBANEL, H. D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, American Physical Society, v. 45, p. 3403–3411, Mar 1992.

- KIFER, D.; BEN-DAVID, S.; GEHRKE, J. Detecting change in data streams. In: *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30*. [S.l.]: VLDB Endowment, 2004. (VLDB '04), p. 180–191.
- KLEM, G.; LÜDERS, H.; JASPER, H.; ELGER, C. The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalography and clinical neurophysiology. Supplement*, v. 52, p. 3–6, 1999.
- KOHAVI, R. Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid. *KDD*, 09 1997.
- KRAWCZYK, B.; MINKU, L. L.; GAMA, J.; STEFANOWSKI, J.; WOŹNIAK, M. Ensemble learning for data stream analysis: A survey. *Information Fusion*, v. 37, p. 132 – 156, 2017. ISSN 1566-2535.
- LI, T. yien; YORKE, J. A. Period three implies chaos. *AMER. MATH. MONTHLY*, v. 82, n. 10, p. 985–992, 1975.
- LIND, D.; MARCUS, B. *An Introduction to Symbolic Dynamics and Coding*. 2. ed. [S.l.]: Cambridge University Press, 2021. (Cambridge Mathematical Library).
- LORENZ, E. N. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, v. 20, n. 2, p. 130–141, 03 1963. ISSN 0022-4928.
- LORENZ, E. N. Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, v. 26, n. 4, p. 636–646, 1969.
- LOSING, V.; HAMMER, B.; WERSING, H. Interactive online learning for obstacle classification on a mobile robot. In: *2015 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2015. p. 1–8.
- LUXBURG, U. von; SCHÖLKOPF, B. Statistical learning theory: Models, concepts, and results. In: GABBAY, D. M.; HARTMANN, S.; WOODS, J. (Ed.). *Inductive Logic*. [S.l.]: North-Holland, 2011, (Handbook of the History of Logic, v. 10). p. 651 – 706.
- MACGREGOR, J.; KOURTI, T. Statistical process control of multivariate processes. *Control Engineering Practice*, v. 3, n. 3, p. 403 – 414, 1995. ISSN 0967-0661.
- MACIEL, B. I. F.; SANTOS, S. G. T. C.; BARROS, R. S. M. MOAManager: a tool to support data stream experiments. *Software: Practice and Experience*, Wiley, v. 50, n. 4, p. 325–334, 2020.
- MARWAN, N.; ROMANO, M. C.; THIEL, M.; KURTHS, J. Recurrence plots for the analysis of complex systems. *Physics reports*, Elsevier, v. 438, n. 5, p. 237–329, 2007.
- MAY, R. Simple mathematical models with very complicated dynamics. *Nature*, v. 26, p. 457, 07 1976.
- MCDIARMID, C. On the method of bounded differences. *Surveys in combinatorics*, Norwich, v. 141, n. 1, p. 148–188, 1989.
- MITCHELL, T. M. *Machine Learning*. USA: McGraw-Hill, Inc., 1997.
- MORETTI, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. 2. ed. [S.l.]: Editora Edgard Blücher Ltda, 2006.

- MORO, S.; CORTEZ, P.; RITA, P. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, v. 62, p. 22–31, 2014. ISSN 0167-9236.
- MOUSS, H.; MOUSS, M.; MOUSS, K.; LINDA, S. Test of page-hinckley, an approach for fault detection in an agro-alimentary production system. In: . [S.l.: s.n.], 2004. p. 815 – 818 Vol.2.
- NEMENYI, P. Distribution-free multiple comparisons. In: *International Biometric Society*. [S.l.]: Biometrics, 1962. v. 18, n. 2, p. 263.
- NISHIDA, K.; YAMAUCHI, K. Detecting concept drift using statistical testing. In: *Proceedings of 10th International Conference on Discovery Science (DS'07)*. [S.l.]: Springer, 2007. (LNCS, v. 4755), p. 264–269.
- OTT, E. *Chaos in Dynamical Systems*. 2. ed. [S.l.]: Cambridge University Press, 2002.
- OZA, N.; RUSSELL, S. Experimental comparisons of online and batch versions of bagging and boosting. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 10 2001.
- ÖZTÜRK, I.; KILIÇ, R. Digitally generating true orbits of binary shift chaotic maps and their conjugates. *Communications in Nonlinear Science and Numerical Simulation*, v. 62, p. 395 – 408, 2018. ISSN 1007-5704.
- PACKARD, N. H.; CRUTCHFIELD, J. P.; FARMER, J. D.; SHAW, R. S. Geometry from a time series. *Phys. Rev. Lett.*, American Physical Society, v. 45, p. 712–716, Sep 1980.
- PAGE, E. S. Continuous inspection schemes. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 41, n. 1-2, p. 100–115, 1954.
- PAGLIOSA, L. d. C.; MELLO, R. F. Applying a kernel function on time-dependent data to provide supervised-learning guarantees. *Expert Systems with Applications*, v. 71, p. 216 – 229, 2017. ISSN 0957-4174.
- PALIT, A. K.; POPOVIC, D. *Computational intelligence in time series forecasting: theory and engineering applications*. [S.l.]: Springer Science & Business Media, 2006.
- PEARS, R.; SAKTHITHASAN, S.; KOH, Y. S. Detecting concept change in dynamic data streams. *Machine Learning*, v. 97, n. 3, p. 259–293, 2014. ISSN 1573-0565.
- PESARANGHADER, A.; VIKTOR, H. L. Fast hoeffding drift detection method for evolving data streams. In: *Machine Learning and Knowledge Discovery in Databases*. Cham: Springer, 2016. p. 96–111.
- PESARANGHADER, A.; VIKTOR, H. L.; PAQUET, E. Mcdiarmid drift detection methods for evolving data streams. In: IEEE. *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2018. p. 1–9.
- POWELL, G.; PERCIVAL, I. A spectral entropy method for distinguishing regular and irregular motion of hamiltonian systems. *Journal of Physics A: Mathematical and General*, IOP Publishing, v. 12, n. 11, p. 2053, 1979.
- QUINLAN, J. R. Induction of decision trees. *Mach. Learn.*, Kluwer Academic Publishers, USA, v. 1, n. 1, p. 81–106, mar. 1986. ISSN 0885-6125.

- RIOS, R. A.; MELLO, R. F. de. Applying empirical mode decomposition and mutual information to separate stochastic and deterministic influences embedded in signals. *Signal Processing*, v. 118, p. 159 – 176, 2016. ISSN 0165-1684.
- ROBERTS, S. W. Control chart tests based on geometric moving averages. *Technometrics*, v. 1, n. 3, p. 239–250, 1959.
- ROSENSTEIN, M. T.; COLLINS, J. J.; De Luca, C. J. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, v. 65, n. 1, p. 117 – 134, 1993. ISSN 0167-2789.
- ROSS, G. J.; ADAMS, N. M.; TASOULIS, D. K.; HAND, D. J. Exponentially weighted moving average charts for detecting concept drift. *Pattern Recognition Letters*, v. 33, n. 2, p. 191 – 198, 2012. ISSN 0167-8655.
- SAUER, T.; YORKE, J. A.; CASDAGLI, M. Embedology. *J. Stat. Phys.*, v. 65, p. 579, 1991.
- SAVI, M. A. *Dinâmica não-linear e caos*. [S.l.]: Editora E-papers, 2006.
- SCHLIMMER, J. C.; GRANGER, R. H. Incremental learning from noisy data. *Machine Learning*, Kluwer Academic Publishers, USA, v. 1, n. 3, p. 317–354, 1986. ISSN 0885-6125.
- SCHMIDT, J. P.; SIEGEL, A.; SRINIVASAN, A. Chernoff–hoeffding bounds for applications with limited independence. *SIAM Journal on Discrete Mathematics*, v. 8, n. 2, p. 223–250, 1995.
- SCHREIBER, T. Interdisciplinary application of nonlinear time series methods. *Physics Reports*, v. 308, n. 1, p. 1 – 64, 1999. ISSN 0370-1573.
- SEBASTIÃO, R.; GAMA, J.; RODRIGUES, P.; BERNARDES, J. Monitoring incremental histogram distribution for change detection in data streams. In: *Knowledge Discovery from Sensor Data*. [S.l.: s.n.], 2010. v. 5840, p. 25–42.
- SEBASTIÃO, R.; GAMA, J.; MENDONÇA, T. Fading histograms in detecting distribution and concept changes. *International Journal of Data Science and Analytics*, v. 3, 05 2017.
- SEMMLOW, J. L.; GRIFFEL, B. *Biosignal and medical image processing*. [S.l.]: CRC press, 2014.
- SETHI, T.; KANTARDZIC, M. Don't pay for validation: Detecting drifts from unlabeled data using margin density. *Procedia Computer Science*, v. 53, p. 103–112, 12 2015.
- SETHI, T. S.; KANTARDZIC, M. On the reliable detection of concept drift from streaming unlabeled data. *Expert Systems with Applications*, v. 82, p. 77 – 99, 2017. ISSN 0957-4174.
- SHANNON, C. E. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, ACM, v. 5, n. 1, p. 3–55, 2001.
- SHUMWAY, R. H.; STOFFER, D. S. *Time series analysis and its applications: with R examples*. [S.l.]: Springer, 2017.
- SLEIGH, J. W.; STEYN-ROSS, D. A.; STEYN-ROSS, M. L.; GRANT, C.; LUDBROOK, G. Cortical entropy changes with general anaesthesia: theory and experiment. *Physiological measurement*, IOP Publishing, v. 25, n. 4, p. 921, 2004.

- SMALL, M. *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*. [S.l.]: WORLD SCIENTIFIC, 2005.
- SPARKS, R. S. Cusum charts for signalling varying location shifts. *Journal of Quality Technology*, Taylor & Francis, v. 32, n. 2, p. 157–171, 2000.
- SPARROW, C. *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*. [S.l.]: Springer New York, 1982. v. 41. ISSN 0066-5452.
- SRIPIRAKAS, S.; PEARS, R.; KOH, Y. S. One pass concept change detection for data streams. In: . [S.l.: s.n.], 2013.
- STAM, C. Nonlinear dynamical analysis of eeg and meg: Review of an emerging field. *Clinical Neurophysiology*, v. 116, n. 10, p. 2266 – 2301, 2005. ISSN 1388-2457.
- STEVENSON, N.; TAPANI, K.; LAURONEN, L.; VANHATALO, S. *A dataset of neonatal EEG recordings with seizures annotations*. [S.l.]: Zenodo, 2018.
- TAKENS, F. Detecting strange attractors in turbulence. In: *Dynamical systems and turbulence, Warwick 1980*. [S.l.]: Springer, 1981. p. 366–381.
- THEILER, J.; EUBANK, S.; LONGTIN, A.; GALDRIKIAN, B.; FARMER, J. D. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, v. 58, n. 1, p. 77 – 94, 1992. ISSN 0167-2789.
- TRULLA, L.; GIULIANI, A.; ZBILUT, J.; JR, C. W. Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A*, Elsevier, v. 223, n. 4, p. 255–260, 1996.
- TSYMBAL, A. The problem of concept drift: definitions and related work. *Computer Science Department, Trinity College Dublin*, Citeseer, v. 106, n. 2, p. 58, 2004.
- TSYMBAL, A.; PECHENIZKIY, M.; CUNNINGHAM, P.; PUURONEN, S. Dynamic integration of classifiers for handling concept drift. *Information Fusion*, Elsevier Science Publishers B. V., v. 9, n. 1, p. 56–68, 2008. ISSN 1566-2535.
- VALIANT, L. G. A theory of the learnable. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 27, n. 11, p. 1134–1142, nov. 1984. ISSN 0001-0782.
- VALLIM, R. M. M.; MELLO, R. F. Proposal of a new stability concept to detect changes in unsupervised data streams. *Expert Systems with Applications*, Elsevier, v. 41, n. 16, p. 7350–7360, 2014.
- VAPNIK, V. *The nature of statistical learning theory*. [S.l.]: Springer science & business media, 2013.
- VAPNIK, V. N. An overview of statistical learning theory. *Trans. Neur. Netw.*, IEEE Press, v. 10, n. 5, p. 988–999, set. 1999. ISSN 1045-9227.
- VERDIER, G.; HILGERT, N.; VILA, J.-P. Adaptive threshold computation for cusum-type procedures in change detection and isolation problems. *Comput. Stat. Data Anal.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 52, n. 9, p. 4161–4174, maio 2008. ISSN 0167-9473.
- VITTER, J. S. Random sampling with a reservoir. *ACM Transactions on Mathematical Software (TOMS)*, ACM, v. 11, n. 1, p. 37–57, 1985.

- WALD, A. *Sequential analysis*. [S.l.]: Courier Corporation, 1973.
- WANG, H.; ABRAHAM, Z. Concept drift detection for streaming data. In: IEEE. *Neural Networks (IJCNN), 2015 International Joint Conference on*. [S.l.], 2015. p. 1–9.
- WIDMER, G. Combining robustness and flexibility in learning drifting concepts. In: *In Proceedings of the 11th European Conference on Artificial Intelligence, ECAI94*. [S.l.]: Wiley & Sons, 1994. p. 468–472.
- WILCOXON, F. Individual comparisons by ranking methods. *Biometrics Bulletin*, v. 1, p. 80–83, 1945.
- WOLF, A.; SWIFT, J. B.; SWINNEY, H. L.; VASTANO, J. A. Determining lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, v. 16, n. 3, p. 285 – 317, 1985. ISSN 0167-2789.
- WOODALL, W. H.; MONTGOMERY, D. C. Research issues and ideas in statistical process control. *Journal of Quality Technology*, American Society for Quality, v. 31, n. 4, p. 376, 1999.
- YEH, A.; MCGRATH, R.; SEMBOWER, M.; SHEN, Q. Ewma control charts for monitoring high-yield processes based on non-transformed observations. *International Journal of Production Research - INT J PROD RES*, v. 46, p. 5679–5699, 10 2008.
- ZIEBA, M.; HÄRDLE, W. K. Beta-boosted ensemble for big credit scoring data. In: *Handbook of Big Data Analytics*. Cham: Springer International Publishing, 2018. p. 523–538.
- ŽLIJBAITĚ, I. Change with delayed labeling: When is it detectable? In: . [S.l.: s.n.], 2011. p. 843 – 850.
- ŽLIJBAITĚ, I.; BIFET, A.; READ, J.; PFAHRINGER, B.; HOLMES, G. Evaluation methods and decision theory for classification of streaming data with temporal dependence. *Machine Learning*, Springer, v. 98, n. 3, p. 455–482, 2015.

APÊNDICE A – RESULTADOS EXPERIMENTAIS DAS BASES ARTIFICIAIS COM O NB

Tabela 26 – Resultados obtidos usando NB, considerando a base artificial Agrawal balanceada, com um intervalo de confiança de 95%.

<i>Agrawal</i> _(B)	<i>Classificador/Detector</i> _(NB)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Majority Class	39,90 ± 0,00					
	Naive Bayes	73,88 ± 0,05					
	Temporally Augmented Classifier	99,97 ± 0,00					
	ADWIN	89,28 ± 0,02	65,87 ± 2,56	42381,97 ± 3288,37	0,00 ± 0,00	651,17 ± 56,28	
	CUSUM	89,24 ± 0,03	553,73 ± 870,38	50255,71 ± 799,78	0,01 ± 0,02	381,76 ± 30,33	
	DDM	88,18 ± 0,73	2246,31 ± 1897,34	32082,77 ± 3322,51	0,02 ± 0,04	51,67 ± 12,46	
	ECDD	86,45 ± 0,04	1249,47 ± 1219,23	811,85 ± 12,62	0,00 ± 0,00	24,81 ± 13,46	
	EDDM	86,84 ± 0,08	3630,70 ± 123,15	9156,11 ± 1977,79	0,00 ± 0,00	2,51 ± 0,54	
	FHDDM	89,01 ± 0,03	523,22 ± 870,37	42218,93 ± 204,11	0,01 ± 0,02	550,28 ± 127,30	
	FTDD	89,30 ± 0,03	53,18 ± 8,01	48083,59 ± 1992,62	0,00 ± 0,00	1041,99 ± 147,36	
	HDDMA	89,22 ± 0,03	133,84 ± 20,50	46952,60 ± 2027,36	0,00 ± 0,00	408,72 ± 61,47	
	MDDMA	89,15 ± 0,03	468,06 ± 871,57	43122,83 ± 247,33	0,01 ± 0,02	1005,95 ± 81,73	
	PHT	89,10 ± 0,03	350,47 ± 12,88	51200,00 ± 0,00	0,00 ± 0,00	147,50 ± 5,56	
	RDDM	89,21 ± 0,03	121,42 ± 21,54	35102,66 ± 3295,04	0,00 ± 0,00	320,30 ± 41,00	
	SEED	89,17 ± 0,04	50,13 ± 5,54	19100,62 ± 3937,09	0,00 ± 0,00	398,73 ± 94,48	
	SEQDRIFT2	89,19 ± 0,02	200,00 ± 0,00	42433,78 ± 3539,96	0,00 ± 0,00	212,17 ± 17,70	
	STEPD	88,63 ± 0,04	1854,52 ± 1645,38	5187,53 ± 322,92	0,00 ± 0,00	222,07 ± 49,41	
	WSTD	89,24 ± 0,03	460,22 ± 873,62	44064,91 ± 1736,57	0,01 ± 0,02	1389,28 ± 170,60	
	Gradual	PEDD	88,66 ± 0,05	924,50 ± 548,53	10922,74 ± 1155,84	0,00 ± 0,00	43,99 ± 10,50
		RQADD	87,21 ± 0,20	5035,78 ± 909,10	49729,51 ± 989,59	0,01 ± 0,02	11,39 ± 1,58
SEDD		89,26 ± 0,03	495,84 ± 871,18	49568,58 ± 996,54	0,01 ± 0,02	691,15 ± 57,14	
Majority Class		31,14 ± 0,00					
Naive Bayes		73,88 ± 0,05					
Temporally Augmented Classifier		98,91 ± 0,01					
ADWIN		88,63 ± 0,03	1539,73 ± 125,04	28396,23 ± 1667,06	0,00 ± 0,00	19,02 ± 1,39	
CUSUM		88,26 ± 0,04	50022,47 ± 935,50	50194,35 ± 708,60	0,24 ± 0,02	0,77 ± 0,04	
DDM		88,30 ± 0,07	26792,67 ± 3089,92	48651,78 ± 1241,21	0,22 ± 0,04	1,62 ± 0,31	
ECDD		85,89 ± 0,05	3765,12 ± 907,36	888,65 ± 21,19	0,00 ± 0,00	0,40 ± 0,14	
EDDM	86,75 ± 0,57	6158,79 ± 2369,12	14222,67 ± 3334,57	0,06 ± 0,05	3,51 ± 1,16		
FHDDM	88,09 ± 0,05	15044,36 ± 2682,08	42192,44 ± 187,16	0,24 ± 0,04	7,99 ± 6,93		
FTDD	88,49 ± 0,05	16403,09 ± 4657,40	48676,77 ± 1416,87	0,22 ± 0,06	64,18 ± 45,87		
HDDMA	88,25 ± 0,13	41298,12 ± 3148,89	48261,96 ± 1225,08	0,24 ± 0,02	0,93 ± 0,09		
MDDMA	88,12 ± 0,05	43754,81 ± 3191,33	42873,99 ± 433,15	0,37 ± 0,05	0,63 ± 0,03		
PHT	88,45 ± 0,03	16775,22 ± 2555,10	51135,25 ± 15,12	0,32 ± 0,05	11,95 ± 19,76		
RDDM	88,20 ± 0,06	39490,48 ± 2973,01	34323,28 ± 1926,89	0,21 ± 0,04	0,72 ± 0,08		
SEED	88,53 ± 0,04	2351,47 ± 1426,57	13279,67 ± 1537,47	0,00 ± 0,00	11,43 ± 2,05		
SEQDRIFT2	88,51 ± 0,06	18683,33 ± 3637,39	40383,33 ± 1737,57	0,08 ± 0,04	3,92 ± 1,97		
STEPD	87,69 ± 0,06	23300,90 ± 2182,81	17763,79 ± 1880,61	0,16 ± 0,05	0,67 ± 0,10		
WSTD	81,33 ± 0,32	36195,49 ± 3029,36	45289,51 ± 1389,69	0,66 ± 0,06	0,49 ± 0,12		
PEDD	87,99 ± 0,05	8342,72 ± 1201,73	11385,82 ± 1084,47	0,00 ± 0,00	1,53 ± 0,24		
RQADD	87,95 ± 0,37	27570,36 ± 3235,47	49649,58 ± 942,27	0,38 ± 0,05	6,73 ± 11,42		
SEDD	88,37 ± 0,05	46645,61 ± 2572,16	49302,14 ± 971,88	0,27 ± 0,02	0,79 ± 0,06		

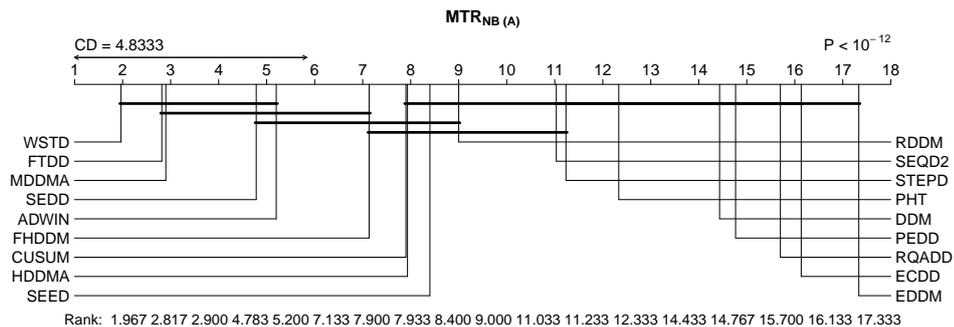
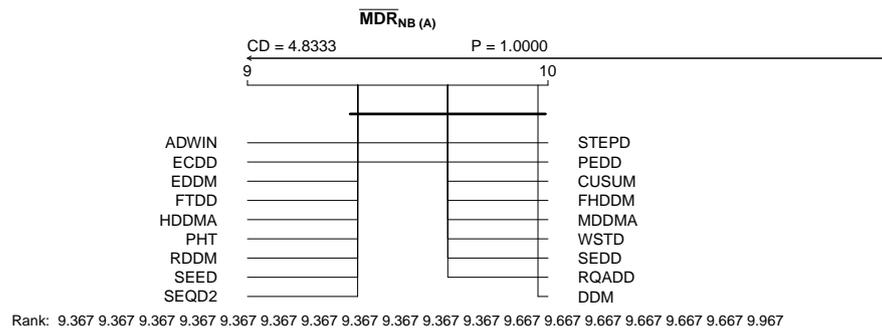
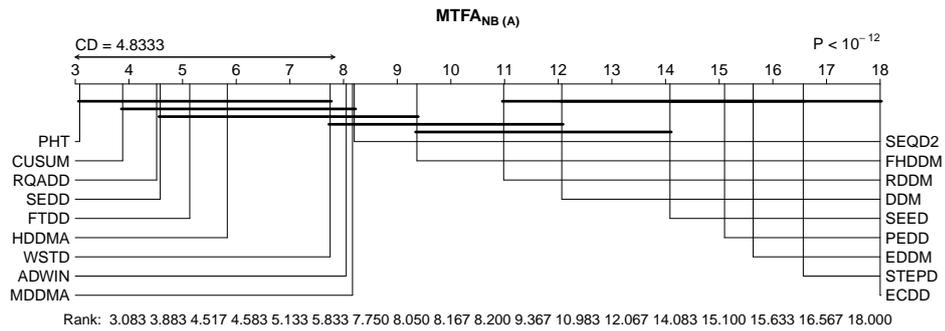
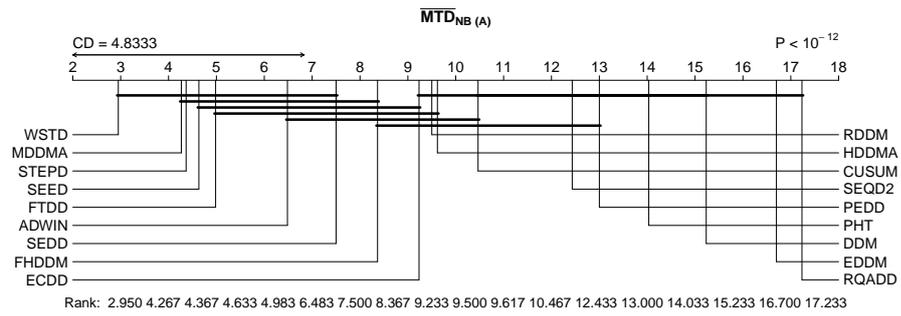
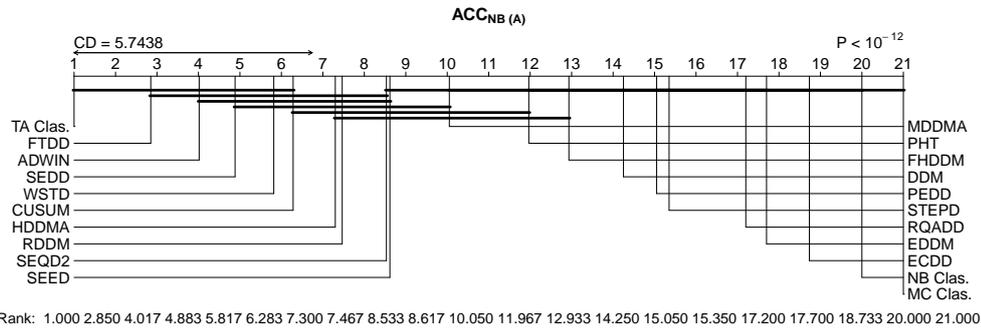
Fonte: Elaborada pelo autor (2021).

Tabela 27 – Resultados obtidos usando NB, considerando a base artificial Agrawal desbalanceada, com um intervalo de confiança de 95%.

<i>Agrawal</i> (<i>D</i>)	<i>Classificador/Detector</i> (<i>NB</i>)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Majority Class	66,94 ± 0,04					
	Naive Bayes	68,91 ± 0,06					
	Temporally Augmented Classifier	86,08 ± 0,35					
	ADWIN	89,11 ± 0,05	62,13 ± 2,90	36515,15 ± 3568,09	0,00 ± 0,00	607,50 ± 83,27	
	CUSUM	89,10 ± 0,04	115,91 ± 6,25	47133,38 ± 1868,12	0,00 ± 0,00	415,61 ± 28,17	
	DDM	88,33 ± 0,36	1095,17 ± 452,13	28291,74 ± 3709,60	0,00 ± 0,00	50,04 ± 12,30	
	ECDD	86,68 ± 0,10	1791,12 ± 1591,70	943,52 ± 693,45	0,01 ± 0,02	30,07 ± 14,55	
	EDDM	86,29 ± 0,22	3852,09 ± 121,45	8380,34 ± 2252,47	0,00 ± 0,00	2,22 ± 0,63	
	FHDDM	88,95 ± 0,04	2164,96 ± 1811,38	41901,26 ± 77,81	0,04 ± 0,04	1186,07 ± 243,52	
	FTDD	89,17 ± 0,04	43,98 ± 8,62	49855,99 ± 1301,76	0,00 ± 0,00	1352,14 ± 183,94	
	HDDMA	89,07 ± 0,05	132,07 ± 28,24	46100,95 ± 1965,63	0,00 ± 0,00	514,96 ± 138,90	
	MDDMA	89,01 ± 0,03	464,44 ± 872,28	42266,00 ± 412,53	0,01 ± 0,02	1097,43 ± 96,79	
	PHT	88,94 ± 0,05	341,63 ± 22,12	51200,00 ± 0,00	0,00 ± 0,00	154,60 ± 10,98	
	RDDM	89,09 ± 0,04	533,94 ± 874,85	28733,70 ± 2733,83	0,00 ± 0,00	277,97 ± 43,26	
	SEED	89,06 ± 0,05	45,60 ± 2,24	25525,60 ± 4224,23	0,00 ± 0,00	568,20 ± 98,62	
	SEQDRIFT2	89,04 ± 0,04	200,00 ± 0,00	39632,33 ± 3381,40	0,00 ± 0,00	198,16 ± 16,91	
	STEPD	88,54 ± 0,05	647,50 ± 861,40	3675,39 ± 711,99	0,00 ± 0,00	150,28 ± 38,71	
	WSTD	89,08 ± 0,06	1310,14 ± 1456,46	43890,40 ± 1322,95	0,02 ± 0,03	1491,99 ± 283,91	
	PEDD	88,20 ± 0,22	1185,57 ± 685,58	11653,72 ± 931,42	0,00 ± 0,00	47,44 ± 14,36	
	RQADD	88,90 ± 0,14	1936,70 ± 1641,06	49743,98 ± 902,54	0,03 ± 0,03	193,02 ± 33,53	
	SEDD	89,13 ± 0,04	67,64 ± 2,56	48530,99 ± 1122,17	0,00 ± 0,00	724,31 ± 30,95	
	Gradual	Majority Class	66,95 ± 0,04				
		Naive Bayes	68,92 ± 0,06				
		Temporally Augmented Classifier	84,96 ± 0,28				
		ADWIN	88,23 ± 0,05	1601,87 ± 883,42	22675,60 ± 1541,95	0,00 ± 0,00	19,15 ± 2,07
		CUSUM	88,29 ± 0,04	43142,20 ± 2323,06	48792,42 ± 1130,25	0,25 ± 0,00	0,86 ± 0,05
		DDM	88,31 ± 0,05	23982,28 ± 4261,43	46516,97 ± 1944,75	0,22 ± 0,03	14,33 ± 18,47
ECDD		86,04 ± 0,05	2276,17 ± 830,49	646,11 ± 18,08	0,00 ± 0,00	1,12 ± 0,54	
EDDM		86,41 ± 0,10	7933,64 ± 2311,65	12527,18 ± 2652,90	0,08 ± 0,04	2,12 ± 0,60	
FHDDM		88,17 ± 0,05	21524,95 ± 4378,21	40189,78 ± 1444,38	0,32 ± 0,06	12,76 ± 13,48	
FTDD		88,35 ± 0,04	27362,18 ± 3456,32	46557,77 ± 1722,59	0,25 ± 0,04	1,45 ± 0,24	
HDDMA		88,34 ± 0,04	37460,73 ± 3580,14	47137,60 ± 1472,92	0,25 ± 0,03	1,04 ± 0,20	
MDDMA		88,21 ± 0,04	46972,13 ± 2795,26	42364,79 ± 155,06	0,28 ± 0,03	0,67 ± 0,06	
PHT		88,41 ± 0,04	23124,01 ± 3403,39	51156,59 ± 13,79	0,32 ± 0,04	1,83 ± 0,34	
RDDM		88,31 ± 0,04	34821,73 ± 3791,38	26892,49 ± 1942,91	0,12 ± 0,05	0,74 ± 0,13	
SEED		88,18 ± 0,07	1028,80 ± 117,11	13433,61 ± 1922,12	0,00 ± 0,00	14,06 ± 2,21	
SEQDRIFT2		88,23 ± 0,06	10930,00 ± 2930,97	32880,94 ± 3088,70	0,00 ± 0,00	11,21 ± 5,12	
STEPD		87,82 ± 0,06	13706,88 ± 2774,54	10575,56 ± 1857,04	0,11 ± 0,05	2,90 ± 3,10	
WSTD		84,94 ± 1,04	13996,75 ± 3623,04	39028,11 ± 2648,43	0,16 ± 0,06	5,62 ± 2,31	
PEDD		87,35 ± 0,41	5773,09 ± 1275,84	12365,81 ± 1210,21	0,00 ± 0,00	3,43 ± 1,62	
RQADD		88,33 ± 0,11	27766,09 ± 5440,30	49741,96 ± 826,42	0,33 ± 0,06	35,92 ± 48,64	
SEDD		88,34 ± 0,04	47889,62 ± 2032,11	48309,98 ± 1050,77	0,26 ± 0,02	0,76 ± 0,04	

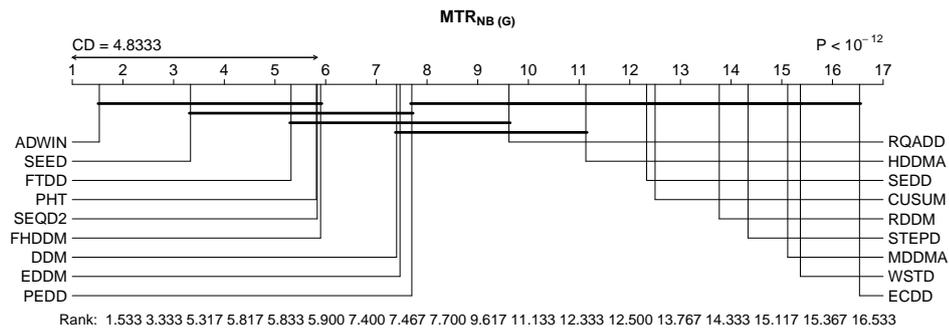
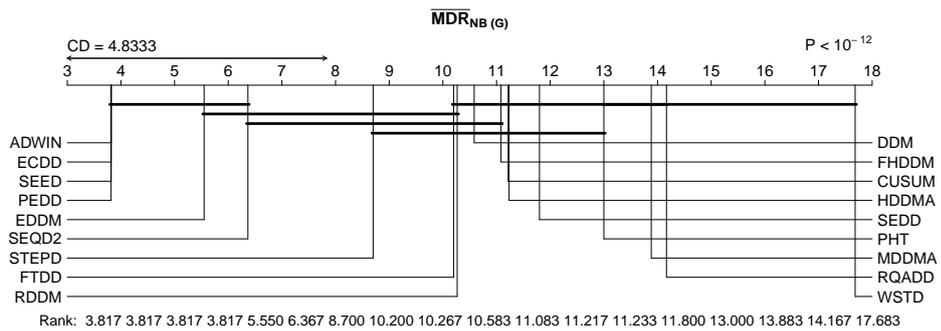
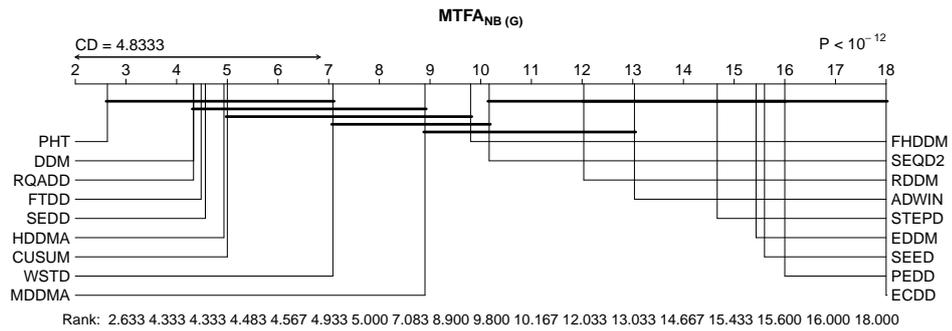
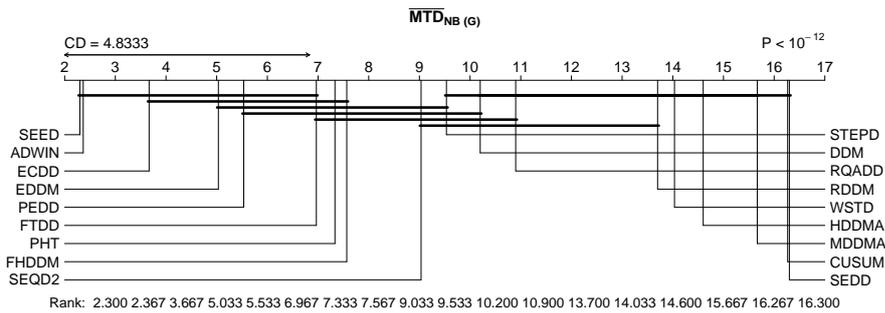
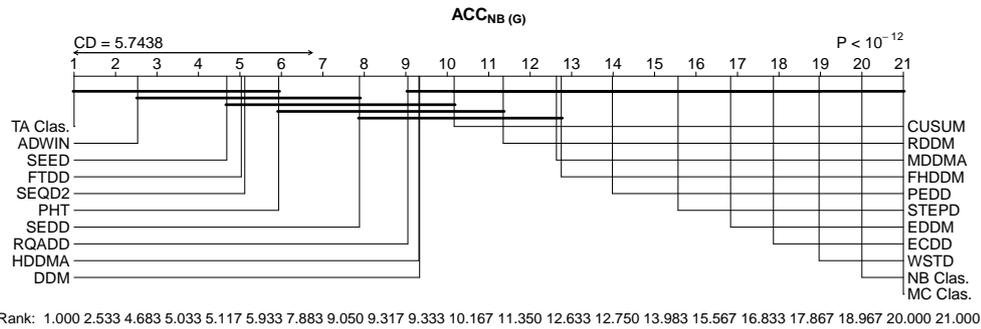
Fonte: Elaborada pelo autor (2021).

Figura 46 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.



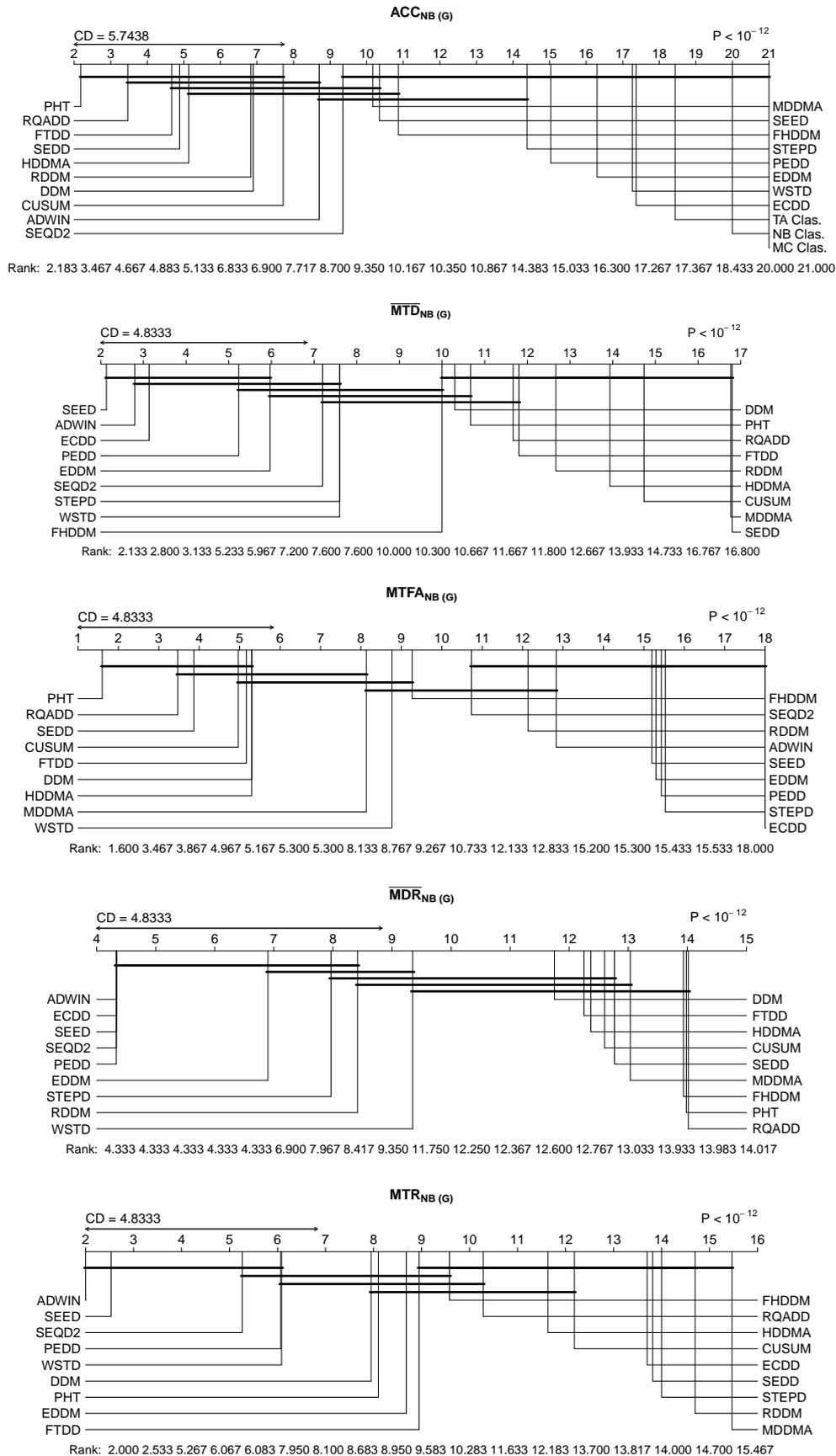
Fonte: Elaborada pelo autor (2021).

Figura 47 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal balanceada, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 49 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base Agrawal desbalanceada, considerando um intervalo de confiança de 95%.



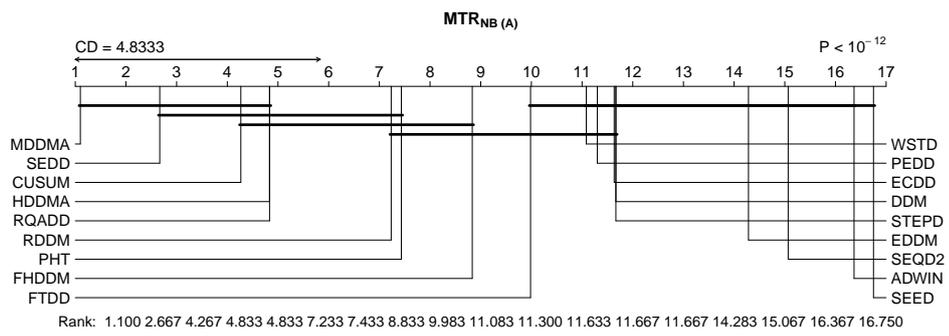
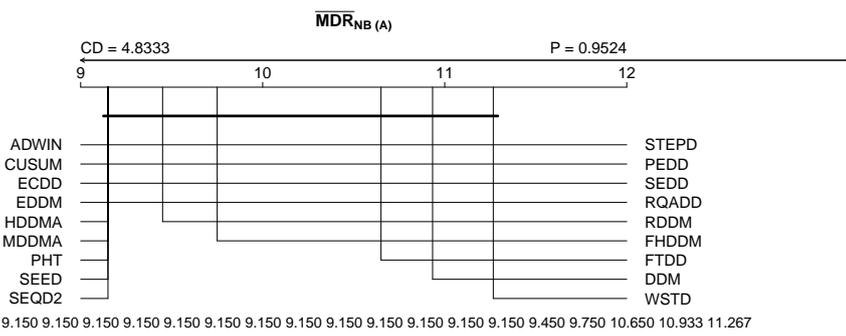
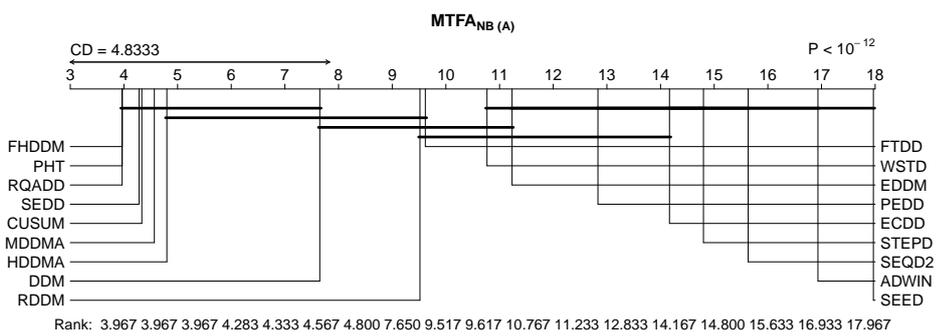
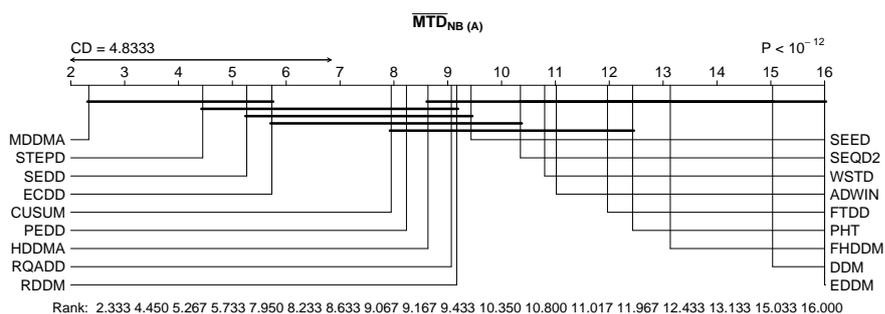
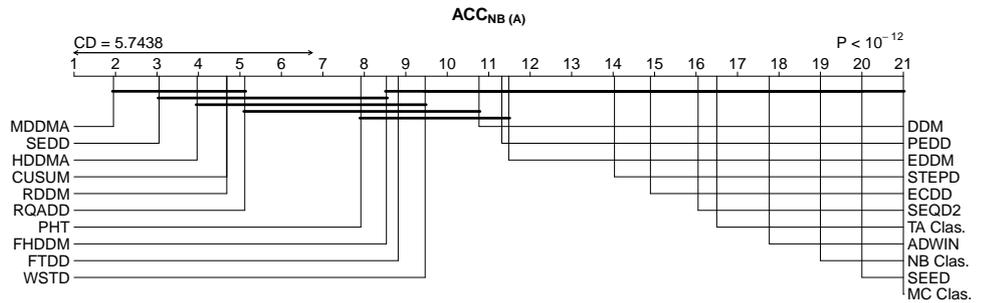
Fonte: Elaborada pelo autor (2021).

Tabela 28 – Resultados obtidos usando NB, considerando a base artificial LED tendo 5% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.

$LED_{(5\%)}$	<i>Classificador/Detector</i> _(NB)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Majority Class	10,11 ± 0,05					
	Naive Bayes	69,38 ± 0,05					
	Temporally Augmented Classifier	80,23 ± 0,38					
	ADWIN	74,66 ± 0,92	4271,20 ± 4335,01	369,07 ± 5,69	0,00 ± 0,00	1,09 ± 0,33	
	CUSUM	86,48 ± 0,06	234,32 ± 22,30	50416,05 ± 1136,79	0,00 ± 0,00	233,87 ± 31,16	
	DDM	85,13 ± 0,54	5981,33 ± 2319,27	38670,17 ± 3925,87	0,05 ± 0,04	17,71 ± 6,86	
	ECDD	82,93 ± 0,12	555,09 ± 307,12	1426,40 ± 44,18	0,00 ± 0,00	26,38 ± 13,42	
	EDDM	85,25 ± 0,11	7013,57 ± 484,57	15591,80 ± 2171,32	0,00 ± 0,00	2,20 ± 0,26	
	FHDDM	85,91 ± 0,25	4209,91 ± 1820,26	51200,00 ± 0,00	0,02 ± 0,02	106,57 ± 64,14	
	FTDD	85,92 ± 0,26	4008,71 ± 1887,60	28777,46 ± 5146,85	0,04 ± 0,04	87,84 ± 48,23	
	HDDMA	86,48 ± 0,06	369,97 ± 126,70	48135,47 ± 3164,05	0,00 ± 0,00	254,38 ± 78,36	
	MDDMA	86,52 ± 0,06	76,88 ± 9,03	49210,73 ± 2309,57	0,00 ± 0,00	697,33 ± 83,84	
	PHT	86,37 ± 0,06	743,83 ± 71,04	51200,00 ± 0,00	0,00 ± 0,00	73,62 ± 7,58	
	RDDM	86,47 ± 0,05	1034,10 ± 1090,57	29679,76 ± 5025,46	0,01 ± 0,02	115,49 ± 27,19	
	SEED	64,00 ± 0,14	11252,80 ± 7041,06	193,08 ± 1,24	0,00 ± 0,00	0,92 ± 0,29	
	SEQDRIFT2	79,40 ± 1,34	10001,67 ± 6920,23	2102,74 ± 2831,91	0,00 ± 0,00	9,83 ± 14,22	
	STEPD	83,50 ± 0,30	1141,44 ± 1311,49	1181,76 ± 155,28	0,00 ± 0,00	18,17 ± 5,41	
	WSTD	85,56 ± 0,42	6263,43 ± 3346,30	21829,44 ± 5157,57	0,08 ± 0,06	73,98 ± 46,45	
	PEDD	85,30 ± 0,08	1000,86 ± 779,76	4121,59 ± 215,97	0,00 ± 0,00	22,32 ± 6,39	
	RQADD	86,48 ± 0,06	279,04 ± 47,89	51200,00 ± 0,00	0,00 ± 0,00	201,13 ± 15,90	
	SEDD	86,51 ± 0,06	541,52 ± 820,29	50609,33 ± 1053,93	0,00 ± 0,00	404,13 ± 53,91	
	Gradual	Majority Class	10,11 ± 0,05				
		Naive Bayes	69,38 ± 0,05				
		Temporally Augmented Classifier	80,08 ± 0,31				
		ADWIN	74,63 ± 0,58	1894,13 ± 1417,57	411,97 ± 36,99	0,00 ± 0,00	0,81 ± 0,23
		CUSUM	86,30 ± 0,05	48312,35 ± 2580,30	50834,41 ± 412,58	0,28 ± 0,03	0,77 ± 0,03
		DDM	85,53 ± 0,50	4851,73 ± 2903,35	43303,86 ± 2445,16	0,03 ± 0,04	46,17 ± 21,69
		ECDD	82,70 ± 0,11	897,59 ± 203,73	1424,00 ± 42,74	0,00 ± 0,00	2,36 ± 0,81
		EDDM	85,16 ± 0,26	7200,20 ± 697,38	15684,77 ± 1644,34	0,01 ± 0,02	2,23 ± 0,28
		FHDDM	85,14 ± 0,23	9195,25 ± 2142,92	51200,00 ± 0,00	0,03 ± 0,03	8,32 ± 2,85
FTDD		84,60 ± 0,30	13215,26 ± 2919,24	30211,60 ± 4209,98	0,12 ± 0,05	2,77 ± 0,82	
HDDMA		86,34 ± 0,06	15103,55 ± 5166,86	48789,19 ± 1837,26	0,19 ± 0,07	32,50 ± 19,58	
MDDMA		86,27 ± 0,05	21827,09 ± 5988,35	49835,58 ± 1266,04	0,21 ± 0,06	16,23 ± 11,66	
PHT		86,27 ± 0,05	2328,50 ± 2037,65	51196,60 ± 4,28	0,03 ± 0,04	83,62 ± 20,25	
RDDM		86,30 ± 0,05	29037,73 ± 4622,35	35457,23 ± 2582,04	0,08 ± 0,04	5,18 ± 8,01	
SEED		63,82 ± 0,07	5239,47 ± 2373,94	193,29 ± 0,77	0,00 ± 0,00	1,01 ± 0,37	
SEQDRIFT2		79,76 ± 0,91	16376,67 ± 4400,80	4261,49 ± 2325,52	0,03 ± 0,04	0,31 ± 0,12	
STEPD		83,48 ± 0,28	4727,12 ± 1189,09	1360,17 ± 177,77	0,00 ± 0,00	0,64 ± 0,37	
WSTD		82,40 ± 0,54	27316,99 ± 2797,80	28674,47 ± 2871,94	0,31 ± 0,06	0,78 ± 0,14	
PEDD		85,10 ± 0,07	3706,92 ± 611,47	4224,73 ± 197,50	0,00 ± 0,00	1,65 ± 0,63	
RQADD		86,20 ± 0,06	8409,77 ± 2845,01	51160,45 ± 23,70	0,13 ± 0,05	21,36 ± 8,63	
SEDD		86,24 ± 0,06	17604,85 ± 5326,84	50691,19 ± 835,92	0,22 ± 0,07	21,49 ± 13,20	

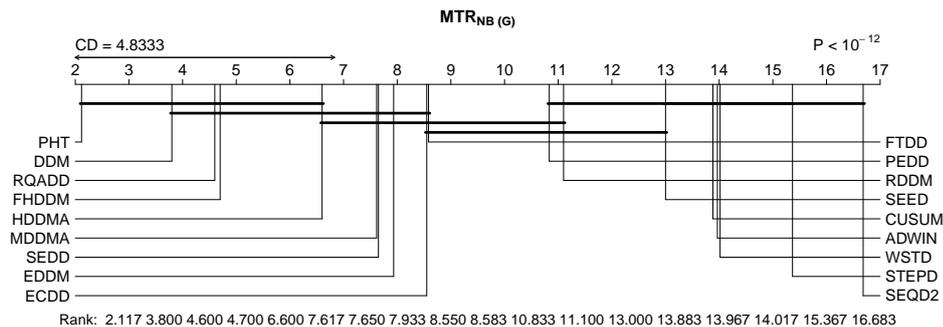
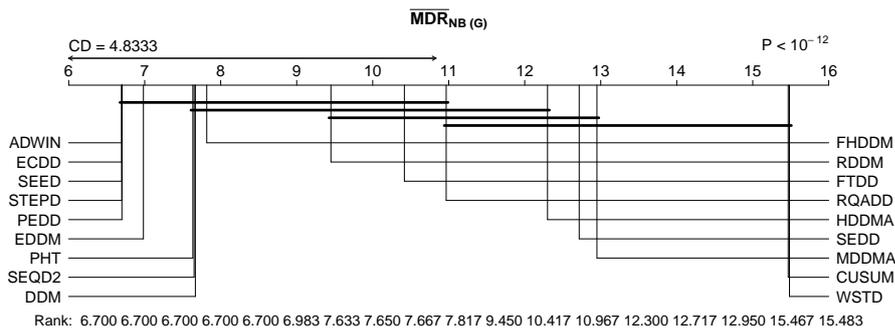
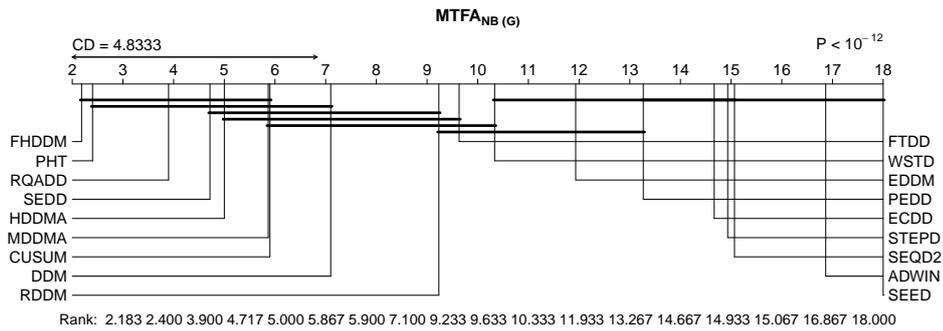
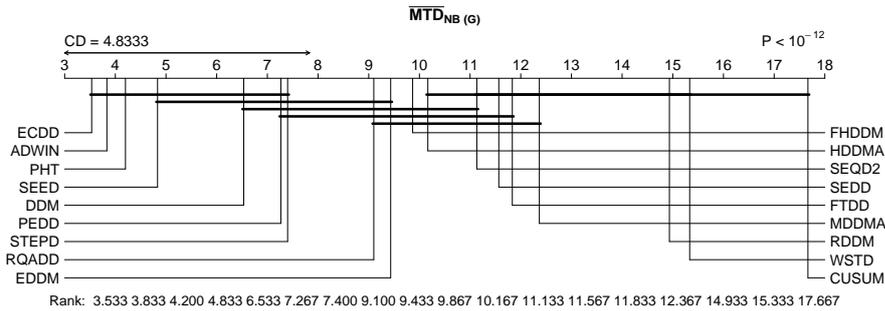
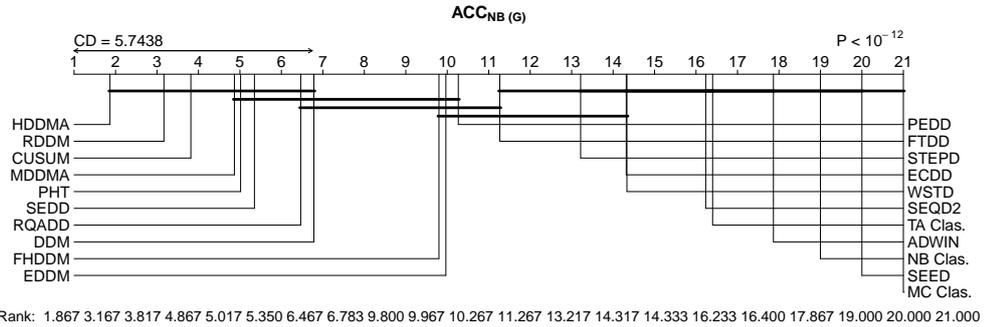
Fonte: Elaborada pelo autor (2021).

Figura 50 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 51 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 5% de ruído, considerando um intervalo de confiança de 95%.



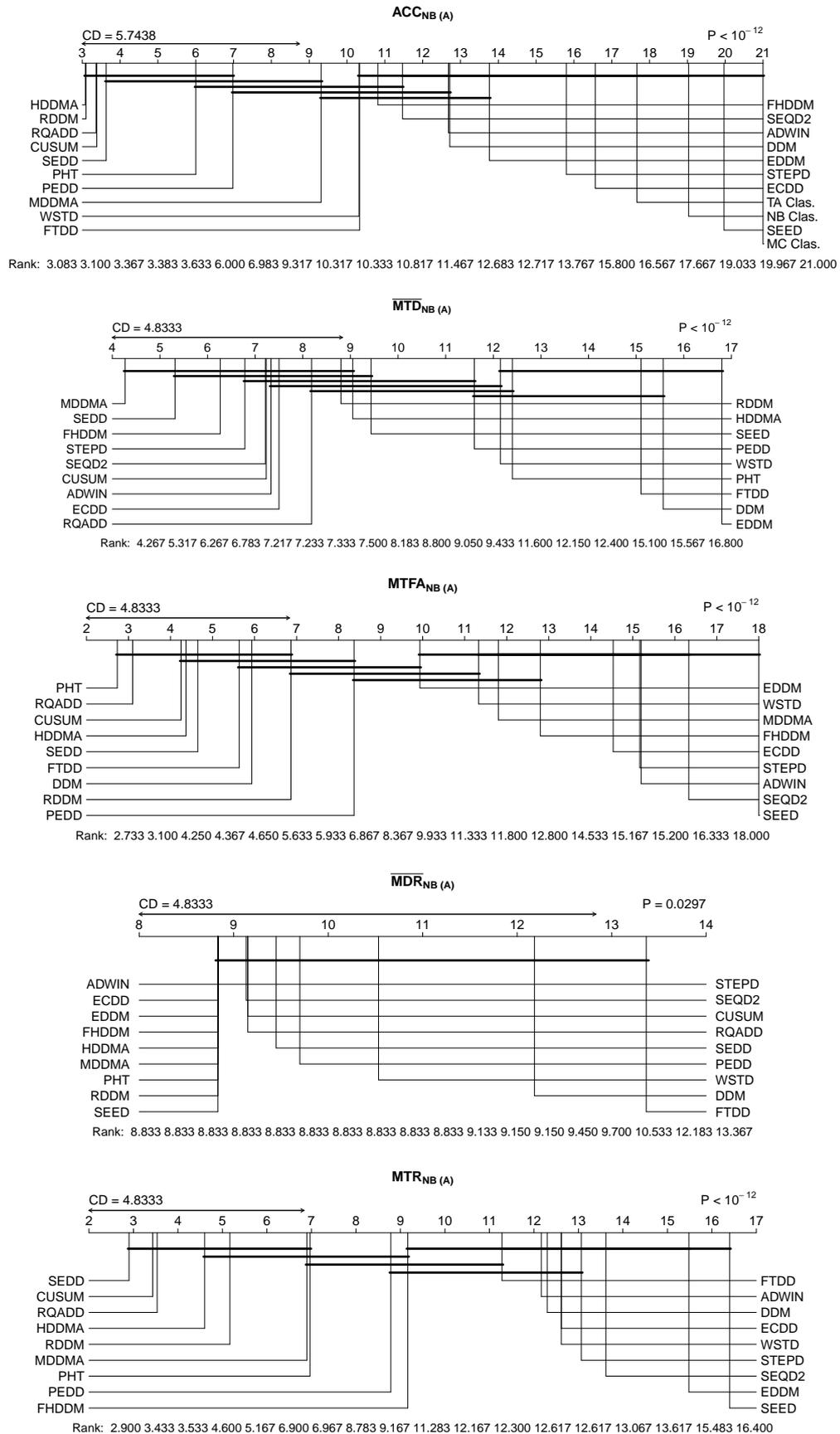
Fonte: Elaborada pelo autor (2021).

Tabela 29 – Resultados obtidos usando NB, considerando a base artificial LED tendo 10% de probabilidade dos atributos serem invertidos (ruído), com um intervalo de confiança de 95%.

$LED_{(10\%)}$	Classificador/Detector _(NB)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Majority Class	10,11 ± 0,05					
	Naive Bayes	59,90 ± 0,07					
	Temporally Augmented Classifier	67,84 ± 0,25					
	ADWIN	71,79 ± 0,56	300,80 ± 70,08	2448,28 ± 1430,06	0,00 ± 0,00	8,68 ± 4,86	
	CUSUM	73,73 ± 0,09	1110,94 ± 1732,60	46590,93 ± 3194,07	0,02 ± 0,03	184,42 ± 28,36	
	DDM	71,20 ± 0,84	11598,53 ± 3788,50	41238,18 ± 3111,63	0,13 ± 0,07	7,71 ± 2,88	
	ECDD	69,15 ± 0,15	586,90 ± 225,43	1294,73 ± 35,25	0,00 ± 0,00	13,48 ± 10,86	
	EDDM	71,55 ± 0,12	15251,33 ± 490,43	19984,32 ± 1726,67	0,00 ± 0,00	1,32 ± 0,13	
	FHDDM	72,84 ± 0,14	953,33 ± 613,23	6818,15 ± 849,00	0,00 ± 0,00	37,75 ± 13,17	
	FTDD	72,52 ± 0,40	10427,44 ± 3578,20	41306,24 ± 3699,26	0,15 ± 0,06	30,79 ± 23,27	
	HDDMA	73,72 ± 0,08	497,32 ± 125,77	46104,25 ± 3310,11	0,00 ± 0,00	158,87 ± 58,80	
	MDDMA	73,18 ± 0,09	1770,40 ± 1874,56	9285,01 ± 995,36	0,00 ± 0,00	90,71 ± 24,37	
	PHT	73,61 ± 0,09	920,55 ± 106,98	51200,00 ± 0,00	0,00 ± 0,00	61,96 ± 8,48	
	RDDM	73,73 ± 0,09	420,28 ± 102,82	37529,88 ± 4229,77	0,00 ± 0,00	123,37 ± 32,43	
	SEED	57,01 ± 0,31	7944,80 ± 5800,54	292,21 ± 3,34	0,00 ± 0,00	1,07 ± 0,38	
	SEQDRIFT2	72,22 ± 0,64	691,67 ± 865,68	2486,60 ± 2883,75	0,01 ± 0,02	9,12 ± 13,86	
	STEPD	69,93 ± 0,46	1030,31 ± 642,89	1057,98 ± 128,25	0,00 ± 0,00	6,73 ± 2,52	
	WSTD	72,78 ± 0,27	7372,04 ± 2961,28	12183,74 ± 2715,16	0,05 ± 0,04	24,53 ± 19,14	
	PEDD	73,50 ± 0,10	4726,57 ± 3077,87	28903,85 ± 5016,07	0,02 ± 0,03	57,04 ± 22,68	
	RQADD	73,71 ± 0,09	1169,58 ± 1732,08	49808,34 ± 1633,01	0,02 ± 0,03	182,09 ± 33,80	
	SEDD	73,70 ± 0,09	1500,55 ± 1914,56	44205,24 ± 3831,71	0,02 ± 0,04	267,17 ± 59,69	
	Gradual	Majority Class	10,11 ± 0,05				
		Naive Bayes	59,90 ± 0,07				
		Temporally Augmented Classifier	67,83 ± 0,27				
		ADWIN	71,13 ± 0,35	2190,93 ± 1567,99	2823,58 ± 1150,16	0,00 ± 0,00	5,34 ± 2,72
		CUSUM	73,58 ± 0,08	37719,22 ± 5736,17	49381,08 ± 1417,34	0,25 ± 0,04	2,73 ± 3,24
		DDM	71,43 ± 0,81	11395,20 ± 3648,94	42492,38 ± 3054,38	0,11 ± 0,07	9,00 ± 3,59
		ECDD	69,01 ± 0,15	878,23 ± 148,26	1304,49 ± 34,01	0,00 ± 0,00	1,87 ± 0,42
		EDDM	71,62 ± 0,12	14920,34 ± 476,06	20888,83 ± 2231,77	0,00 ± 0,00	1,40 ± 0,14
		FHDDM	72,64 ± 0,14	3273,07 ± 629,88	7015,16 ± 790,94	0,00 ± 0,00	2,88 ± 0,77
FTDD		71,19 ± 0,33	21536,91 ± 3322,78	42957,93 ± 3074,68	0,29 ± 0,07	2,46 ± 1,76	
HDDMA		73,59 ± 0,08	10061,39 ± 4976,07	44871,80 ± 2961,60	0,11 ± 0,05	39,90 ± 16,83	
MDDMA		72,99 ± 0,09	7056,12 ± 1959,70	11024,40 ± 1563,60	0,00 ± 0,00	2,74 ± 1,05	
PHT		73,54 ± 0,08	1652,68 ± 1154,46	51198,32 ± 2,51	0,02 ± 0,02	69,43 ± 15,54	
RDDM		73,62 ± 0,09	27662,70 ± 5618,83	38911,56 ± 2655,42	0,11 ± 0,05	6,51 ± 6,00	
SEED		57,13 ± 0,38	6235,73 ± 2705,18	317,33 ± 41,67	0,00 ± 0,00	0,74 ± 0,32	
SEQDRIFT2		71,96 ± 0,38	4906,67 ± 1913,68	7062,17 ± 2896,27	0,02 ± 0,03	4,59 ± 3,43	
STEPD		69,77 ± 0,40	2295,97 ± 1051,41	1112,63 ± 156,12	0,00 ± 0,00	1,00 ± 0,35	
WSTD		71,98 ± 0,31	15752,38 ± 3098,26	15864,49 ± 3097,02	0,07 ± 0,04	1,05 ± 0,25	
PEDD		73,33 ± 0,12	18842,35 ± 3734,39	34199,62 ± 3943,84	0,10 ± 0,05	3,49 ± 2,35	
RQADD		73,45 ± 0,11	13291,42 ± 4552,29	48950,40 ± 1542,06	0,17 ± 0,07	17,68 ± 10,11	
SEDD		73,45 ± 0,08	12367,90 ± 3643,38	44536,38 ± 2615,65	0,14 ± 0,06	13,07 ± 9,63	

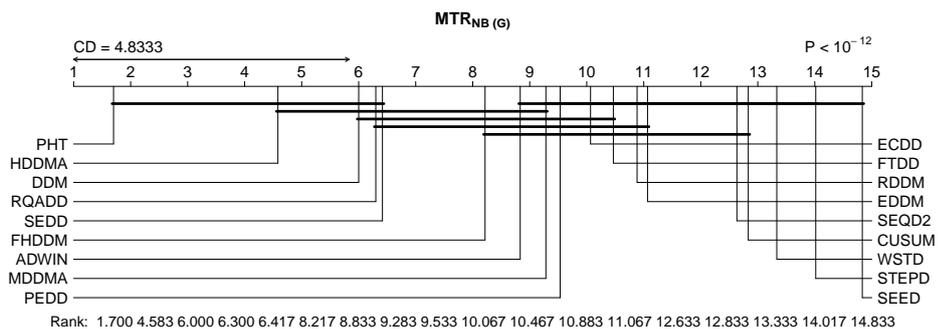
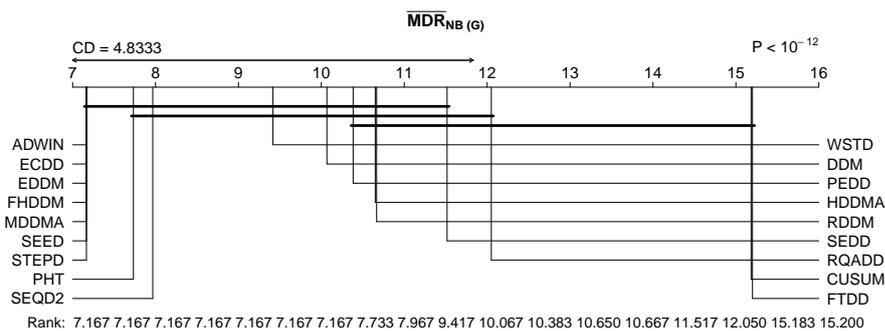
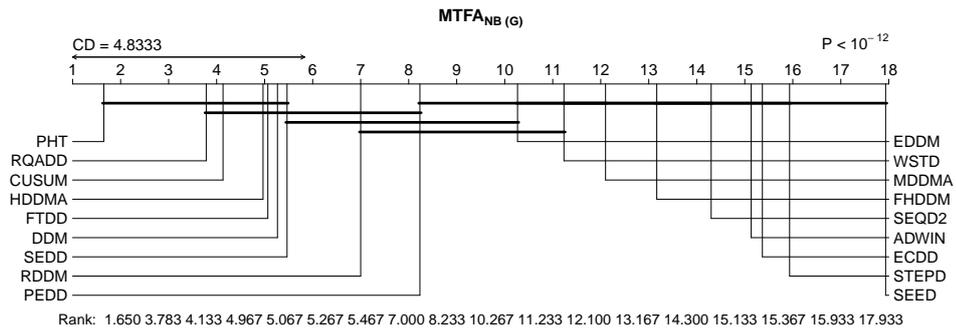
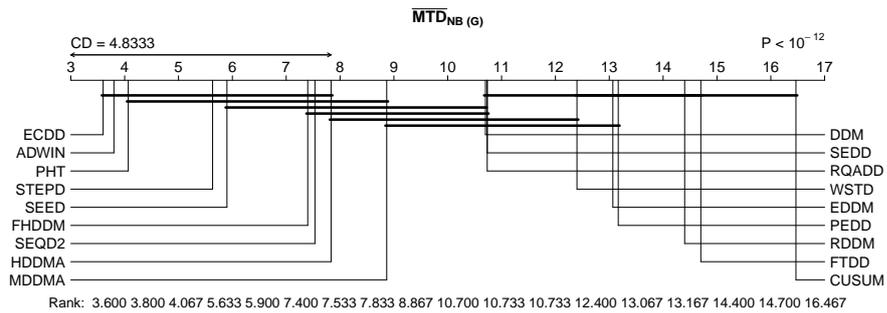
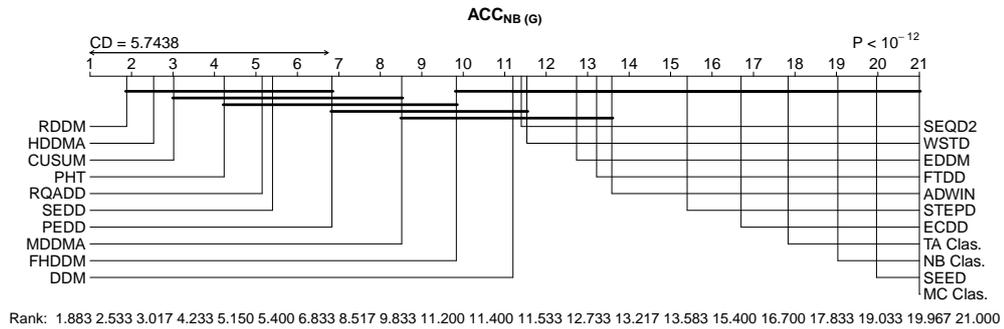
Fonte: Elaborada pelo autor (2021).

Figura 52 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 53 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base LED com 10% de ruído, considerando um intervalo de confiança de 95%.



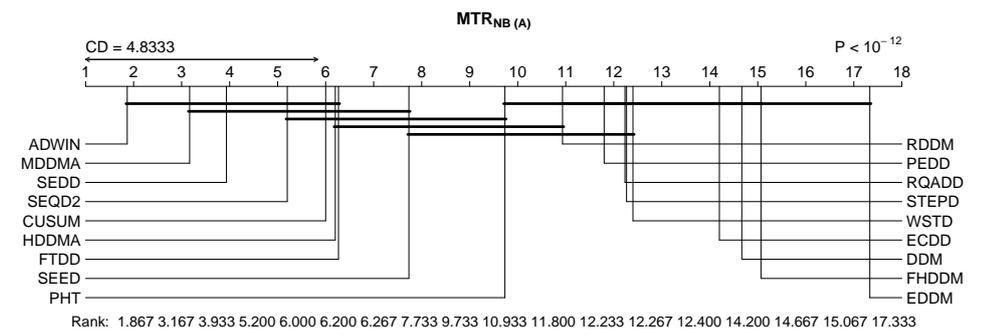
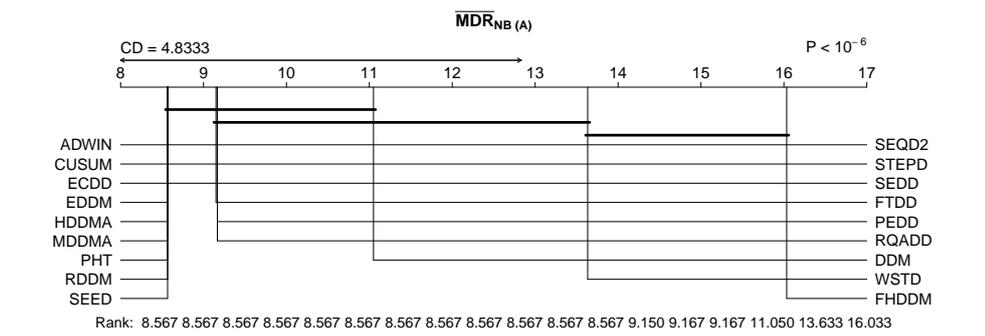
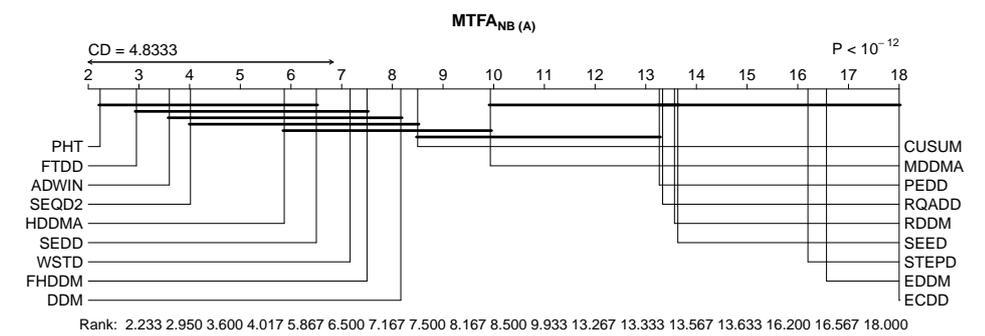
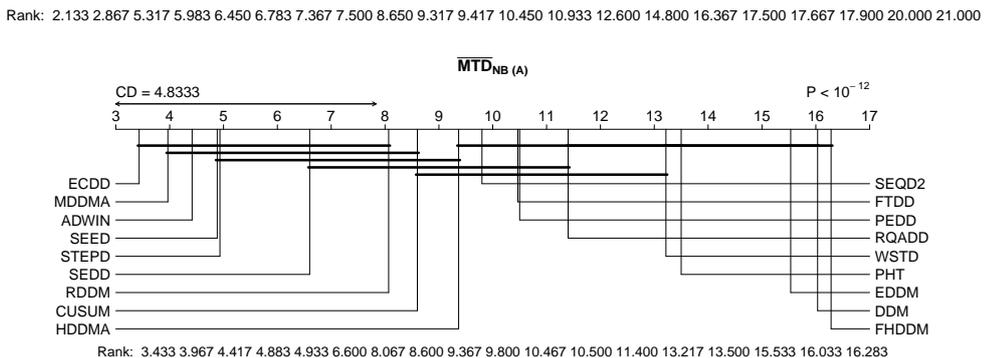
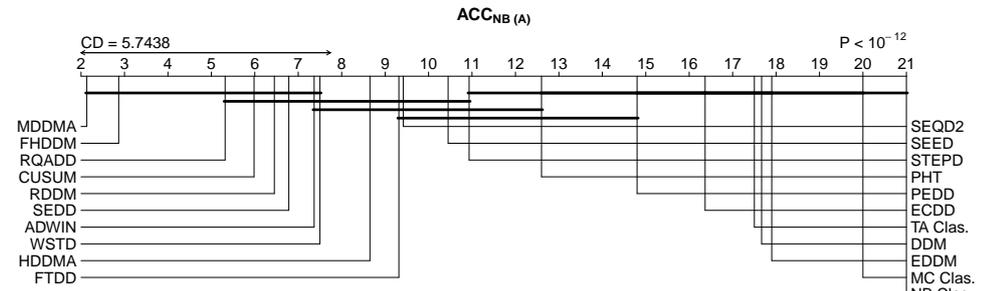
Fonte: Elaborada pelo autor (2021).

Tabela 30 – Resultados obtidos usando NB, considerando a base artificial RBF com duas classes, com um intervalo de confiança de 95%.

$RBF_{(2)}$	<i>Classificador/Detector</i> _(NB)	Acurácia(%)	MTD	MTFA	MDR	MTR	
Abrupta	Majority Class	70,01 ± 0,03					
	Naive Bayes	69,59 ± 0,04					
	Temporally Augmented Classifier	82,11 ± 0,11					
	ADWIN	83,00 ± 0,04	105,60 ± 8,43	39868,03 ± 2359,87	0,00 ± 0,00	392,86 ± 37,34	
	CUSUM	83,04 ± 0,04	184,83 ± 8,81	28873,49 ± 1636,49	0,00 ± 0,00	159,26 ± 13,13	
	DDM	79,96 ± 1,34	7958,22 ± 4261,98	29229,60 ± 3783,93	0,12 ± 0,09	15,25 ± 7,32	
	ECDD	82,38 ± 0,03	88,28 ± 33,53	454,02 ± 6,18	0,00 ± 0,00	16,24 ± 7,47	
	EDDM	81,95 ± 0,12	3304,61 ± 255,75	2783,12 ± 1425,01	0,00 ± 0,00	0,82 ± 0,41	
	FHDDM	83,13 ± 0,03	10769,27 ± 1783,91	31783,49 ± 76,22	0,21 ± 0,04	25,32 ± 21,40	
	FTDD	82,97 ± 0,04	1388,24 ± 1289,01	41895,07 ± 2270,32	0,02 ± 0,02	212,37 ± 76,98	
	HDDMA	82,97 ± 0,04	241,60 ± 43,95	34638,77 ± 2939,47	0,00 ± 0,00	177,48 ± 37,00	
	MDDMA	83,16 ± 0,04	100,53 ± 21,59	25714,19 ± 1468,98	0,00 ± 0,00	340,42 ± 69,07	
	PHT	82,87 ± 0,04	523,87 ± 26,57	44635,86 ± 2370,79	0,00 ± 0,00	86,80 ± 6,62	
	RDDM	83,02 ± 0,05	206,35 ± 57,41	11692,43 ± 1894,50	0,00 ± 0,00	68,91 ± 12,53	
	SEED	82,94 ± 0,04	1363,47 ± 1455,34	12373,82 ± 2173,00	0,00 ± 0,00	152,48 ± 42,42	
	SEQDRIFT2	82,96 ± 0,04	630,00 ± 868,90	38982,67 ± 2325,77	0,00 ± 0,00	183,92 ± 17,40	
	STEPD	82,92 ± 0,04	247,48 ± 158,20	2024,75 ± 54,05	0,00 ± 0,00	53,70 ± 16,24	
	WSTD	83,00 ± 0,04	8205,41 ± 2214,09	31111,27 ± 1469,54	0,14 ± 0,05	122,01 ± 86,86	
	PEDD	82,70 ± 0,04	1322,47 ± 1278,82	13146,35 ± 132,70	0,02 ± 0,02	57,43 ± 11,31	
	RQADD	82,94 ± 0,36	2270,07 ± 1940,29	12947,70 ± 2948,84	0,02 ± 0,02	52,58 ± 15,20	
	SEDD	83,01 ± 0,03	150,08 ± 26,69	32741,51 ± 1690,58	0,00 ± 0,00	262,69 ± 41,77	
	Gradual	Majority Class	70,00 ± 0,03				
		Naive Bayes	69,58 ± 0,04				
		Temporally Augmented Classifier	82,03 ± 0,09				
		ADWIN	82,69 ± 0,04	15058,93 ± 2983,06	38480,04 ± 2425,37	0,11 ± 0,05	5,26 ± 3,53
		CUSUM	82,67 ± 0,03	16302,21 ± 1580,67	27289,88 ± 1507,60	0,26 ± 0,02	1,31 ± 0,14
		DDM	80,90 ± 1,05	10741,00 ± 3624,24	29593,01 ± 3460,71	0,17 ± 0,07	8,06 ± 3,89
ECDD		82,09 ± 0,04	472,43 ± 154,91	459,23 ± 6,20	0,00 ± 0,00	1,40 ± 0,26	
EDDM		81,72 ± 0,45	4882,11 ± 1873,98	4034,70 ± 2213,22	0,03 ± 0,03	0,73 ± 0,39	
FHDDM		82,77 ± 0,05	19303,12 ± 2404,69	31149,14 ± 953,34	0,33 ± 0,06	1,25 ± 0,23	
FTDD		82,60 ± 0,07	18223,30 ± 3400,32	46935,91 ± 1724,61	0,23 ± 0,07	3,65 ± 1,92	
HDDMA		82,65 ± 0,04	14892,38 ± 3490,73	37354,56 ± 2058,08	0,18 ± 0,07	11,33 ± 9,49	
MDDMA		82,80 ± 0,04	7733,92 ± 2389,87	25545,26 ± 1320,21	0,12 ± 0,05	14,56 ± 6,58	
PHT		82,66 ± 0,04	21431,33 ± 3175,59	46409,53 ± 1659,32	0,33 ± 0,06	3,20 ± 3,24	
RDDM		82,72 ± 0,04	11388,67 ± 2621,81	12751,65 ± 1884,45	0,01 ± 0,02	2,00 ± 0,72	
SEED		82,59 ± 0,05	5041,60 ± 1449,69	11998,80 ± 1456,86	0,01 ± 0,02	3,87 ± 1,10	
SEQDRIFT2		82,64 ± 0,04	24306,67 ± 2995,47	37944,98 ± 2352,25	0,20 ± 0,05	2,28 ± 1,95	
STEPD		82,54 ± 0,05	2456,41 ± 701,37	2185,09 ± 91,77	0,00 ± 0,00	1,35 ± 0,32	
WSTD		82,23 ± 0,20	18560,97 ± 2825,33	32720,92 ± 1169,27	0,27 ± 0,06	1,80 ± 0,62	
PEDD		82,26 ± 0,10	11302,90 ± 2311,89	13365,98 ± 136,82	0,12 ± 0,05	1,79 ± 0,59	
RQADD		82,71 ± 0,35	2916,22 ± 2476,45	5389,77 ± 1445,60	0,02 ± 0,03	13,88 ± 4,27	
SEDD		82,62 ± 0,05	13432,15 ± 3177,67	32849,89 ± 1833,09	0,18 ± 0,06	6,92 ± 4,13	

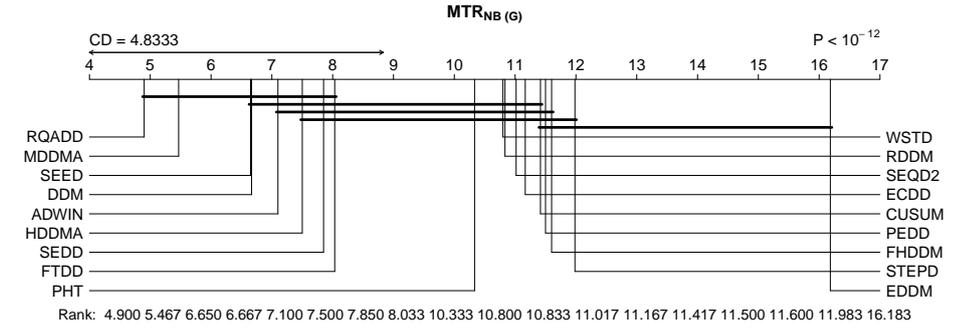
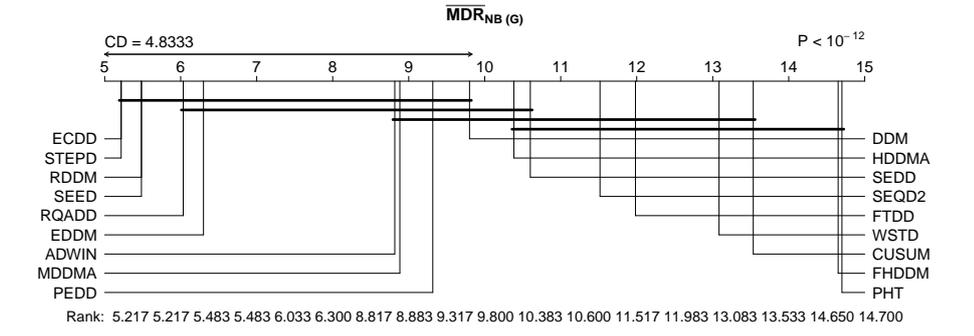
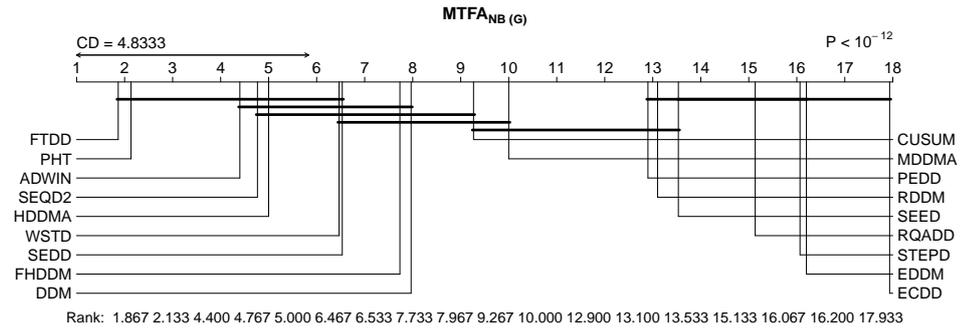
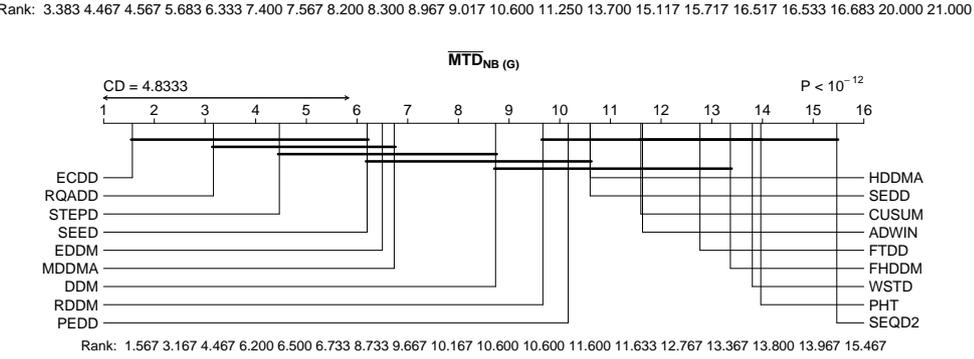
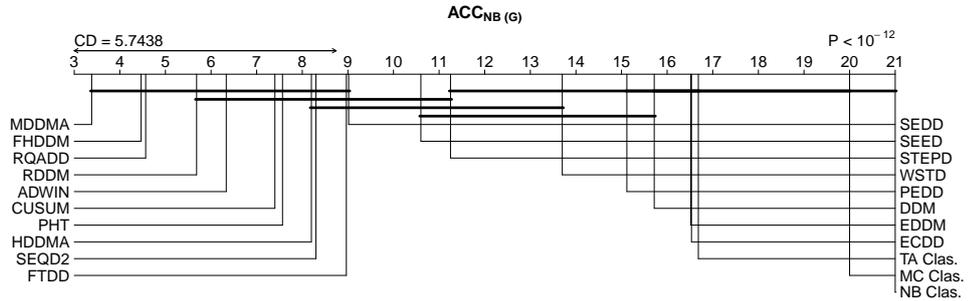
Fonte: Elaborada pelo autor (2021).

Figura 54 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com duas classes, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 55 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com duas classes, considerando um intervalo de confiança de 95%.



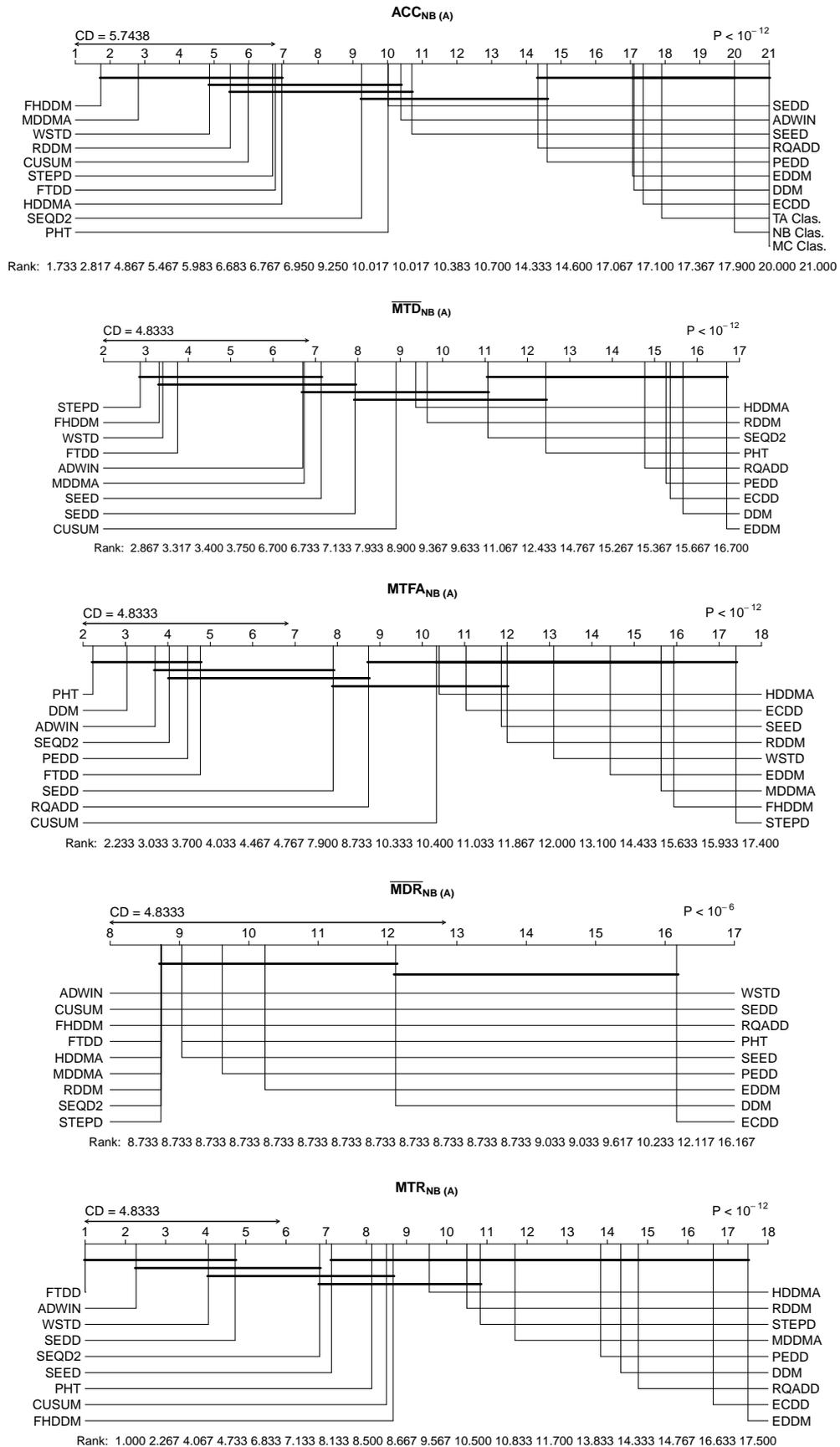
Fonte: Elaborada pelo autor (2021).

Tabela 31 – Resultados obtidos usando NB, considerando a base artificial RBF com cinco classes, com um intervalo de confiança de 95%.

$RBF_{(5)}$	Classificador/Detector _(NB)	Acurácia(%)	MTD	MTFA	MDR	MTR	
	Majority Class	32,98 ± 0,02					
	Naive Bayes	43,13 ± 0,04					
	Temporally Augmented Classifier	61,02 ± 0,25					
Abrupta	ADWIN	64,64 ± 0,07	67,20 ± 3,81	32825,89 ± 1633,79	0,00 ± 0,00	497,37 ± 34,37	
	CUSUM	64,84 ± 0,06	292,31 ± 396,42	12689,01 ± 1821,54	0,00 ± 0,00	124,76 ± 17,47	
	DDM	60,94 ± 1,22	8182,77 ± 3612,88	36422,05 ± 3308,25	0,12 ± 0,07	21,80 ± 9,57	
	ECDD	61,52 ± 0,47	10701,92 ± 1800,36	12335,70 ± 2510,91	0,21 ± 0,04	9,13 ± 14,27	
	EDDM	61,97 ± 0,54	7965,98 ± 1657,14	6574,43 ± 2342,58	0,04 ± 0,04	0,84 ± 0,37	
	FHDDM	65,11 ± 0,06	97,92 ± 66,93	2486,80 ± 99,38	0,00 ± 0,00	129,53 ± 26,85	
	FTDD	64,82 ± 0,06	27,41 ± 3,04	29754,34 ± 2540,33	0,00 ± 0,00	1129,84 ± 106,27	
	HDDMA	64,80 ± 0,08	169,09 ± 49,86	14052,36 ± 2391,91	0,00 ± 0,00	145,47 ± 68,73	
	MDDMA	65,03 ± 0,06	127,33 ± 82,51	2600,67 ± 120,36	0,00 ± 0,00	66,31 ± 15,34	
	PHT	64,69 ± 0,06	792,24 ± 876,82	38251,17 ± 2000,03	0,01 ± 0,02	134,47 ± 18,82	
	RDDM	64,87 ± 0,10	183,22 ± 67,01	9955,42 ± 1796,34	0,00 ± 0,00	80,58 ± 17,30	
	SEED	64,65 ± 0,07	1369,87 ± 1453,65	9801,42 ± 1622,21	0,01 ± 0,02	202,00 ± 51,41	
	SEQDRIFT2	64,71 ± 0,06	200,00 ± 0,00	32447,70 ± 2812,20	0,00 ± 0,00	162,24 ± 14,06	
	STEPD	64,85 ± 0,08	47,92 ± 33,19	1592,20 ± 50,43	0,00 ± 0,00	79,29 ± 12,53	
	WSTD	64,92 ± 0,08	55,66 ± 49,81	7302,09 ± 1122,58	0,00 ± 0,00	336,72 ± 66,64	
		PEDD	63,95 ± 0,09	3405,56 ± 1494,64	30249,19 ± 934,93	0,02 ± 0,03	16,54 ± 3,33
		RQADD	63,80 ± 0,23	1957,33 ± 503,64	19058,86 ± 4006,83	0,00 ± 0,00	10,52 ± 2,27
	SEDD	64,68 ± 0,06	134,57 ± 115,40	20467,40 ± 1625,51	0,00 ± 0,00	257,47 ± 29,37	
	Majority Class	32,98 ± 0,02					
	Naive Bayes	43,13 ± 0,04					
	Temporally Augmented Classifier	60,93 ± 0,24					
Gradual	ADWIN	64,10 ± 0,06	2510,13 ± 881,26	34518,69 ± 2343,17	0,00 ± 0,00	18,17 ± 2,91	
	CUSUM	64,25 ± 0,06	20655,57 ± 2877,10	26241,46 ± 2767,42	0,04 ± 0,04	1,32 ± 0,18	
	DDM	62,06 ± 1,11	18619,44 ± 4382,06	42881,59 ± 2628,65	0,21 ± 0,07	14,68 ± 13,11	
	ECDD	60,38 ± 0,13	26730,67 ± 1103,68	20396,11 ± 589,85	0,48 ± 0,02	0,40 ± 0,02	
	EDDM	62,32 ± 0,37	8052,93 ± 1741,62	3398,42 ± 1666,84	0,04 ± 0,04	0,32 ± 0,12	
	FHDDM	64,49 ± 0,06	1547,14 ± 237,14	2419,52 ± 90,82	0,00 ± 0,00	1,83 ± 0,29	
	FTDD	64,14 ± 0,07	18405,58 ± 3527,34	37940,82 ± 1849,71	0,08 ± 0,04	5,89 ± 5,19	
	HDDMA	64,24 ± 0,06	8623,37 ± 2517,70	20211,14 ± 2657,88	0,00 ± 0,00	4,17 ± 1,33	
	MDDMA	64,44 ± 0,07	2073,58 ± 387,77	2615,33 ± 107,97	0,00 ± 0,00	1,73 ± 0,43	
	PHT	64,14 ± 0,05	30671,97 ± 4054,92	50086,25 ± 898,71	0,31 ± 0,05	1,32 ± 0,25	
	RDDM	64,22 ± 0,07	21575,38 ± 2857,12	18703,80 ± 2329,84	0,05 ± 0,04	0,97 ± 0,23	
	SEED	64,08 ± 0,07	1981,07 ± 1172,78	9220,91 ± 1154,28	0,00 ± 0,00	8,20 ± 1,63	
	SEQDRIFT2	64,14 ± 0,09	18531,67 ± 2747,65	34559,38 ± 2256,01	0,03 ± 0,03	2,03 ± 0,27	
	STEPD	64,24 ± 0,08	1369,26 ± 320,75	1638,27 ± 54,56	0,00 ± 0,00	1,64 ± 0,39	
	WSTD	63,95 ± 0,16	5082,04 ± 1089,18	8192,88 ± 1064,56	0,00 ± 0,00	2,25 ± 0,71	
		PEDD	63,34 ± 0,11	21698,50 ± 2490,32	33892,78 ± 1187,98	0,22 ± 0,03	1,38 ± 0,25
		RQADD	63,65 ± 0,14	13259,77 ± 2768,83	21886,04 ± 4376,95	0,11 ± 0,05	2,36 ± 1,32
	SEDD	64,00 ± 0,07	17473,61 ± 2164,69	22798,49 ± 1428,38	0,12 ± 0,05	1,54 ± 0,70	

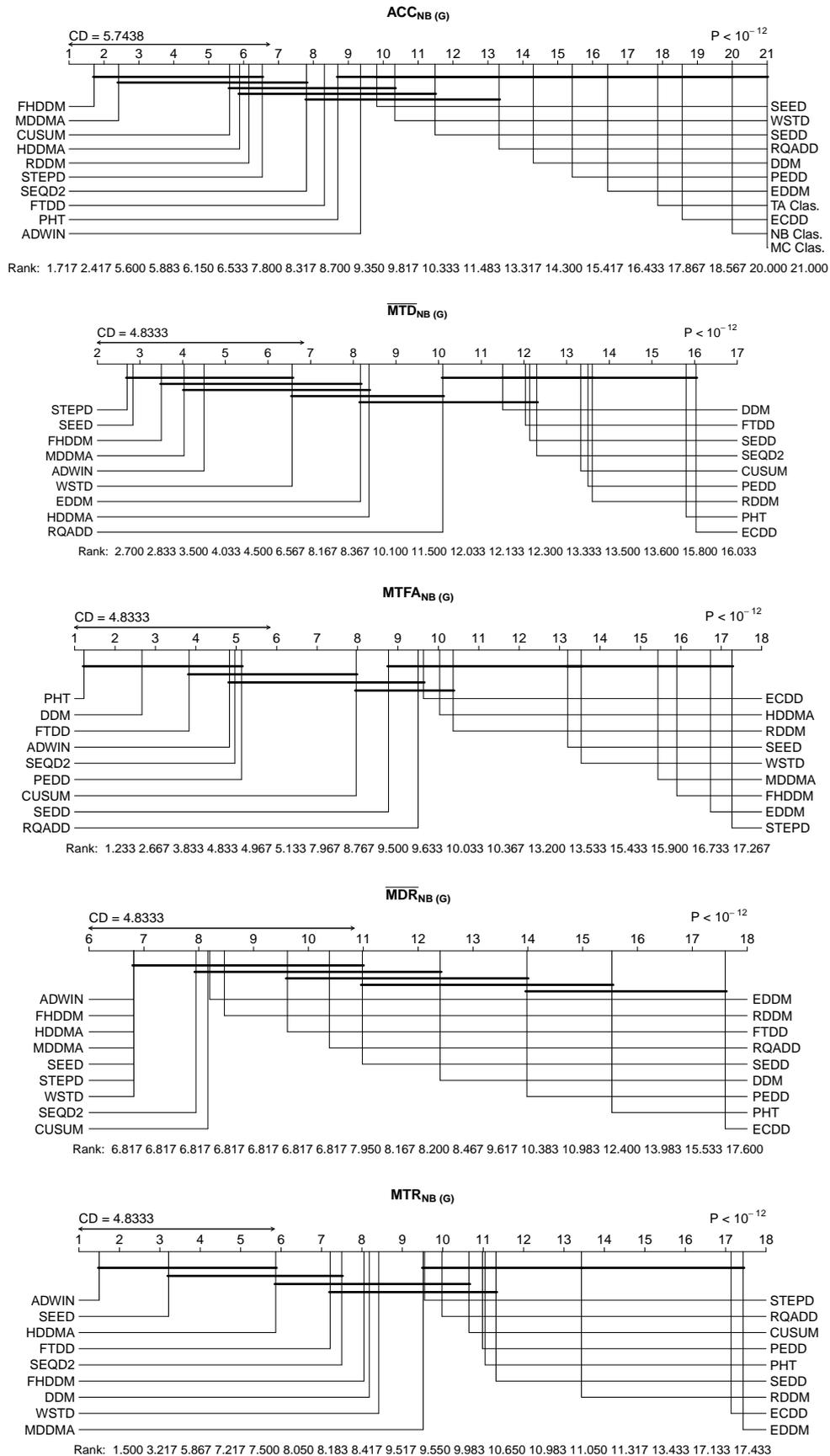
Fonte: Elaborada pelo autor (2021).

Figura 56 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados abruptos (A) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

Figura 57 – Comparação dos classificadores e detectores com NB usando o teste post-hoc de Nemenyi em conjuntos de dados graduais (G) da base RBF com cinco classes, considerando um intervalo de confiança de 95%.



Fonte: Elaborada pelo autor (2021).

APÊNDICE B – RESULTADOS EXPERIMENTAIS COM BASES REAIS

Tabela 32 – Resultados com as acurácias (%) obtidas usando HT, considerando as 20 bases reais.

	ADWIN	CUSUM	DDM	ECDD	EDDM	FHDDM	FTDD	HDDMA	HT	MDDMA	MC Class,	PHT	RDDM	SEED	SEQD2	STEPD	TA Class,	WSTD	PEDD	RQADD	SEDD
Bank Marketing	72.30	68.90	68.40	65.20	70.10	71.40	69.50	69.00	58.40	72.00	51.10	71.00	70.10	69.50	72.30	67.50	61.60	71.70	72.40	71.90	70.80
Census-Income (KDD)	94.13	95.03	95.63	95.13	95.10	94.97	94.90	95.53	94.27	94.37	93.90	93.90	94.70	94.40	93.87	95.13	97.90	94.77	94.10	94.27	93.97
Chess (King-Rook vs. King)	98.60	94.70	99.90	99.90	97.40	98.50	99.00	99.40	22.30	99.80	0.00	92.90	99.90	98.60	88.00	99.90	99.90	98.90	97.10	79.40	96.60
Chess (King-Rook vs. King) (Emb.)	34.80	35.10	35.10	35.10	35.10	33.80	35.10	35.10	35.10	35.50	14.90	35.10	35.10	17.00	35.10	16.60	35.50	34.90	35.20	17.40	35.10
Connect-4 Data Set	65.60	66.00	68.00	66.30	64.70	65.80	65.10	64.80	61.50	64.30	53.40	65.10	65.80	64.50	66.10	66.00	58.80	63.40	65.90	66.10	65.80
Forest Covertypes (Norm.)	83.92	83.87	88.58	90.48	87.53	86.28	86.92	87.22	82.25	85.15	38.90	82.87	88.48	85.02	81.60	87.82	96.28	86.95	81.15	77.15	82.78
Forest Covertypes (Norm. e Emb.)	75.02	76.07	80.43	74.53	79.98	74.88	74.85	76.45	78.70	75.40	46.80	78.87	75.30	74.88	75.92	75.03	76.98	75.92	75.87	75.22	76.35
EEG Eye State	97.00	92.80	99.40	99.40	96.10	97.10	97.60	98.20	79.00	94.70	90.70	90.70	97.40	97.10	93.80	98.20	99.70	96.30	90.70	90.60	92.90
Electricity (Norm.)	78.80	76.00	86.20	88.80	84.30	85.00	85.70	89.90	81.60	82.00	53.30	86.30	87.20	77.60	76.90	86.80	90.60	85.50	76.20	79.30	77.10
Electricity (Norm. e Emb.)	75.10	75.10	75.10	70.50	77.10	73.60	75.10	76.00	75.10	75.40	59.30	75.10	75.10	71.50	75.10	75.70	75.70	76.00	75.70	75.00	75.10
Give me Some Credit	92.90	92.90	92.60	91.95	92.65	92.90	92.90	92.90	92.90	92.90	92.65	92.90	92.40	92.35	92.90	92.55	92.95	92.90	92.55	92.90	92.90
Letter Recognition	39.00	60.20	60.20	47.80	60.20	58.90	60.60	60.20	60.20	58.20	4.10	60.20	60.20	28.30	37.80	23.70	60.00	57.10	60.20	60.20	60.20
Nomao	81.60	81.00	82.30	82.20	80.50	80.40	81.80	82.10	85.60	82.00	30.60	83.40	81.10	81.60	80.40	83.10	84.30	81.80	81.60	80.90	84.40
Outdoor Objects	57.70	57.90	60.40	54.60	53.90	58.40	58.30	56.80	57.80	55.30	0.00	57.80	54.30	55.60	53.50	54.80	90.00	55.00	57.80	57.80	53.30
Poker-Hand (Norm.)	75.78	74.65	70.24	78.93	77.68	77.64	77.00	77.00	80.54	76.85	57.93	69.15	76.96	76.59	73.79	77.95	90.64	77.14	72.44	62.81	75.63
Santander Customer Satisfaction	95.20	95.20	95.60	94.60	95.70	95.20	95.20	95.20	95.20	95.20	95.70	95.20	95.50	95.70	95.50	95.70	95.20	95.20	95.60	95.60	95.20
Sick	94.00	94.00	93.50	93.00	93.30	94.00	94.00	94.00	94.00	94.00	93.40	94.00	94.00	94.00	94.00	94.00	94.10	94.00	94.00	94.00	94.00
Spam Data	96.90	96.10	97.70	98.00	96.30	97.20	95.00	95.40	96.70	97.20	95.20	96.10	97.80	96.70	95.50	97.50	97.50	96.50	95.10	97.20	97.30
Usenet 1 e 2 (Unidos)	69.90	69.20	68.60	67.10	70.60	69.60	70.00	68.90	67.70	69.50	66.60	67.70	68.60	69.40	70.00	68.70	83.20	69.60	70.10	68.70	70.20
Weather	66.50	67.20	65.30	70.10	69.70	67.60	68.40	65.90	65.60	67.30	63.40	65.60	66.40	65.80	66.00	64.50	67.80	67.20	62.60	62.90	65.50

Fonte: Elaborada pelo autor (2021).

Tabela 33 – Resultados com as acurácias (%) obtidas usando NB, considerando as 20 bases reais.

	ADWIN	CUSUM	DDM	ECDD	EDDM	FHDDM	FTDD	HDDMA	MDDMA	MC Class.	NB	PHT	RDDM	SEED	SEQD2	STEPD	TA Class.	WSTD	PEDD	RQADD	SEDD
Bank Marketing	70.50	71.10	71.30	67.00	71.40	71.50	71.50	70.60	72.10	51.10	63.10	71.90	71.40	70.20	71.10	70.70	61.60	71.70	70.40	70.60	70.50
Census-Income (KDD)	85.13	87.23	90.60	94.00	90.10	89.17	93.87	92.23	85.80	93.90	82.53	94.43	90.13	89.23	79.87	92.83	97.90	91.70	81.67	94.27	92.80
Chess (King-Rook vs. King)	99.70	94.80	99.90	99.90	99.70	96.40	98.90	99.50	99.70	0.00	5.00	93.90	99.90	97.50	88.00	99.90	99.90	98.50	89.40	85.90	96.50
Chess (King-Rook vs. King) (Emb.)	34.60	35.10	35.10	35.10	35.30	33.80	35.10	35.10	35.50	14.90	35.10	35.10	35.10	18.20	35.40	14.10	35.50	34.90	34.90	35.60	35.10
Connect-4 Data Set	66.00	66.00	66.10	66.60	64.20	65.20	65.40	64.90	64.80	53.40	48.10	66.90	66.10	65.10	66.10	66.00	58.80	65.80	65.20	66.10	65.80
Forest Covertype (Norm.)	83.95	83.65	89.65	91.05	88.05	86.53	87.20	87.33	85.52	38.90	63.30	82.58	88.68	85.17	81.77	88.38	96.28	87.28	79.42	63.65	83.13
Forest Covertype (Norm. e Emb.)	70.45	72.33	68.15	71.37	64.85	70.83	68.68	71.87	71.05	46.80	64.38	71.20	68.13	70.42	70.50	70.58	78.98	72.63	71.70	69.18	71.43
EEG Eye State	97.00	93.90	99.10	99.20	96.10	97.10	97.60	98.10	96.60	90.70	9.30	91.50	97.50	96.80	93.80	98.20	99.70	96.60	92.30	9.30	94.10
Electricity (Norm.)	78.80	76.20	86.60	89.90	86.20	85.00	85.90	90.20	82.00	53.30	75.30	76.10	87.20	76.90	77.00	87.30	90.60	85.60	76.70	77.00	78.20
Electricity (Norm. e Emb.)	70.70	75.00	75.60	70.00	75.80	71.70	74.00	74.20	75.10	59.30	74.00	74.00	75.80	75.50	74.00	73.30	75.70	76.00	74.00	72.80	74.10
Give me Some Credit	92.55	92.55	92.55	78.65	92.60	92.55	92.55	92.55	92.55	92.65	92.55	92.55	91.70	90.50	92.55	90.95	92.95	92.55	90.70	92.55	92.55
Letter Recognition	39.10	60.20	60.20	47.90	60.20	58.90	60.80	60.20	58.20	4.10	60.20	60.20	60.20	30.40	37.80	23.60	60.00	57.00	60.20	60.20	60.20
Nomao	82.10	79.40	82.60	75.20	78.80	77.70	77.20	76.00	77.60	30.60	55.40	84.30	79.40	81.60	76.00	80.60	84.30	77.20	79.30	81.10	79.30
Outdoor Objects	57.70	57.90	60.40	54.60	56.80	58.90	57.70	57.40	55.50	0.00	57.80	57.80	55.30	55.70	53.60	55.50	90.00	55.30	57.80	57.80	53.40
Poker-Hand (Norm.)	76.34	73.53	64.45	79.53	77.29	77.29	76.94	77.85	76.39	57.93	59.74	66.86	76.64	76.45	73.31	77.74	90.64	76.76	71.38	63.05	73.88
Santander Customer Satisfaction	28.20	61.50	95.70	95.70	77.60	75.40	25.30	95.70	75.30	95.70	95.70	63.50	70.70	75.00	64.70	20.20	95.20	20.10	20.20	53.90	95.70
Sick	93.30	93.30	93.30	92.40	93.40	93.30	93.30	93.40	93.30	93.40	93.30	93.30	93.30	94.30	93.30	93.30	94.10	93.30	92.60	93.30	93.30
Spam Data	96.30	96.00	96.00	98.00	95.40	95.20	95.50	87.70	95.60	95.20	96.60	96.00	97.80	96.30	95.50	97.70	97.50	95.30	95.40	95.60	96.00
Usenet 1 e 2 (Unidos)	71.50	69.20	71.40	74.50	71.90	71.90	70.80	72.40	70.00	66.60	67.70	67.70	74.50	70.80	72.10	72.90	83.20	71.00	67.70	69.20	71.80
Weather	67.00	61.70	59.80	70.50	70.70	67.60	66.10	65.30	64.90	63.40	59.20	61.60	66.10	66.10	61.10	69.00	67.80	66.70	59.40	61.20	65.20

Fonte: Elaborada pelo autor (2021).

Tabela 34 – Resultados com as medidas ARL obtidas usando HT, considerando as 20 bases reais.

	ADWIN	CUSUM	DDM	ECDD	EDDM	FHDDM	FTDD	HDDMA	MDDMA	PHT	RDDM	SEED	SEQD2	STEPD	WSTD	PEDD	RQADD	SEDD
Bank Marketing	2192,00	3975,90	1106,22	180,18	386,72	3010,36	3331,33	5010,50	3504,58	6950,83	2232,56	1547,08	2475,00	619,47	2637,38	1643,33	1395,75	4937,75
Census-Income (KDD)	593,21	3125,22	329,87	112,00	291,95	424,70	309,16	221,26	608,19	16943,75	298,30	344,37	698,60	75,42	391,76	1423,32	149642,00	1563,17
Chess (King-Rook vs. King)	1153,33	1736,63	813,74	813,74	1265,23	1164,92	998,32	1263,05	1444,06	1743,63	864,59	1153,33	1544,44	47,29	1164,75	1271,91	1632,43	1389,75
Chess (King-Rook vs. King) (Emb.)	705,88	14028,00	14028,00	14028,00	14028,00	1740,79	14028,00	14028,00	2789,25	14028,00	14028,00	457,60	680,00	268,89	3952,00	617,57	536,15	14028,00
Connect-4 Data Set	510,55	1037,61	533,63	156,90	133,96	382,66	429,51	401,00	420,96	3239,80	354,89	308,99	465,22	195,68	358,32	1159,11	893,42	624,31
Forest Covertype (Norm.)	344,13	1267,69	66,76	44,74	117,61	161,52	117,22	78,41	235,58	3175,77	108,47	188,23	690,00	75,52	133,59	2702,99	1496,00	699,43
Forest Covertype (Norm. e Emb.)	2240,50	5004,85	14518,26	186,82	487,79	548,53	3769,64	1217,38	584,54	27238,70	966,11	804,62	1654,86	325,82	758,40	3222,31	161,27	1695,28
EEG Eye State	456,50	588,64	349,50	334,72	456,84	456,47	406,67	349,83	477,38	647,10	382,58	442,44	489,50	67,93	448,59	649,20	886,00	518,46
Electricity (Norm.)	511,64	1000,41	133,31	52,63	118,35	250,02	214,65	107,29	346,25	2692,08	157,34	261,77	586,84	127,31	206,82	1506,43	1099,42	933,19
Electricity (Norm. e Emb.)	22656,00	22656,00	22656,00	260,13	199,33	3634,70	22656,00	13931,50	2828,94	22656,00	22656,00	3688,00	22656,00	1925,50	13933,00	10700,00	8590,25	22656,00
Give me Some Credit	75000,00	75000,00	5087,60	226,50	392,60	75000,00	75000,00	75000,00	5548,12	75000,00	5548,12	48080,00	75000,00	1347,23	75000,00	2401,56	75000,00	75000,00
Letter Recognition	181,04	10000,00	10000,00	1068,78	10000,00	1499,50	171,67	10000,00	2001,50	10000,00	10000,00	110,58	200,00	67,99	484,84	10000,00	10000,00	10000,00
Nomao	670,08	1403,29	441,42	202,29	390,31	854,80	589,59	518,00	854,75	3157,50	438,19	465,33	633,33	183,26	683,88	1006,18	1421,46	1197,42
Outdoor Objects	458,67	685,00	124,50	108,11	57,09	107,61	237,50	101,63	247,38	2000,00	151,77	293,33	800,00	109,64	141,04	2000,00	2000,00	520,83
Poker-Hand (Norm.)	497,98	671,54	396,11	61,43	86,23	220,85	272,40	161,52	294,05	1022,14	164,94	324,96	522,57	132,96	226,40	1248,25	2021,71	598,74
Santander Customer Satisfaction	38010,00	38010,00	3795,15	198,47	1302,12	38010,00	38010,00	38010,00	2653,38	38010,00	2653,38	6432,00	6500,00	1353,17	38010,00	8265,75	8556,25	38010,00
Sick	1886,00	1886,00	25,00	238,64	427,50	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00	1886,00
Spam Data	660,57	274,75	115,40	69,94	242,95	605,67	662,43	461,80	261,75	275,75	256,47	418,91	1033,33	105,22	534,90	923,00	369,80	269,75
Usenet 1 e 2 (Unidos)	192,00	222,50	129,86	150,75	103,13	182,40	152,42	181,50	184,20	1500,00	151,75	178,29	250,00	151,75	152,17	514,00	475,00	155,75
Weather	820,36	1766,20	1287,86	134,07	86,17	226,14	491,08	290,15	237,11	9079,50	204,88	290,58	556,25	153,85	339,98	1543,50	772,64	322,34

Fonte: Elaborada pelo autor (2021).

Tabela 35 – Resultados com as medidas ARL obtidas usando NB, considerando as 20 bases reais.

	ADWIN	CUSUM	DDM	ECDD	EDDM	FHDDM	FTDD	HDDMA	MDDMA	PHT	RDDM	SEED	SEQD2	STEPPD	WSTDD	PEDD	RQADD	SEDD
Bank Marketing	1518,77	1979,80	1988,60	190,87	476,58	1515,36	1657,71	2195,22	2548,06	5072,13	1326,20	827,26	2200,00	523,46	1622,54	1645,29	823,58	2820,43
Census-Income (KDD)	404,37	1644,15	273,55	104,49	189,62	341,27	212,20	207,48	471,85	22304,63	307,89	258,18	520,91	47,49	255,29	1186,00	27462,83	1092,13
Chess (King-Rook vs. King)	1032,62	1155,25	813,74	813,74	1033,27	926,83	864,84	822,00	1033,04	1166,83	864,59	932,27	1158,33	47,29	911,10	1157,50	1269,00	1076,08
Chess (King-Rook vs. King) (Emb.)	710,59	14028,00	14028,00	14028,00	33,50	1740,79	14028,00	2789,25	14028,00	14028,00	14028,00	377,95	200,00	214,22	3952,00	770,54	177,75	14028,00
Connect-4 Data Set	535,62	1037,73	526,09	152,64	133,96	354,29	453,66	426,38	415,76	2634,13	351,20	311,85	452,11	184,93	366,11	1149,63	845,16	602,02
Forest Covertype (Norm.)	330,53	1013,53	62,52	45,10	119,92	156,37	115,11	88,25	219,78	2475,89	106,00	184,06	645,43	77,65	132,07	3108,99	926,56	655,80
Forest Covertype (Norm. e Emb.)	1637,97	3393,60	7113,36	154,43	215,57	363,83	1791,74	739,79	462,37	13727,83	1083,07	603,21	1695,32	258,90	535,26	5368,40	161,40	1449,47
EEG Eye State	486,30	606,50	349,50	321,17	486,67	486,27	406,67	366,93	456,41	631,29	382,58	449,03	459,50	69,33	448,59	632,67	7499,50	521,14
Electricity (Norm.)	416,89	786,00	157,55	53,12	111,03	206,44	186,26	106,78	261,70	1703,15	137,19	229,71	557,50	116,76	177,51	1327,82	1066,25	859,83
Electricity (Norm. e Emb.)	21280,00	3402,83	2508,50	260,11	138,67	2514,61	22656,00	19486,50	6060,50	22656,00	919,67	8624,00	22656,00	1599,79	5720,00	22656,00	10987,00	16824,00
Give me Some Credit	480,00	75000,00	75000,00	343,90	172,70	75000,00	75000,00	75000,00	75000,00	75000,00	1966,00	8520,00	500,00	407,11	75000,00	1132,91	75000,00	75000,00
Letter Recognition	184,45	10000,00	10000,00	1088,78	10000,00	1499,50	113,00	10000,00	2001,50	10000,00	10000,00	110,93	200,00	80,06	541,82	10000,00	10000,00	10000,00
Nomao	625,78	1143,97	371,62	172,63	311,40	715,38	474,94	520,62	899,74	2385,79	328,71	372,27	657,69	224,26	505,38	777,50	947,64	950,25
Outdoor Objects	458,67	684,50	47,33	83,91	43,43	107,61	246,88	107,19	247,38	2000,00	151,73	293,33	800,00	131,57	151,85	2000,00	2000,00	520,83
Poker-Hand (Norm.)	461,38	628,74	166,54	59,74	85,20	214,89	245,88	158,43	278,07	845,91	160,65	303,53	501,09	128,03	216,01	1307,30	1801,96	550,24
Santander Customer Satisfaction	173,32	155,98	94,30	45,53	73,35	95,31	833,45	181,87	122,73	162,57	103,03	105,81	425,84	92,96	1512,89	3771,17	170,73	197,38
Sick	1886,00	1886,00	25,00	117,29	86,00	1886,00	1886,00	26,50	1886,00	1886,00	1886,00	616,00	1886,00	1886,00	1886,00	1763,50	1886,00	1886,00
Spam Data	921,60	535,00	137,13	109,90	210,93	1534,83	1151,00	513,39	881,00	537,00	242,97	460,80	775,00	100,64	767,17	1538,33	308,17	535,50
Usenet 1 e 2 (Unidos)	192,00	220,00	124,40	142,22	116,90	176,83	152,42	175,79	177,00	1500,00	152,31	174,00	300,00	134,83	155,36	1500,00	747,50	155,75
Weather	601,60	827,95	178,43	134,07	90,01	226,14	401,82	287,75	243,61	1804,70	215,27	250,22	725,00	160,54	294,63	4535,25	621,25	291,15

Fonte: Elaborada pelo autor (2021).