



Pós-Graduação em Ciência da Computação

Danilo Rafael de Lima Cabral

TESTES ESTATÍSTICOS E DETECÇÕES DE MUDANÇAS DE CONCEITOS EM FLUXOS DE DADOS



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
<www.cin.ufpe.br/~posgraduacao>

Recife
2017

Danilo Rafael de Lima Cabral

**TESTES ESTATÍSTICOS E DETECÇÕES DE MUDANÇAS
DE CONCEITOS EM FLUXOS DE DADOS**

Trabalho apresentado ao Programa de Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: *Prof. Roberto Souto Maior de Barros*

Recife
2017

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

C117t Cabral, Danilo Rafael de Lima
Testes estatísticos e detecções de mudanças de conceitos em fluxos de dados / Danilo Rafael de Lima Cabral . – 2017.
107 f.: il., fig., tab.

Orientador: Roberto Souto Maior de Barros.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2017.
Inclui referências e apêndices.

1. Ciência da computação. 2. Aprendizagem de máquina. I. Barros, Roberto Souto Maior (orientador). II. Título.

004 CDD (23. ed.) UFPE- MEI 2017-93

Danilo Rafael de Lima Cabral

Testes Estatísticos e Detecções de Mudanças de Conceitos em Fluxos de Dados

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação

Aprovado em: 03/02/2017.

BANCA EXAMINADORA

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

Prof. Dr. Marco Antonio de Oliveira Domingues
Departamento de Análise e Desenvolvimento de Sistemas / IFPE

Prof. Dr. Roberto Souto Maior de Barros
Centro de Informática / UFPE
(Orientador)

Este trabalho é dedicado a todos aqueles que, a despeito das evidências, acreditaram que um tecnólogo se tornaria um cientista e mestre.

AGRADECIMENTOS

Agradeço ao meu orientador, o professor Roberto Souto Maior de Barros, pelo apoio, ensino e dedicação, desde o momento em que me recebeu como o seu orientando. Ele, com a sua ampla experiência como educador e o seu vasto conhecimento em minha área de pesquisa, foi presente nos momentos essenciais com a sua orientação. Sem a sua ajuda, eu não teria chegado até aqui.

Agradeço a todos os professores/funcionários do Centro de Informática (CIn-UFPE) que participaram direta ou indiretamente da minha formação.

Por fim, e não menos importante, agradeço a todos os meus parentes e amigos. Em especial a minha esposa Jaqueline Cabral, que sempre foi compreensiva quanto a minha necessidade de dedicação a esta pós-graduação; ao meu irmão George Cabral, pelo seu belo exemplo como mestre e pesquisador; a minha mãe Gilvaneide Cabral, por ter me ensinado os princípios básicos da vida; e ao meu pai João Cabral, por ter me ensinado a importância da sabedoria e do conhecimento.

*“Mais vale a sabedoria do que a força.”
(Bíblia Sagrada, Eclesiastes 9:16)*

RESUMO

O aprendizado online objetiva a extração de informações a partir de fluxos de dados compostos de uma grande quantidade de exemplos. Esses fluxos frequentemente contêm mudanças de conceitos que na maioria dos casos são caracterizadas como modificações nas distribuições dos dados. Métodos detectores de mudanças de conceitos são algoritmos projetados para trabalharem conjuntamente com um – ou mais – classificador base, a fim de estimarem as posições das mudanças e quando necessário substituírem o preditor, com o objetivo de melhorar a sua acurácia. DDM, EDDM e STEPDP são exemplos de detectores simples, eficientes e bem-conceituados. Apesar de sua eficácia em bases pequenas, o DDM tende a perder precisão quando apresentado a conjuntos de dados consideravelmente grandes. Por outro lado, o EDDM funciona bem com bases de dados graduais, porém alcança baixos índices de acurácia em bases com mudanças de conceitos abruptas. O STEPDP, por sua vez, foi projetado para a detecção de mudanças de conceitos através do teste de hipóteses entre duas proporções, entretanto, estatisticamente, esse teste não é recomendado para amostras pequenas e/ou desbalanceadas. Este trabalho propõe quatro novos detectores (formando o total de sete versões) que objetivam melhorar o DDM, EDDM e STEPDP. Todos os métodos propostos são baseados em testes estatísticos, sendo o EMZD baseado no teste de hipóteses entre médias de duas amostras independentes e, o FPDD, FSDD e FTDD baseados no teste exato de Fisher. Os experimentos realizados, com dois classificadores base, usando 36 conjuntos de dados artificiais e três bases de dados reais, demonstraram a eficácia e eficiência dos detectores propostos. No que diz respeito a avaliação dos detectores, uma das versões do EMZD obteve as melhores acurácias e o FPDD foi o mais preciso na análise das detecções de mudanças de conceitos.

Palavras-chave: Estatística. Aprendizagem de Máquina. Mudanças de Conceitos em Fluxos de Dados.

ABSTRACT

Online learning aims to extract information from data streams composed of a large number of examples. These flows often contain concept drifts that in most cases are characterized as changes in data distributions. Concept drifts detectors are algorithms designed to work with one or more base classifier in order to estimate the change positions and, when necessary, replace the predictor to improve its accuracy. DDM, EDDM and STEPDP are simple, efficient and well-known detectors. Despite its effectiveness on small bases, DDM tends to lose accuracy when faced with considerably large data sets. On the other hand, EDDM works well with gradual databases, but achieves low accuracy on bases with abrupt drifts. STEPDP was designed to detect changes in distribution using a hypothesis test between two proportions, however, statistically this test is not recommended for small and/or imbalanced samples. This work proposes four new detectors (seven versions in total) that aim to improve DDM, EDDM and STEPDP. All the proposed methods are inspired by statistical tests, where EMZD is based on hypothesis test between means of two independent samples and FPDD, FSDD and FTDD are based on Fisher's exact test. Experiments with two base classifiers using 36 artificial data sets and three real-world datasets demonstrated the effectiveness and efficiency of the proposed detectors. Regarding the evaluation of detectors, one of the versions of the EMZD obtained the best accuracy and the FPDD was the most accurate in the analysis of the concept drifts detections.

Keywords: Statistic. Machine Learning. Concept Drifts in Data Streams.

LISTA DE ILUSTRAÇÕES

Figura 1 – Comparação das Acurácias dos Detectores, Utilizando o NB, através do Teste de <i>Friedman</i> e o Pós-Teste <i>Nemenyi</i> , com 95% de intervalo de confiança.	57
Figura 2 – Comparação das Acurácias dos Detectores, Utilizando o HT, através do Teste de <i>Friedman</i> e o Pós-Teste <i>Nemenyi</i> , com 95% de intervalo de confiança.	57
Figura 3 – Comparação do MCC dos Detectores, Utilizando o NB, através do Teste de <i>Friedman</i> e o Pós-Teste <i>Nemenyi</i> , com 95% de intervalo de confiança.	64
Figura 4 – Comparação do MCC dos Detectores, Utilizando o HT, através do Teste de <i>Friedman</i> e o Pós-Teste <i>Nemenyi</i> , com 95% de intervalo de confiança.	64
Figura 5 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando NB - Parte 1. . . .	68
Figura 6 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando NB - Parte 2. . . .	69
Figura 7 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando HT - Parte 1. . . .	70
Figura 8 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando HT - Parte 2. . . .	71

LISTA DE TABELAS

Tabela 1 – Tabela de Contigência para o Teste Exato de Fisher	23
Tabela 2 – Tabela de Contingência Generalizada para o Teste Exato de Fisher . .	23
Tabela 3 – Tabela de Contingência Generalizada para o Teste do Qui-Quadrado de Homogeneidade das Proporções	24
Tabela 4 – Tabela com as Principais Características dos Detectores	33
Tabela 5 – Médias – em Porcentagem (%) – das Acurácias dos Detectores em 40 Repetições, Utilizando NB nos Conjuntos de Dados Artificiais, com 95% de Intervalo de Confiança.	55
Tabela 6 – Médias – em Porcentagem (%) – das Acurácias dos Detectores em 40 Repetições, Utilizando HT nos Conjuntos de Dados Artificiais, com 95% de Intervalo de Confiança.	56
Tabela 7 – Acurácias – em Porcentagem (%) – dos Detectores, Utilizando NB nos Conjuntos de Dados Reais.	58
Tabela 8 – Acurácias – em Porcentagem (%) – dos Detectores, Utilizando HT nos Conjuntos de Dados Reais.	58
Tabela 9 – Análise das Detecções de Mudanças de Conceitos Abruptas, Em 40 Repetições, Utilizando NB nos Conjuntos de Dados Artificiais.	62
Tabela 10 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT nos Conjuntos de Dados Artificiais.	63
Tabela 11 – Médias das Distâncias entre as Mudanças de Conceitos Abruptas e as Detecções dos Métodos, Utilizando NB nos Conjuntos de Dados Artificiais.	85
Tabela 12 – Médias das Distâncias entre as Mudanças de Conceitos Abruptas e as Detecções dos Métodos, Utilizando HT nos Conjuntos de Dados Artificiais.	86
Tabela 13 – Taxas de Falsos Positivos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	87
Tabela 14 – Taxas de Falsos Positivos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	88
Tabela 15 – Taxas de Falsos Negativos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	89
Tabela 16 – Taxas de Falsos Negativos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	90
Tabela 17 – Taxas de Verdadeiros Positivos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	91

Tabela 18 – Taxas de Verdadeiros Positivos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	92
Tabela 19 – Taxas de Verdadeiros Negativos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	93
Tabela 20 – Taxas de Verdadeiros Negativos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.	94
Tabela 21 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados <i>Agrawal</i> .	96
Tabela 22 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados <i>Agrawal</i> .	97
Tabela 23 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados <i>LED</i> .	98
Tabela 24 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados <i>LED</i> .	99
Tabela 25 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados <i>Mixed</i> .	100
Tabela 26 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados <i>Mixed</i> .	101
Tabela 27 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados <i>Random RBF</i> .	102
Tabela 28 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados <i>Random RBF</i> .	103
Tabela 29 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados <i>SEA</i> .	104
Tabela 30 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados <i>SEA</i> .	105
Tabela 31 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados <i>Sine</i> .	106
Tabela 32 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados <i>Sine</i> .	107

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivos	16
1.2	Metodologia	17
1.3	Organização	18
2	ESTATÍSTICA	19
2.1	Estatística Paramétrica e Não Paramétrica	19
2.2	Princípios Básicos dos Testes de Hipóteses	19
2.3	Testes Estatísticos Paramétricos	20
2.3.1	Teste de Hipóteses entre Proporções de Duas Amostras Independentes	20
2.3.2	Teste de Hipóteses entre Médias de Duas Amostras Independentes	21
2.4	Testes Estatísticos Não Paramétricos	22
2.4.1	Teste Exato de Fisher	22
2.4.2	Teste de Qui-Quadrado de Homogeneidade das Proporções	24
2.5	Considerações Finais	25
3	APRENDIZAGEM DE MÁQUINA E FLUXOS DE DADOS	26
3.1	Aprendizagem de Máquina	26
3.2	Aprendizagem Supervisionada	26
3.3	Aprendizagem Não Supervisionada	27
3.4	Fluxos de Dados	27
3.5	Mudanças de Conceitos em Fluxos de Dados	28
3.6	Detectores de Mudanças de Conceitos	29
3.6.1	DDM	30
3.6.2	EDDM	30
3.6.3	STEPD	31
3.6.4	ADWIN	32
3.6.5	ECDD	32
3.6.6	Resumo dos Detectores de Mudanças de Conceitos	33
3.7	Algoritmos de Aprendizagem Supervisionada em Fluxos de Dados	34
3.7.1	Naive Bayes	34
3.7.2	Hoeffding Tree	34
3.8	Considerações Finais	34

4	MÉTODOS BASEADOS NO TESTE DE HIPÓTESES ENTRE DUAS MÉDIAS	35
4.1	Descrição Geral dos Métodos Baseados no Teste de Hipóteses entre Duas Médias	35
4.2	EMZD1	36
4.3	EMZD2	38
4.4	Considerações Finais	39
5	MÉTODOS BASEADOS NO TESTE EXATO DE FISHER	40
5.1	Implementação do Teste Exato de Fisher	40
5.2	Descrição Geral dos Métodos Baseados no Teste Exato de Fisher	41
5.3	FPDD1 e FPDD2	44
5.4	FSDD1 e FSDD2	45
5.5	FTDD	47
5.6	Considerações Finais	48
6	EXPERIMENTOS	49
6.1	Configuração dos Experimentos	49
6.2	Bases Artificiais	50
6.2.1	Agrawal	51
6.2.2	LED	51
6.2.3	Mixed	52
6.2.4	Random RBF	52
6.2.5	SEA	52
6.2.6	Sine	53
6.3	Bases Reais	53
6.3.1	Covertime	53
6.3.2	Pokerhand	53
6.3.3	Sensor	54
6.4	Comparação de Acurácias	54
6.5	Análise das Detecções de Mudanças de Conceitos	60
6.6	Gráficos de Acurácia x MCC	66
7	CONCLUSÕES	72
7.1	Publicações	74
7.2	Limitações	74
7.3	Trabalhos Futuros	75
	ACRÔNIMOS	76

REFERÊNCIAS	78
APÊNDICES	83
APÊNDICE A – TABELAS DETALHADAS, POR MÉTRICAS, COM AS ANÁLISES DAS DETECÇÕES DE MUDANÇAS DE CONCEITOS	84
APÊNDICE B – TABELAS DETALHADAS, POR BASES DE DADOS, COM AS ANÁLISES DAS DETECÇÕES DE MUDANÇAS DE CONCEITOS	95

1 INTRODUÇÃO

A extração de conhecimento em ambientes com fluxo contínuo de dados é uma atividade que vem crescendo progressivamente. Nesse contexto, existem muitos exemplos de aplicações que utilizam o aprendizado online (ŽLIOBAITĖ; PECHENIZKIY; GAMA, 2016). Dentre eles, pode-se citar: monitoramento de sensores de movimento (LEE; WANG; RYU, 2007), filtros de *spams* (KATAKIS; TSOUMAKAS; VLAHAVAS, 2010), detecção de intrusões (LANE; BRODLEY, 1998), análise de sentimentos na área financeira (SMAILOVIC et al., 2014), dentre outros.

Fluxos de dados são ambientes que, frequentemente, contém uma grande quantidade de dados (possivelmente infinita) fluindo rapidamente e continuamente. Assim, os algoritmos propostos para aprenderem a partir desses fluxos, funcionam, em sua grande maioria, utilizando o protocolo de aprendizado online. Além disso, tais algoritmos precisam lidar com restrições quanto ao uso de memória e tempo de execução, visto que, nesse tipo de ambiente, a leitura da mesma instância mais de uma vez costuma não ser possível. Adicionalmente, nesse cenário, os conceitos das distribuições podem mudar ao longo do tempo, gerando um fenômeno conhecido como mudança de conceito (BRZEZINSKI; STEAFNOWSKI, 2016, p. 1–2).

Existem diferentes maneiras de se categorizar as mudanças de conceitos em fluxos de dados. Uma delas, é de acordo com a velocidade das mudanças, que podem ser: abruptas ou graduais (MINKU; WHITE; YAO, 2010, p. 731). Outra forma de categorização, é dividir as mudanças de conceitos de acordo com as reações que elas provocam nas distribuições dos dados. Nesse contexto, as mudanças podem ser: reais ou virtuais (GONÇALVES JR.; BARROS, 2013, p. 1018).

Diversas abordagens têm sido propostas para lidarem com as mudanças de conceitos em fluxos de dados. Dentre elas, uma das principais estratégias é baseada em métodos detectores de mudanças de conceitos, os quais são algoritmos propostos para monitorarem os resultados das predições de um classificador base, com o objetivo de identificar, e sinalizar, possíveis mudanças nas distribuições dos dados (BRZEZINSKI; STEAFNOWSKI, 2016, p. 7).

Dentre os principais métodos detectores de mudanças de conceitos da atualidade, pode-se citar: *Drift Detection Method* (DDM) (GAMA et al., 2004), *Early Drift Detection Method* (EDDM) (BAENA-GARCIA et al., 2006), *Statistical Test of Equal Proportions* (STEPD) (NISHIDA; YAMAUCHI, 2007), *Adaptative Windowing* (ADWIN) (BIFET; GAVALDÀ, 2007) e *EWMA for Concept Drift Detection* (ECDD) (ROSS et al., 2012).

Uma conhecida dificuldade enfrentada pelo **DDM** reside em amostras com conceitos muito longos. Nesse tipo de cenário, o método tende a perder precisão (**SALPERWYCK; BOULLÉ; LEMAIRE, 2015**, p. 2). Não obstante, o **EDDM** – proposto para melhorar as detecções do **DDM** em conceitos muito longos – não funciona bem na presença de mudanças de conceitos abruptas (**DU et al., 2014**, p. 2).

Por outro lado, um dos problemas com o **STEPD** está na utilização do teste estatístico entre proporções de duas amostras independentes para a detecção de mudanças de conceitos, mesmo para os casos de amostras com o tamanho pequeno e/ou desbalanceado, gerando, assim, perda de acurácia (**MEHTA; PATEL, 1996**, p. 1). Os seus autores alertaram sobre tal problema, entretanto, optaram por não utilizar o teste exato de Fisher (que é o mais recomendado em tais situações) (**AGRESTI, 1992**, p. 132) pelo seu alto custo computacional.

Diante do exposto anteriormente, torna-se perceptível a necessidade de explorar possíveis melhorias nos detectores citados. Dessa forma, esta pesquisa investiga como o **DDM**, **EDDM** e **STEPD** podem ser melhorados através do uso de testes estatísticos mais adequados, propondo, com base nisso, novos métodos para a detecção de mudanças de conceitos em fluxos de dados.

1.1 Objetivos

Inspirado no **DDM** e o **EDDM**, e tendo por base que esses detectores monitoram – respectivamente – a taxa de erro (**GAMA et al., 2004**, p. 289) e a distância média entre os erros do classificador (**BAENA-GARCIA et al., 2006**, p. 80), este trabalho propõe duas versões do *Equal Means Z – Test Concept Drift Detector* (**EMZD**), o qual utiliza o teste de hipóteses entre médias de duas amostras independentes para sinalizar as mudanças de conceitos, com o intuito de melhorar a acurácia dos métodos originais em bases de dados com mudanças abruptas e graduais, independentemente dos tamanhos de seus conceitos. Mais especificamente, a fim de detectar mudanças de conceitos, o **EMZD** utiliza o teste z para a comparação de duas médias, visto que, estatisticamente, esse teste é o mais recomendado nas comparações entre médias de duas amostras independentes suficientemente grandes (**LARSON; FARBER, 2010**, p. 354).

Além disso, inspirado no **STEPD** e tendo em vista que em cenários com amostras de tamanhos pequenos, com dados esparsos e/ou desequilibrados, o teste de hipóteses entre duas proporções tende a perder precisão (**MEHTA; PATEL, 1996**, p. 1), este trabalho propõe uma eficiente implementação do teste exato de Fisher – conhecido por manter a precisão independente dos tamanhos das amostras (**MCDONALD, 2014**, p. 65) – e sugere três diferentes aplicações (com o total de cinco versões) desse teste no problema de detecção de mudanças de conceitos, com o objetivo de melhorar a capacidade de detecção

do método original em cenários com amostras desbalanceadas e, dessa forma, alcançar melhores acurácias. Os três métodos propostos são:

1. *Fisher Proportions Drift Detector* (FPDD), o qual é uma variação do *STEPD*, utilizando o teste exato de Fisher apenas quando as amostras são pequenas;
2. *Fisher Square Drift Detector* (FSDD), que é similar ao *FPDD*, porém utiliza o teste do qui-quadrado para homogeneidade de proporções;
3. *Fisher Test Drift Detector* (FTDD), que utiliza apenas o teste exato de Fisher, independente do tamanho das amostras.

1.2 Metodologia

Levando em consideração os objetivos pretendidos, foi efetuada uma pesquisa exploratória, com o objetivo de realizar estudos críticos e comparativos, e identificar possíveis soluções para a melhoria do desempenho dos detectores de mudanças de conceitos pesquisados. Segue uma lista das atividades que foram realizadas no decorrer do projeto:

1. Apreciação do estado da arte dos testes estatísticos paramétricos e não paramétricos;
2. Apreciação do estado da arte dos detectores de mudanças de conceitos;
3. Levantamento bibliográfico sobre testes estatísticos paramétricos e não paramétricos;
4. Levantamento bibliográfico sobre detectores de mudanças de conceitos em fluxos de dados;
5. Levantamento bibliográfico sobre bases de dados reais e artificiais que contenham mudanças de conceitos em fluxos de dados;
6. Investigação do funcionamento detalhado do *DDM*, *EDDM* e *STEPD*, a fim de verificar pontos a serem melhorados;
7. Investigação dos testes estatísticos mais indicados para a realização das melhorias nos detectores analisados;
8. Implementação das melhorias nos detectores investigados, propondo os métodos *EMZD*, *FPDD*, *FSDD* e *FTDD*;
9. Realização dos experimentos comparativos, utilizando o ambiente *Massive Online Analysis* (MOA) (BIFET et al., 2010), com bases de dados reais e artificiais, a fim de avaliar o desempenho dos detectores propostos no que diz respeito às acurácias e precisão nas detecções de mudanças de conceitos;
10. Análise dos resultados dos experimentos comparativos e conclusões da pesquisa.

1.3 Organização

O restante deste trabalho está dividido da seguinte forma:

- Os capítulos 2, 3 compõem a revisão da literatura deste trabalho. Tal parte, é responsável por realizar um levantamento do estado da arte dos tópicos relevantes aos testes estatísticos aplicados as mudanças de conceitos em fluxos de dados. Mais especificamente, o capítulo 2 revisa conceitos importantes sobre os testes estatísticos utilizados neste trabalho. Não obstante, o capítulo 3 faz uma breve revisão sobre os principais conceitos de aprendizagem de máquina, com foco em fluxos de dados e mudanças de conceitos.
- Os capítulos 4, 5 e 6 constituem os métodos propostos e resultados desta pesquisa. Essa parte apresenta 4 novos detectores, divididos em 7 versões. Além disso, os resultados também apresentam experimentos objetivando a comparação entre os métodos propostos e alguns dos principais detectores da atualidade. De forma mais específica, no capítulo 4, são apresentadas duas versões de um detector que utiliza o teste de hipóteses entre médias de duas amostras independentes para realizar as detecções das mudanças de conceitos. Não obstante, o capítulo 5 trás 3 métodos, divididos em 7 versões, que detectam mudanças de conceitos através do teste exato de Fisher. Ademais, o capítulo 6 demonstra os preparativos, execuções e resultados dos testes comparativos, envolvendo todos os detectores apresentados neste trabalho.
- Por fim, o capítulo 7 tece as conclusões dessa dissertação. Esta última parte contém o detalhamento das considerações finais a respeito desta pesquisa, cita os artigos que foram escritos com base nos métodos propostos nesta dissertação e em pesquisas correlatas, além de descrever alguns possíveis trabalhos futuros.

2 ESTATÍSTICA

Este capítulo apresenta um estudo acerca dos testes estatísticos utilizados neste trabalho. Além disso, também são definidos os princípios básicos para a realização de um teste de hipóteses e destacadas as principais diferenças entre os testes paramétricos e não paramétricos.

2.1 Estatística Paramétrica e Não Paramétrica

Os testes estatísticos podem ser divididos em paramétricos e não paramétricos. De acordo com [Bluman \(2014, p. 672–673\)](#), os testes paramétricos foram projetados para inferências que dependam do conhecimento de parâmetros populacionais, tais como: médias, variâncias e proporções. Assim, para que um teste estatístico paramétrico possa ser aplicado, é necessário que as amostras a serem testadas sejam suficientemente grandes para se assumir o conhecimento de algum parâmetro populacional. Além disso, pode-se aplicar um teste paramétrico sempre que as populações que originaram as amostras analisadas forem normalmente distribuídas.

Por outro lado, ainda segundo [Bluman \(2014, p. 672–673\)](#), os testes estatísticos não paramétricos podem ser utilizados em inferências que não dependam do conhecimento dos parâmetros populacionais. Dessa forma, a execução de um teste não paramétrico não depende do conhecimento de nenhum parâmetro populacional, nem da garantia de normalidade nas distribuições das populações envolvidas.

2.2 Princípios Básicos dos Testes de Hipóteses

O principal objetivo de um teste de hipóteses é verificar se os parâmetros de duas, ou mais, populações são estatisticamente iguais, ou não, através de análises baseadas em suas amostras ([DAVIS; MUKAMAL, 2006, p. 1078](#)). Dessa forma, esse tipo de teste simplifica os questionamentos referentes às amostras analisadas em apenas duas hipóteses: nula e alternativa.

Um teste de hipóteses pode ser executado por meio dos seguintes passos ([ALLUA; THOMPSON, 2009, p. 109](#)):

1. Formulação das hipóteses nula e alternativa, as quais representam respectivamente a igualdade e a diferença entre o que está sendo comparado;

2. Definição de um nível de significância (α), o qual representa a probabilidade de incorretamente se rejeitar a hipótese nula, quando ela é verdadeira;
3. Definição e cálculo da estatística do teste;
4. Análise e conclusão baseada nas hipóteses nula e/ou alternativa.

2.3 Testes Estatísticos Paramétricos

Este trabalho utilizou, como referenciais, os seguintes testes estatísticos paramétricos: teste de hipóteses entre proporções de duas amostras independentes e teste de hipóteses entre médias de duas amostras independentes.

2.3.1 Teste de Hipóteses entre Proporções de Duas Amostras Independentes

De acordo com [Larson e Farber \(2010, p. 377–378\)](#), quando se deseja testar a diferença entre duas proporções populacionais (p_1 e p_2) através de uma proporção amostral de cada população (\hat{p}_1 e \hat{p}_2), deve-se utilizar uma variação do teste z para duas amostras independentes.

Entretanto, para que o teste z possa ser aplicado no teste entre proporções, [Larson e Farber \(2010, p. 377–378\)](#) também citam a necessidade de que alguns pré-requisitos sejam satisfeitos:

1. Aleatoriedade na seleção das amostras;
2. Independência entre as amostras;
3. O número de elementos de cada classe – em cada uma das duas amostras – deve ser maior ou igual a 5.

Uma vez que todos os pré-requisitos sejam atendidos, a execução do teste de hipóteses entre proporções de duas amostras independentes pode ser realizada através dos seguintes passos ([ALLUA; THOMPSON, 2009, p. 109](#)) ([LARSON; FARBER, 2010, p. 377–378](#)) ([BLUMAN, 2014, p. 504–507](#)):

1. Formulação das hipóteses nula (H_0) e alternativa (H_1), onde, $H_0 : p_1 = p_2$ e $H_1 : p_1 \neq p_2$ (para testes bicaudais) ou, ainda, $H_1 : p_1 < p_2$ ou $H_1 : p_1 > p_2$ (ambas, para testes unicaudais);
2. Definição de um nível de significância (α), o qual deverá ser um valor entre 0 e 1;

3. Definição e cálculo da estatística do teste, de acordo com a equação 2.1.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (2.1)$$

Onde: $\bar{p} = (x_1 + x_2)/(n_1 + n_2)$ e $\bar{q} = 1 - \bar{p}$. Além disso, x_1 e x_2 representam – respectivamente – o número de sucessos contidos nas amostras n_1 e n_2 .

Uma vez que a hipótese nula indica que $p_1 = p_2$, então, assume-se que $p_1 - p_2$ é igual a 0.

4. Análise e conclusão baseada nas hipóteses nula e/ou alternativa. Neste caso, deve-se encontrar o valor p – da tabela normal padrão – que está relacionado ao resultado de z – na equação 2.1 – e compará-lo com o nível de significância adotado. Caso o teste seja bicaudal, o valor de p deverá ser dobrado. Caso o valor p seja menor do que α , a hipótese $H_0 : p_1 = p_2$ será rejeitada.

2.3.2 Teste de Hipóteses entre Médias de Duas Amostras Independentes

Quando o objetivo de um teste de hipóteses – para duas amostras independentes – é verificar se existe diferença entre duas médias populacionais (μ_1 e μ_2), tomando por base uma média amostral de cada população (\bar{x}_1 e \bar{x}_2), pode-se utilizar os testes z ou t , caso as amostras satisfaçam as seguintes condições (LARSON; FARBER, 2010, p. 354–364):

1. Aleatoriedade na seleção das amostras;
2. Independência entre as amostras;
3. Cada amostra deve possuir pelo menos 30 elementos (teste z).
4. As amostras devem ter sido retiradas de uma população com distribuição normal (teste t).

A escolha de qual teste será utilizado para a comparação de duas médias populacionais deve ser feita da seguinte forma (BLUMAN, 2014, p. 484): se as amostras forem consideravelmente grandes (com pelo menos 30 elementos), recomenda-se o uso do teste z . Caso contrário, recomenda-se a aplicação do teste t se houver comprovações de que as amostras foram retiradas de uma população com distribuição normal. Com base nisso, uma vez que – neste trabalho – não será aplicado o teste de hipóteses entre duas médias que utilize amostras com menos de 30 exemplos, apenas o teste z foi tomado como referência.

Assim, caso todas as condições exigidas para a aplicação do teste z sejam atendidas, a execução do teste de hipóteses entre médias de duas amostras independentes segue

os passos abaixo (ALLUA; THOMPSON, 2009, p. 109) (LARSON; FARBER, 2010, p. 354–355) (BLUMAN, 2014, p. 473–475):

1. Formulação das hipóteses nula (H_0) e alternativa (H_1). Neste caso, a hipótese nula é dada por $H_0 : \mu_1 = \mu_2$. Já as hipóteses alternativas podem ser: $H_1 : \mu_1 \neq \mu_2$ (para testes bicaudais) ou, ainda, $H_1 : \mu_1 < \mu_2$ ou $H_1 : \mu_1 > \mu_2$ (ambas, para testes unicaudais);
2. Definição de um nível de significância (α). Uma vez que este nível representa uma porcentagem, o seu valor deverá sempre estar entre 0 e 1;
3. Definição e cálculo da estatística do teste, de acordo com a equação 2.2.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (2.2)$$

Onde: σ_1^2 e σ_2^2 representam, respectivamente, as variâncias das populações que geraram as amostras n_1 e n_2 . Além disso, como a hipótese nula indica que $\mu_1 = \mu_2$, então, é assumido que $\mu_1 - \mu_2$ é igual a 0.

4. Análise e conclusão baseada nas hipóteses nula e/ou alternativa. Para este caso, o valor p que está relacionado ao resultado da equação 2.2, deve ser encontrado na tabela da normal padrão e comparado com o nível de significância escolhido. Caso o teste seja bicaudal, o valor de p deverá ser dobrado. A hipótese nula, representada por $H_0 : \mu_1 = \mu_2$, será rejeitada apenas se o valor p for menor do que α .

2.4 Testes Estatísticos Não Paramétricos

Este trabalho utilizou, como referenciais, os seguintes testes estatísticos não paramétricos: teste exato de Fisher e teste de qui-quadrado de homogeneidade das proporções.

2.4.1 Teste Exato de Fisher

Proposto com a ideia de realizar a análise de tabelas de contingência e sendo diferente da maioria dos testes estatísticos que dependem de uma aproximação que tenderá a se tornar exata à medida que o tamanho das amostras tenderem ao infinito, o teste exato de Fisher (AGRESTI, 1992, p. 132) é recomendado para a análise de amostras de quaisquer tamanhos.

Esse teste, é chamado de exato porque a sua estatística permite encontrar diretamente os p -values (AGRESTI, 1992, p. 134–135), sem a necessidade de se recorrer a nenhuma tabela.

Em um cenário com amostras de tamanhos pequenos, com dados esparsos e/ou desequilibrados, os testes que dependem de aproximação tendem a perder precisão (MEHTA; PATEL, 1996, p. 1). Por outro lado, o teste exato de Fisher consegue manter a precisão independente dos tamanhos das amostras (MCDONALD, 2014, p. 65).

Embora possa ser utilizado para análise em tabelas com dimensões generalizadas para $m \times n$ (MEHTA; PATEL, 1983, p. 427), o teste exato de Fisher é, normalmente, recomendado para tabelas de contingência com tamanho 2×2 , se adequando perfeitamente ao exemplo demonstrado a seguir.

Supondo um conjunto de dados binomial – contendo os erros e acertos de um determinado classificador – dividido em duas amostras, onde, a amostra 1 possui 30 exemplos e a amostra 2 possui 100 exemplos, pode-se ter, por exemplo, a tabela 1.

Tabela 1 – Tabela de Contingência para o Teste Exato de Fisher

Classes/Amostras	Amostra 1	Amostra 2	Total
Erros	10	52	62
Acertos	20	48	68
Total	30	100	130

Uma pergunta que se pode fazer sobre estes dados é: sabendo que 62 destes 130 exemplos são rotulados como erros e que 30 dos 130 pertencem a amostra 1 e, além disso, assumindo a hipótese nula de que as amostras, 1 e 2, são igualmente propensas aos erros, qual é a probabilidade de que estes 62 erros estejam desigualmente distribuídos entre as duas amostras? Se forem escolhidos 62 exemplos ao acaso, qual será a probabilidade de que 52 ou mais deles estejam entre os 100 exemplos da amostra 2 e que 10 ou menos deles estejam entre os 30 exemplos da amostra 1?

Antes de seguir para o teste exato de Fisher, é interessante generalizar os dados, conforme apresentado na tabela 2.

Tabela 2 – Tabela de Contingência Generalizada para o Teste Exato de Fisher

Classes/Amostras	Amostra 1	Amostra 2	Total
Erros	a	b	a + b
Acertos	c	d	c + d
Total	a + c	b + d	a + b + c + d = n

Fisher demonstrou que a probabilidade de se obter qualquer conjunto de valores vindos da tabela 2 é dada pela equação 2.3 (FISHER, 1934, p. 99–101), onde o “!” representa o símbolo de *fatorial*.

$$p = \frac{\binom{a+b}{a} \times \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)! \times (c+d)! \times (a+c)! \times (b+d)!}{a! \times b! \times c! \times d! \times n!} \quad (2.3)$$

Dessa forma, a hipótese nula de que as amostras descritas no exemplo anterior, são igualmente propensas aos erros, só deverá ser rejeitada se o valor de p - aplicado a tabela de contingência do exemplo - for menor do que o nível de significância adotado.

Assim, para o exemplo das janelas, apresentado na tabela 1, formula-se a expressão dada pela equação 2.4.

$$p = \frac{(62)! \times (68)! \times (30)! \times (100)!}{10! \times 52! \times 20! \times 48! \times 130!} \cong 0,03 \quad (2.4)$$

Como a ideia principal é verificar se a amostra 1 é mais propensa a erros do que a amostra 2, ou vice-versa, o teste em questão é considerado como bicaudal. Dessa forma, o valor de p deve ser dobrado (ARMITAGE; BERRY; MATTHEWS, 2002, p. 136). Ficando, então: $p \cong 0,06$.

Adotando, por exemplo, um nível de significância de 0,05, não pode-se rejeitar a hipótese nula, uma vez que $0,06 > 0,05$. Entretanto, se o nível de significância adotado for de 0,10, a hipótese nula será rejeitada, pois: $0,06 < 0,10$.

2.4.2 Teste de Qui-Quadrado de Homogeneidade das Proporções

O teste do qui-quadrado (χ^2) pode ser utilizado com vários objetivos diferentes. Uma de suas aplicações envolve a análise de tabelas de contingência e consiste na verificação da existência, ou não, de homogeneidade entre as proporções de duas, ou mais, amostras vindas de populações diferentes (LARSON; FARBER, 2010, p. 477).

Ao realizar a análise de um conjunto de dados dividido em duas amostras, com dados binomiais referentes aos erros e acertos de um classificador, pode-se ter, por exemplo, a tabela de contingência 3.

Tabela 3 – Tabela de Contingência Generalizada para o Teste do Qui-Quadrado de Homogeneidade das Proporções

Classes/Amostras	Amostra 1	Amostra 2	Total
Erros	a	b	a + b
Acertos	c	d	c + d
Total	a + c	b + d	a + b + c + d = n

Antes da aplicação do teste do qui-quadrado de homogeneidade das proporções, é preciso se verificar a satisfação dos seguintes requisitos (BLUMAN, 2014, p. 613):

1. Os dados devem ser obtidos a partir de amostras aleatórias;
2. O valor esperado em cada célula deve ser, pelo menos, 5.

Uma vez que todos os requisitos para a aplicação do teste do qui-quadrado de homogeneidade das proporções sejam satisfeitos, a sua execução seguirá os seguintes passos (ALLUA; THOMPSON, 2009, p. 109) (LARSON; FARBER, 2010, p. 477) (BLUMAN, 2014, p. 613):

1. Formulação das hipóteses nula ($H_0 : p_1 = p_2$) e alternativa ($H_1 : p_1 \neq p_2$), onde, p_1 e p_2 representam as proporções das duas populações analisadas;
2. Definição de um nível de significância (α), com um valor entre 0 e 1;
3. Cálculo dos graus de liberdade, valores esperados e da estatística do teste, de acordo com as equações 2.5, 2.6 e 2.7, respectivamente.

$$\text{graus de liberdade} = (r - 1) \times (c - 1) \quad (2.5)$$

$$E = \frac{\text{soma das linhas} \times \text{soma das colunas}}{\text{tamanho da amostra}} \quad (2.6)$$

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad (2.7)$$

Onde: r e c representam o número de linhas e o número de colunas, respectivamente. Além disso, o E representa o valor esperado e o O a frequência obtida.

4. Análise e conclusão baseada na hipótese nula. Neste caso, o valor p relacionado ao resultado da equação 2.7 deverá ser encontrado na tabela χ^2 (de acordo com os graus de liberdade definidos na equação 2.5) e comparado com o nível de significância adotado. A hipótese nula ($H_0 : p_1 = p_2$) será rejeitada apenas se o valor p for menor do que α . Além disso, é importante salientar que este teste estatístico é sempre unicaudal (BLUMAN, 2014, p. 594).

2.5 Considerações Finais

Este capítulo descreveu os testes estatísticos paramétricos e não paramétricos que serviram de base para a realização desta pesquisa.

O capítulo 3 apresenta os principais tópicos referentes a aprendizagem de máquina e detecções de mudanças de conceitos em fluxos de dados.

3 APRENDIZAGEM DE MÁQUINA E FLUXOS DE DADOS

Este capítulo realiza um levantamento do estado da arte referente à detecção de mudanças de conceitos em fluxos de dados. Nesse contexto, são apresentados alguns dos principais detectores de mudanças de conceitos da publicados na literatura acadêmica. Além disso, também são revisados alguns conceitos básicos sobre aprendizagem de máquina e fluxos de dados

3.1 Aprendizagem de Máquina

Segundo [Mitchell \(1997, p. 2–3\)](#), a [Aprendizagem de Máquina \(AM\)](#) é definida como a capacidade de um programa computacional ter o seu desempenho melhorado, na execução de uma determinada tarefa, através da experiência.

Quanto ao tipo de paradigma utilizado no aprendizado, as tarefas da [AM](#) podem ser divididas em: preditivas ou descritivas. Dessa forma, os algoritmos que realizam a indução de modelos preditivos, através da classificação ou regressão, seguem o paradigma do aprendizado supervisionado. Por outro lado, a meta do aprendizado não supervisionado é a exploração e descrição de um conjunto de dados por meio de agrupamento, associação ou sumarização ([FACELI et al., 2011, p. 5-6](#)).

Também é possível se dividir a [AM](#) quanto aos seus protocolos de treinamentos, dos quais, segundo [Duda, Hart e Stork \(2001, p. 293-294\)](#), os três mais conhecidos são: estocástico, lote ou *online*. No protocolo estocástico, os dados do conjunto de treinamento são apresentados aos classificadores de forma aleatória. Já no aprendizado em lote, todos os exemplos de treinamento são apresentados aos preditores antes do início da aprendizagem. Por último, no treinamento online, cada instância é apresentada apenas uma única vez aos algoritmos de aprendizagem, sem o armazenamento dos exemplos de treinamento.

3.2 Aprendizagem Supervisionada

Os algoritmos que seguem o paradigma do aprendizado supervisionado, conforme já mencionado, são propostos para a realização da tarefa de predição. De acordo com [Faceli et al. \(2011, p. 6\)](#), o objetivo da tarefa de predição é, tendo por base os atributos de entrada dos elementos pertencentes ao conjunto de dados de treinamento – e conhecendo os seus respectivos atributos de saída (rótulos) –, encontrar determinada função que seja capaz de prever um rótulo que represente um novo exemplo.

Assim, dada uma sequência de exemplos na forma de pares (\vec{x}_i, y_i) , onde \vec{x}_i é um vetor de atributos e y_i é a sua classe correspondente, para cada exemplo, o classificador – que trabalha de forma supervisionada – irá realizar uma predição (\hat{y}_i) que será comparada com o resultado atual (y_i), a fim de decidir se a predição foi correta ($\hat{y}_i = y_i$), ou não (GAMA et al., 2004, p. 288).

3.3 Aprendizagem Não Supervisionada

O aprendizado não supervisionado, ou descritivo, refere-se à identificação de informações relevantes nos dados, sem o direcionamento de nenhum elemento externo (FACELI et al., 2011, p. 178). Ou seja, nesse tipo de paradigma não se tem o conhecimento das classes referentes a cada instância do conjunto de dados analisado (DAYAN, 1999, p. 857–858).

O objetivo deste protocolo de aprendizagem é a representação dos dados de entrada através da identificação de suas propriedades. Dessa forma, ainda segundo Faceli et al. (2011, p. 178), as tarefas do aprendizado não supervisionado podem ser divididas em:

1. Sumarização: com o objetivo de encontrar uma descrição simples e compacta dos dados;
2. Associação: com o objetivo de encontrar padrões frequentes de associações entre os atributos do conjunto de dados;
3. Agrupamento: com o objetivo de identificar grupos existentes nos dados, de acordo com a similaridade entre os objetos.

3.4 Fluxos de Dados

No contexto de AM, um fluxo de dados possui uma sequência ordenada, e potencialmente ilimitada, de exemplos que fluem continuamente em alta velocidade. Dentre as suas principais características, estão incluídos (BRZEZINSKI; STEAFNOWSKI, 2016, p. 1–2):

1. Fluxo contínuo;
2. Alto volume de dados (possivelmente tendendo ao infinito);
3. Rápida taxa de surgimento dos seus exemplos;
4. Suscetibilidade à mudança de conceito em suas distribuições de dados.

3.5 Mudanças de Conceitos em Fluxos de Dados

Ainda segundo Brzezinski e Steafnowski (2016, p. 2), uma vez que os conceitos das distribuições de dados dos fluxos podem ser modificados ao longo do tempo, esse fenômeno, chamado de mudança de conceito, tende a deteriorar a acurácia dos algoritmos de aprendizagem. Isso porque um classificador já treinado com um determinado conceito perderá a sua precisão quando for apresentado a um conceito novo.

Assim, de acordo com Bifet et al. (2010, p. 1601), os algoritmos desenvolvidos para fins de classificação em fluxos de dados, devem atender aos seguintes requisitos:

1. Processar um exemplo por vez – através do protocolo online –, seguindo uma sequência, num determinado intervalo de tempo e sem realizar a releitura dos dados;
2. Utilizar uma quantidade limitada de memória computacional;
3. Estar pronto para efetuar a predição em qualquer tempo;
4. Ser capaz de reagir às mudanças de conceitos em casos de evolução dos fluxos de dados.

As mudanças de conceitos em fluxos de dados podem ser categorizadas de diferentes formas. De acordo com a velocidade, as mudanças podem ser divididas em abruptas, quando a distribuição é modificada num único intervalo de tempo, ou em graduais, onde as modificações nos conceitos são mais sutis e necessitam de um maior número de instâncias para ocorrerem (MINKU; WHITE; YAO, 2010, p. 731).

Além disso, as mudanças de conceitos podem ser categorizadas como: reais ou virtuais. As mudanças reais acontecem quando um conjunto de exemplos de um fluxo de dados, em um intervalo de tempo, possui determinada distribuição de classes e, em outro momento, passa a ter as suas classes distribuídas de forma diferente (KOLTER; MALOOF, 2007, p. 2759). Por outro lado, nas mudanças de conceitos virtuais, rótulos das classes permanecem os mesmos, porém, a distribuição dos dados muda (DELANY et al., 2005, p. 188).

Existem diferentes tipos de abordagens para o tratamento de mudanças de conceitos. Gonçalves Jr. e Barros (2013, p. 1018–1019) citam quatro das principais metodologias. São elas:

1. Adaptação de classificadores originalmente propostos para funcionarem em modo *batch*;
2. Detecção de mudança de conceito, seguida pela criação de um novo classificador a fim de representar o novo contexto;

3. Combinação de classificadores;
4. Manutenção dos dados relativos aos conceitos aprendidos pelos classificadores com a ideia de lidar com mudanças de conceitos recorrentes.

Diversos exemplos de aplicações precisam lidar com mudanças de conceitos em fluxos de dados. Dentre eles, pode-se citar:

1. Gerenciamento em tempo real de fluxos de dados de sensores de movimento (LEE; WANG; RYU, 2007);
2. Captura de contextos recorrentes em filtros de *spam* através da combinação de classificadores (KATAKIS; TSOUMAKAS; VLAHAVAS, 2010);
3. Abordagens para detecção de intrusões no campo da segurança computacional (LANE; BRODLEY, 1998);
4. Análise de sentimentos na área financeira através do treinamento de máquinas de vetores de suporte em textos retirados do *Twitter* contendo discussões sobre determinadas companhias e seus produtos (SMAILOVIC et al., 2014).

3.6 Detectores de Mudanças de Conceitos

Os algoritmos propostos para funcionarem como detectores de mudanças de conceitos seguem a ideia de monitorar os resultados de um determinado classificador base – que funcione de forma supervisionada –, a fim de que, baseados na distribuição de erros e acertos do preditor, detectem e sinalizem quando o conceito do fluxo de dados analisado for modificado (BRZEZINSKI; STEAFNOWSKI, 2016, p. 7).

Em geral, os detectores trabalham com dois níveis de alarmes: *warning* e *drift* (ATTAR et al., 2012, p. 158), onde, o *drift* representa um nível maior de mudança na distribuição analisada e simboliza que, de fato, ocorreu uma modificação de conceito. Assim sendo, quando o nível de *warning* é sinalizado, uma nova instância do classificador base é criada e mantida em paralelo com o classificador antigo. Caso o nível de *drift* seja alcançado, o detector exclui o antigo classificador e mantém apenas o novo. Por outro lado, caso o sinal de *warning* passe a ser considerado um alarme falso, a nova instância do classificador é excluída.

Para este trabalho, foram tomados como referências os seguintes detectores: DDM (GAMA et al., 2004), EDDM (BAENA-GARCIA et al., 2006) e STEP (NISHIDA; YAMAUCHI, 2007). Além desses, também foram utilizados outros métodos para fins de comparação. São eles: ADWIN (BIFET; GAVALDÀ, 2007) e ECDD (ROSS et al., 2012).

3.6.1 DDM

O **DDM** assume que, de acordo com o modelo de aprendizagem provavelmente aproximadamente correta (MITCHELL, 1997, p. 201–223), se um determinado algoritmo preditor estiver recebendo – como entradas – exemplos de uma distribuição estacionária, então a sua taxa de erro irá decrescer quando o número de exemplos aumentar. Assim, um significativo aumento nos erros do classificador sugere uma mudança na distribuição aprendida e que o modelo de aprendizagem atual não é mais apropriado.

Uma vez que o detector foi projetado para trabalhar de forma conjunta com um classificador supervisionado, o método supõe que os erros e acertos do preditor seguem a forma de uma distribuição binomial, onde, para cada ponto i da sequência de exemplos aprendida, a taxa de erro é representada por p_i e o seu desvio padrão é dado por: $s_i = \sqrt{\frac{p_i(1-p_i)}{i}}$.

O algoritmo de detecção proposto gerencia duas variáveis – p_{min} e s_{min} – durante o treinamento. Propositalmente, o método inicializa p_{min} e s_{min} com um número real de alto valor. Assim, a cada instância avaliada, é comparado se $p_i + s_i < p_{min} + s_{min}$. Caso a expressão anterior seja verdadeira, p_{min} e s_{min} recebem, respectivamente, os valores de p_i e s_i .

Conforme mencionado anteriormente, é comum que os detectores de mudanças de conceitos trabalhem com dois níveis de alarmes (*warning* e *drift*). Para o **DDM**, os níveis de *warning* e *drift* são alcançados, respectivamente, quando $p_i + s_i \geq p_{min} + w \times s_{min}$ e $p_i + s_i \geq p_{min} + d \times s_{min}$. Onde, w e d são dois parâmetros configurados pelo usuário, os quais – por padrão – são 2 e 3, respectivamente. Além disso, o detector também possui o parâmetro n , o qual é responsável por determinar o número mínimo de instâncias que devem ser analisadas antes que o método passe a realizar as comparações citadas.

3.6.2 EDDM

O **EDDM** foi desenvolvido sob a mesma estrutura básica do **DDM**, porém, o seu principal objetivo é a melhoria da detecção na presença de drifts graduais.

A única diferença entre os dois métodos é o fato de que o **EDDM** considera a distância entre os erros de classificação ao invés da média do número de erros. Dessa forma, o método supõe que, enquanto a distribuição aprendida permanecer estacionária, o classificador irá errar menos e a distância entre os seus erros de predição irá aumentar. Caso contrário, o algoritmo irá sinalizar que houve uma mudança de conceito.

Diferente do **DDM**, o **EDDM** calcula a distância média entre dois erros (p'_i) e o seu desvio padrão (s'_i). O detector também mantém mais duas variáveis, p'_{max} e s'_{max} , as quais, são iniciadas – propositalmente – com um número real de alto valor. A cada novo erro do classificador base, o algoritmo de detecção irá comparar se: $p'_i + 2 \times s'_i > p'_{max} + 2 \times s'_{max}$.

Caso a expressão anterior seja verdadeira, p'_{max} e s'_{max} , irão receber – respectivamente – os valores de p'_i e s'_i .

O método também trabalha com dois níveis de alarmes: *warning* e *drift*. O nível de *warning* é alcançado, quando: $p'_i + 2 \times s'_i / p'_{max} + 2 \times s'_{max} < \alpha$. Já o gatilho do *drift* é disparado – anunciando que houve uma mudança de conceito – quando: $p'_i + 2 \times s'_i / p'_{max} + 2 \times s'_{max} < \beta$. Note-se que os valores de α e β são parâmetros definidos pelo usuário, os quais são configurados – respectivamente – com os seguintes valores, por padrão: 0,95 e 0,90. Além disso, o detector também possui o parâmetro e , o qual é responsável por determinar o número mínimo de erros cometidos pelo classificador antes de que o método passe a sinalizar possíveis mudanças de conceitos.

3.6.3 STEPD

O princípio básico do **STEPD** é considerar duas acurácias: recente e geral. O método supõe duas coisas: a precisão de um classificador para w exemplos recentes será igual à precisão geral (medida desde o início do processo de aprendizagem), desde que não haja mudança do conceito alvo. Além disso, uma diminuição significativa na precisão da janela recente sugere uma mudança de conceito.

Para realizar a comparação entre as precisões das duas janelas, o detector utiliza a estatística demonstrada na equação 3.1. r_o é o número de predições corretas sobre todos os n_o exemplos, com exceção dos W exemplos recentes, e o r_r é o número de predições corretas sobre todos os w (n_r) exemplos. Além disso, $\hat{p} = (r_o + r_r) / (n_o + n_r)$.

$$T(r_o, r_r, n_o, n_r) = \frac{|r_o/n_o - r_r/n_r| - 0,5(1/n_o + 1/n_r)}{\sqrt{\hat{p}(1-\hat{p})(1/n_o + 1/n_r)}} \quad (3.1)$$

A estatística da equação 3.1 é, na verdade, um teste de hipóteses para comparação entre proporções de duas amostras independentes. Dessa forma, o resultado dessa equação deve ser utilizado para encontrar o valor p através da tabela da distribuição normal padrão.

Caso o valor p seja menor do que o nível de significância adotado (α_w para *warning* e α_d para *drift*), a hipótese nula ($r_o/n_o = r_r/n_r$) será rejeitada e o detector entrará em *warning* ou em *drift* (de acordo com o valor de *alpha* comparado).

Segundo os autores do método, a estatística que mede a diferença entre as proporções dos acertos não é eficiente nos casos onde os tamanhos das janelas são extremamente pequenos. Para esses casos, o teste exato de Fisher seria o recomendado, porém, não foi utilizado sob a justificativa do alto custo computacional.

3.6.4 ADWIN

O **ADWIN** tem como objetivo a detecção de mudanças de conceitos através de uma janela (W) com tamanho dinamicamente ajustável.

A ideia principal do **ADWIN** é, ao analisar os dados de retorno de um classificador base (erros e acertos), manter em W apenas as instâncias pertencentes a uma mesma distribuição. Dessa forma, quanto mais tempo os dados analisados permanecerem numa determinada distribuição de probabilidade, maior será o tamanho de W . Por outro lado, a janela W terá seu tamanho reduzido na presença de mudanças de conceitos.

A fim de verificar a existência de possíveis mudanças de conceitos na distribuição de erros – e acertos – de um determinado preditor, o **ADWIN** divide W em duas sub-janelas (W_0 e W_1) e realiza a comparação entre as suas médias ($\hat{\mu}_{W_0}$ e $\hat{\mu}_{W_1}$). Assim, a janela W terá o seu tamanho reduzido apenas se a diferença entre $\hat{\mu}_{W_0}$ e $\hat{\mu}_{W_1}$ for maior – ou igual – ao limiar ϵ_{cut} , o qual é definido pela equação 3.2. Nessa equação, m representa a média harmônica de n_0 e n_1 , os quais, são os tamanhos de W_0 e W_1 , respectivamente. Além disso, σ_W^2 é a variância observada nos elementos de W , $\delta \in [0, 1]$ refere-se ao nível de confiança (definido como 0,002 por padrão), e δ' é dado por: $\delta' = \delta / \log(n)$.

$$\epsilon_{cut} = \sqrt{\frac{2}{m} \cdot \sigma_W^2 \cdot \log\left(\frac{2}{\delta'}\right)} + \frac{2}{3m} \cdot \log\left(\frac{2}{\delta'}\right) \quad (3.2)$$

3.6.5 ECDD

O **ECDD** é uma modificação do *Exponentially Weighted Moving Average (EWMA)* (ROBERTS, 1959) com o objetivo de trabalhar com tarefas de classificação em fluxos de dados e ser capaz de reagir diante de possíveis mudanças de conceitos.

Ao analisar os erros e acertos de um determinado classificador base, o **ECDD** – para o instante de tempo t – armazena a saída do preditor analisado na variável X_t e define a sua probabilidade de erro como p_t . Assim, a detecção de possíveis mudanças de conceitos é realizada monitorando quaisquer variações em p_t .

Para sinalizar *warning* e *drift*, o **ECDD** utiliza as inequações definidas, respectivamente, nas fórmulas 3.3 e 3.4. No instante de tempo t , Z_t representa a estimativa da média dos acertos do classificador base analisado e $\hat{p}_{0,t}$ corresponde a variação na sua probabilidade de erro, sendo – respectivamente – definidos pelas equações 3.5 e 3.6. Além disso, L_t é um limiar que define a distância necessária para a detecção de uma mudança de conceito no instante de tempo t . Por fim, $\hat{\sigma}_{Z_t}$ refere-se ao desvio padrão de Z_t e tem o

seu valor definido pela equação 3.7.

$$Z_t > \hat{p}_{0,t} + \frac{L_t \cdot \hat{\sigma}_{Z_t}}{2} \quad (3.3)$$

$$Z_t > \hat{p}_{0,t} + L_t \cdot \hat{\sigma}_{Z_t} \quad (3.4)$$

$$Z_t = (1 - \lambda)Z_{t-1} + \lambda X_t \quad (3.5)$$

$$\hat{p}_{0,t} = \frac{t}{t+1} \hat{p}_{0,t-1} + \frac{1}{t+1} X_t \quad (3.6)$$

$$\hat{\sigma}_{Z_t} = \sqrt{\frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2t})} \hat{\sigma}_{X_t} \quad (3.7)$$

Nas equações 3.5 e 3.7, λ é um parâmetro definido para ponderar o aumento das instâncias atuais em relação às mais antigas. Também na equação 3.7, $\hat{\sigma}_{X_t}$ representa o desvio padrão de X_t .

O ECDD também possui o parâmetro n responsável por determinar o número mínimo de instâncias que devem ser analisadas antes que o método passe a detectar possíveis mudanças de conceitos.

3.6.6 Resumo dos Detectores de Mudanças de Conceitos

A fim de facilitar a compreensão do funcionamento dos os detectores de mudanças de conceitos referenciados neste trabalho, a tabela 4 apresenta um resumo contendo as suas principais características.

Tabela 4 – Tabela com as Principais Características dos Detectores

Método	Características	Parâmetros	Detecção	
			Warning	Drift
DDM	Monitora a taxa de erro do classificador.	$w = 2$ $d = 3$ $n = 30$	$p_i + s_i \geq p_{min} + w \times s_{min}$	$p_i + s_i \geq p_{min} + d \times s_{min}$
EDDM	Monitora a distância média entre os erros do classificador.	$\alpha = 0,95$ $\beta = 0,90$ $e = 30$	$p'_i + 2 \times s'_i / p'_{max} + 2 \times s'_{max} < \alpha$	$p'_i + 2 \times s'_i / p'_{max} + 2 \times s'_{max} < \beta$
STEPD	Divide as predições do classificador em duas janelas e compara as proporções de erros e acertos entre elas.	$\alpha_w = 0,05$ $\alpha_d = 0,003$ $w = 30$	$\alpha_w \leq \text{valor } p$	$\alpha_d \leq \text{valor } p$
ADWIN	Mantém uma janela dinâmica com as predições do classificador.	$\delta = 0,002$	$ \hat{\mu}_{W_0} - \hat{\mu}_{W_1} \geq \epsilon_{cut}$	
ECDD	Monitora variações na probabilidade de erro do classificador.	$\lambda = 0,2$ $n = 30$	$Z_t > \hat{p}_{0,t} + \frac{L_t \cdot \hat{\sigma}_{Z_t}}{2}$	$Z_t > \hat{p}_{0,t} + L_t \cdot \hat{\sigma}_{Z_t}$

3.7 Algoritmos de Aprendizagem Supervisionada em Fluxos de Dados

Uma vez que o foco deste trabalho são os detectores de mudanças de conceitos – e esses métodos funcionam monitorando os resultados de um classificador base (BRZEZINSKI; STEAFNOWSKI, 2016, p. 7) – esta dissertação tomou como referências apenas algoritmos propostos para funcionarem através do protocolo de aprendizagem supervisionada em fluxos de dados. Foram eles: *Naive Bayes* (NB) e *Hoeffding Tree* (HT).

3.7.1 Naive Bayes

O NB (JOHN; LANGLEY, 1995) é um algoritmo de classificação conhecido pela sua simplicidade e o seu baixo custo computacional. O classificador utiliza-se do teorema de *Bayes* para realizar as predições, assumindo que todas as entradas são independentes.

Para cada instância i não rotulada, o NB prediz uma classe C , com base na probabilidade *a posteriori* da classe C , dada a instância i .

3.7.2 Hoeffding Tree

Sendo um algoritmo incremental, capaz de formar uma árvore de decisão indutiva e aprender a partir de um fluxo de dados, o HT (HULTEN; SPENCER; DOMINGOS, 2001) assume que os exemplos da distribuição com os dados analisados não se alteram ao longo do tempo.

Este classificador explora o fato de que uma pequena amostra pode ser suficiente para a escolha de um atributo ótimo para a divisão. Tal ideia é apoiada, matematicamente, pelo *hoeffding bound*, o qual quantifica o número de instâncias necessárias para estimar as estatísticas necessárias para o aprendizado.

3.8 Considerações Finais

Neste capítulo foram apresentados os principais tópicos referentes a aprendizagem de máquina e detecções de mudanças de conceitos em fluxos de dados.

O capítulo 4 descreve duas novas versões de detectores de mudanças de conceitos baseados no teste de hipóteses entre médias de duas amostras diferentes.

4 MÉTODOS BASEADOS NO TESTE DE HIPÓTESES ENTRE DUAS MÉDIAS

Neste capítulo, são propostas duas novas versões de detectores de mudanças de conceitos baseados no teste de hipóteses entre médias de duas amostras independentes. Cada versão é apresentada através de uma descrição geral e um pseudo-código detalhando o processo de detecção de mudanças de conceitos.

4.1 Descrição Geral dos Métodos Baseados no Teste de Hipóteses entre Duas Médias

Conforme apresentado na revisão da literatura deste trabalho, o [DDM](#) – assim como o [EDDM](#) – são métodos que funcionam monitorando as respostas de um determinado classificador base e estimando as posições dos *warnings* e *drifts*, a fim de detectarem possíveis mudanças de conceitos na distribuição dos dados analisados.

O [DDM](#) realiza a detecção de mudanças de conceitos com base na taxa de erro média do classificador monitorado. Dessa forma, uma conhecida dificuldade enfrentada por esse método reside em amostras com conceitos muito longos. Nesse tipo de cenário, o método tende a perder precisão ([SALPERWYCK; BOULLÉ; LEMAIRE, 2015](#), p. 2). Por outro lado, o [EDDM](#) detecta possíveis modificações nos conceitos através da distância média entre os erros do preditor analisado. Não obstante, esse detector não funciona bem na presença de mudanças de conceitos abruptas ([DU et al., 2014](#), p. 2).

Tendo por base que o teste z é o mais recomendado nas comparações entre médias de duas amostras independentes suficientemente grandes ([LARSON; FARBER, 2010](#), p. 354), este trabalho propõe um novo método, [EMZD](#), o qual utiliza esse teste de hipóteses para a detecção de mudanças de conceitos em fluxos de dados, com o objetivo de melhorar a eficiência do [DDM](#) e do [EDDM](#). Para isso, foram propostas duas versões do [EMZD](#): o *Equal Means Z – Test Concept Drift Detector 1* ([EMZD1](#)) e o *Equal Means Z – Test Concept Drift Detector 2* ([EMZD2](#)).

O [EMZD1](#) funciona de forma similar ao [DDM](#), monitorando a taxa de erro média do classificador. Já o [EMZD2](#) foi projetado com base no [EDDM](#), analisando a distância média entre os erros do preditor monitorado. Dessa forma, o fluxo geral das duas versões do [EMZD](#) segue de maneira semelhante aos métodos originais ([DDM](#) e [EDDM](#)), diferindo deles apenas nos detalhes dos testes estatísticos aplicados para as detecções de mudanças de conceitos.

O [EMZD](#), em suas duas versões, também monitora as predições de um classificador base, a fim de sinalizar possíveis *warnings* e *drifts*. Além disso, o [EMZD1](#) e o [EMZD2](#), utilizam os seguintes parâmetros que podem ser configurados pelo usuário: o número mínimo de instâncias para a aplicação dos testes estatísticos ($n = 30$), o número mínimo de erros antes das detecções de possíveis mudanças de conceitos ($e = 30$ – apenas para o [EMZD2](#)) e os níveis de significância para as sinalizações de *warnings* e *drifts* ($\alpha_w = 0,25$ e $\alpha_d = 0,22$).

A escolha do melhor conjunto de parâmetros foi realizada através de uma série de experimentos exploratórios utilizando o algoritmo genético proposto por [Santos, Barros e Gonçalves Jr. \(2015\)](#). Foi verificado, por meio de experimentos realizados sobre um grande número de conjuntos de dados, que as duas versões do [EMZD](#) não funcionam bem com baixos níveis de significância. Não obstante, a variação no número mínimo de instâncias para a aplicação dos testes estatísticos, não produziu resultados significativamente melhores.

4.2 EMZD1

Com o objetivo de realizar a detecção de mudanças de conceitos, o [DDM](#) sinaliza a ocorrência de *warnings* e *drifts* quando, respectivamente, $p_i + s_i \geq p_{min} + w \times s_{min}$ e $p_i + s_i \geq p_{min} + d \times s_{min}$, onde a taxa de erro é representada por p_i e o seu desvio padrão é s_i . Além disso, a menor taxa de erro registrada – e o seu desvio padrão – são, respectivamente: p_{min} e s_{min} . Por fim, w e d são dois parâmetros configurados pelo usuário, os quais representam o escore padrão em relação a p_{min} .

Proposto com o objetivo de aprimorar o processo de detecção de mudanças de conceitos do [DDM](#), o [EMZD1](#) realiza a comparação entre a menor taxa de erro média registrada (p_{min}) e a taxa de erro atual (p_i) por meio do teste de hipóteses entre médias de duas amostras independentes, conforme demonstrado pelo algoritmo 1.

As linhas 1–2 apresentam um resumo dos dados que precisam ser inicializados juntamente com o método.

A parte principal do algoritmo reside nas linhas 3–26. Após a detecção de uma mudança de conceito (linhas 4–6), os ajustes necessários são implementados pelo [EMZD1](#) apenas quando ele recebe a primeira instância do novo conceito. O valor da expressão *changeDetected*, na linha 4, é definido pela análise da instância anterior.

É importante notar que, no [MOA](#), as modificações realizadas no classificador base – após a detecção de uma mudança de conceito – não são implementadas diretamente no códigos dos detectores. Eles apenas sinalizam a posição dessas mudanças para outras classes compartilhadas no *framework*.

Algorithm 1: *Equal Means Z – Test Drift Detector 1*

Input: Fluxo de Dados S , Número Mínimo de Instâncias Antes do *Drift* n , Nível do *Drift* α_d , Nível do *Warning* α_w

```

1  $instances \leftarrow instances_{min} \leftarrow errors \leftarrow 0$ 
2  $mean_{min} \leftarrow 1$ 
3 foreach  $prediction$  in  $S$  do
4   if  $changeDetected$  then
5      $instances \leftarrow instances_{min} \leftarrow errors \leftarrow 0$ 
6      $mean_{min} \leftarrow 1$ 
7      $changeDetected \leftarrow false$ 
8    $isWarningZone \leftarrow false$ 
9    $instances \leftarrow instances + 1$ 
10  if  $prediction = false$  then
11     $errors \leftarrow errors + 1$ 
12   $mean \leftarrow errors / instances$ 
13   $variance \leftarrow (instances \times errors - errors^2) / (instances \times (instances - 1))$ 
14  if  $mean < mean_{min}$  and  $errors \neq 0$  then
15     $mean_{min} \leftarrow mean$ 
16     $variance_{min} \leftarrow variance$ 
17     $instances_{min} \leftarrow instances$ 
18  if  $instances > n$  and  $instances_{min} \neq 0$  then
19     $Z \leftarrow |mean - mean_{min}|$ 
20     $Z \leftarrow Z / \text{sqrt}(variance / instances + variance_{min} / instances_{min})$ 
21     $pValue \leftarrow \text{normalProbability}(Z)$ 
22     $pValue \leftarrow (1 - pValue)$ 
23    if  $pValue < \alpha_d$  then
24       $changeDetected \leftarrow true$ 
25    else if  $pValue < \alpha_w$  then
26       $isWarningZone \leftarrow true$ 

```

As linhas 9–22 referem-se a implementação do cálculo do valor de p pelo teste z . As linhas 9–11 atualizam os números de instâncias processadas e de erros, enquanto que as linhas 12–13 atualizam a média e a variância sobre essas instâncias processadas. Além disso, as linhas 14–17 são responsáveis pelo armazenamento da média, da variância e da quantidade de instâncias analisadas, quando a taxa de erro médio alcança o menor valor do conceito atual. Não obstante, a linha 18 garante que a detecção de *warnings* e *drifts* seja realizada apenas após o processamento de n instâncias.

As linhas 19–20 demonstram a execução do teste de hipóteses entre médias de amostras independentes. Ademais, nas linhas 21–22, o cálculo do valor de p é realizado através de uma função nativa do MOA – *normalProbability()*.

Finalmente, os *drifts* e *warnings* são detectados nas linhas 23–26.

4.3 EMZD2

Para a detecção de mudanças de conceitos, o **EDDM** comunica a existência de *warnings* e *drifts*, respectivamente, quando $p'_i + 2 \times s'_i / p'_{max} + 2 \times s'_{max} < \alpha$ e $p'_i + 2 \times s'_i / p'_{max} + 2 \times s'_{max} < \beta$, onde, p'_i é a distância média entre dois erros e s'_i refere-se ao seu desvio padrão. Ademais, p'_{max} e s'_{max} representam, respectivamente, o maior registro da distância média entre os erros e o seu desvio padrão. Finalmente, α e β referem-se a dois parâmetros que podem ser configurados pelo usuário e representam os intervalos de confiança contendo a distribuição das distâncias entre os erros do classificador analisado.

O **EMZD2** é um método proposto com a finalidade de aperfeiçoar a tarefa de detecção de mudanças de conceitos do **EDDM**. Dessa forma, a segunda versão do **EMZD** compara o maior registro da distância média entre os erros (p'_{max}) com a atual distância média entre os erros (p'_i) através do teste de hipóteses entre médias de duas amostras independentes. O pseudo-código desse processo é demonstrado pelo algoritmo 2.

Uma abstração dos dados que devem ser inicializados junto com o método é apresentada nas linhas 1–2.

As linhas 3–24 contêm a parte principal do algoritmo. Após detectar uma mudança de conceito, o **EMZD2** realiza as modificações necessárias apenas na ocasião em que recebe a primeira instância do novo conceito, nas linhas 4–6. Na linha 4, *changeDetected* tem o seu valor definido pela análise da instância anterior.

Após a detecção de uma mudança de conceito, conforme explicado anteriormente, as modificações realizadas no classificador base não são implementadas diretamente nos códigos dos detectores. Ao invés disso, no **MOA**, os detectores apenas indicam as posições das mudanças para outras classes compartilhadas no *framework*.

A implementação do cálculo do valor de p , através do teste z , é demonstrada nas linhas 9–29. A linha 9 atualiza o número de instâncias processadas. Já as linhas 11–13 registram a atualização do número de erros, da posição do último erro antes do atual, e da posição do erro atual. Não obstante, as linhas 14–15 atualizam a distância e o quadrado da distância, entre as posições do último erro antes do atual e do erro atual.

As linhas 16–20 computam a média e a variância sobre as distâncias entre os erros do classificador analisado. Além disso, as linhas 22–25 armazenam a média e variância das distâncias entre os erros, além da quantidade de erros, quando essa média alcança o maior valor do conceito atual (caso o algoritmo já tenha processado, pelo menos, n instâncias).

Após o processamento de n instâncias, a linha 26 garante que a execução do teste estatístico, bem como a detecção de *warnings* e *drifts*, ocorram apenas depois de e erros.

As linhas 27–30 referem-se à execução do teste de hipóteses entre médias de duas amostras independentes e ao cálculo do valor de p , o qual é realizado por intermédio da

Algorithm 2: *Equal Means Z – Test Drift Detector 2*

Input: Fluxo de Dados S , Número Mínimo de Instâncias Antes do *Drift* n , Número Mínimo de Erros Antes do *Drift* e , Nível do *Drift* α_d , Nível do *Warning* α_w

```

1  $instances \leftarrow errors \leftarrow actualErrors \leftarrow meanErrors_{max} \leftarrow 0$ 
2  $distance \leftarrow squaredDistance \leftarrow 0$ 
3 foreach  $prediction$  in  $S$  do
4   if  $changeDetected$  then
5      $instances \leftarrow errors \leftarrow actualErrors \leftarrow meanErrors_{max} \leftarrow 0$ 
6      $distance \leftarrow squaredDistance \leftarrow 0$ 
7      $changeDetected \leftarrow false$ 
8    $isWarningZone \leftarrow false$ 
9    $instances \leftarrow instances + 1$ 
10  if  $prediction = false$  then
11     $errors \leftarrow errors + 1$ 
12     $lastError \leftarrow actualError$ 
13     $actualError \leftarrow instances$ 
14     $distance \leftarrow distance + (actualError - lastError)$ 
15     $squaredDistance \leftarrow squaredDistance + (actualError - lastError)^2$ 
16     $meanErrors \leftarrow distance / errors$ 
17    if  $errors > 1$  then
18       $varianceErrors \leftarrow (errors \times squaredDistance - distance^2) / (errors \times (errors - 1))$ 
19    else
20       $varianceErrors \leftarrow 0$ 
21    if  $instances > n$  then
22      if  $meanErrors > meanErrors_{max}$  then
23         $meanErrors_{max} \leftarrow meanErrors$ 
24         $varianceErrors_{max} \leftarrow varianceErrors$ 
25         $errors_{max} \leftarrow errors$ 
26      else if  $errors > e$  then
27         $Z \leftarrow |meanErrors_{max} - meanErrors|$ 
28         $Z \leftarrow Z / \text{sqrt}(varianceErrors / errors + varianceErrors_{max} / errors_{max})$ 
29         $pValue \leftarrow \text{normalProbability}(Z)$ 
30         $pValue \leftarrow (1 - pValue)$ 
31        if  $pValue < \alpha_d$  then
32           $changeDetected \leftarrow true$ 
33        else if  $pValue < \alpha_w$  then
34           $isWarningZone \leftarrow true$ 

```

função $normalProbability()$, pertencente ao MOA.

Por fim, as linhas 31–34 demonstram o processo de detecção de *drifts* e *warnings*.

4.4 Considerações Finais

Neste capítulo foram propostas duas novas versões de detectores de mudanças de conceitos baseados no teste de hipóteses entre médias de duas amostras diferentes.

O capítulo 5 apresenta três novos detectores de mudanças de conceitos (formando um total de sete versões) baseados no teste exato de Fisher.

5 MÉTODOS BASEADOS NO TESTE EXATO DE FISHER

Este capítulo propõe três novos detectores de mudanças de conceitos, divididos em sete versões, baseados no teste exato de Fisher. Primeiramente, é apresentada uma descrição detalhada acerca da implementação desse teste estatístico para os métodos propostos neste capítulo. Em seguida, apresenta-se uma descrição geral dos detectores baseados no teste exato de Fisher, acrescida de um pseudo-código detalhando de forma abstrata o fluxo dos métodos. Por fim, cada versão dos detectores propostos é apresentada por meio de uma descrição e um pseudo-código contendo o fluxo detalhado do processo de detecção de mudanças de conceitos.

5.1 Implementação do Teste Exato de Fisher

Conforme mencionado na descrição do [STEPD](#), [Nishida e Yamauchi \(2007, p. 266\)](#) evitaram o uso do teste exato de Fisher, na implementação do método, apenas baseados no seu alto custo computacional.

Considerando o cálculo do valor de p apresentado na equação [2.3](#) (p. [23](#)), torna-se óbvio que são os numerosos cálculos dos fatoriais que tornam o teste computacionalmente custoso, uma vez que a sua implementação literal implicaria num cálculo médio de 9 fatoriais por instância analisada.

Todavia, olhando mais detalhadamente para o problema da análise das proporções de erros – e acertos – entre as duas janelas do [STEPD](#), é possível se observar que todos os fatoriais calculados referem-se a números no intervalo $[0..n]$, onde, n é o número de exemplos visto.

Dado que a janela recente do método em questão possui um tamanho fixo, uma possível forma de diminuir o custo computacional do cálculo dos fatoriais seria através da limitação do tamanho da sua janela antiga. Entretanto, a adoção dessa ideia significaria ignorar uma parte dos dados antigos.

Dessa forma, a fim de implementar uma solução para simplificar os cálculos dos fatoriais, torná-los computacionalmente eficientes e não descartar nenhum dado antigo, este trabalho ajustou – proporcionalmente – o tamanho da janela antiga do [STEPD](#) para se tornar o mesmo da janela com os dados recentes. Assim, considerando que a janela recente contém w exemplos, pôde-se assumir que $n = 2 \times w$ e, então, o cálculo dos fatoriais ficou restrito ao seguinte intervalo: $[0, 2w]$.

A implementação da janela com os dados antigos, utilizando uma proporção do seu tamanho, foi realizada através das equações 5.1 e 5.2. w_p e r_p representam, respectivamente, o número de erros e acertos da janela antiga com o tamanho proporcional. Além disso, w_o e r_o são os números de predições incorretas e corretas, respectivamente, sobre todos os n_o exemplos, com exceção dos w exemplos recentes.

$$w_p = w_o \times w/n_o \quad (5.1)$$

$$r_p = w - w_p \quad (5.2)$$

Uma vez que a janela com os dados antigos foi diminuída mantendo a sua proporção original, assumiu-se que esse ajuste não comprometeria a precisão dos testes estatísticos a serem utilizados neste trabalho, visto que, como será demonstrado nas seções seguintes, tanto o teste exato de Fisher como os demais testes aplicados na criação de novos detectores inspirados no **STEPD** tomam por base as frequências e proporções dos dados nas distribuições analisadas. Adicionalmente, foi observado experimentalmente – através de testes realizados com todas as bases de dados utilizadas neste trabalho – que o uso da janela antiga com o tamanho proporcionalmente reduzido não afetou os resultados apresentados nesta dissertação.

Assumindo que muitos cálculos de fatoriais seriam repetidamente necessários, o próximo passo, para providenciar uma implementação eficiente do teste exato de Fisher, foi calcular previamente – e armazenar num *array* – todos os possíveis resultados dos fatoriais que necessariamente precisariam ser calculados para a execução do teste. Dessa forma, todos os fatoriais necessários puderam ser calculados ao custo de apenas uma execução do fatorial de $2 \times w$, através do armazenamento de seus resultados intermediários. Assim, mesmo que se decidisse utilizar valores altos para w (o qual, por padrão, no **STEPD** é 30), o custo adicional de memória e tempo de execução, na prática, seriam pouco relevantes.

5.2 Descrição Geral dos Métodos Baseados no Teste Exato de Fisher

De acordo com a descrição do método proposto por [Nishida e Yamauchi \(2007\)](#), apresentada na revisão da literatura deste trabalho, o **STEPD** mantém estatísticas de duas janelas de dados: recente e antiga. Além disso, o método mantém a janela recente com um tamanho limitado, o qual – por padrão – é 30.

Assim, é muito comum que – em alguns casos – o número de erros (ou o número de predições corretas) na janela com os dados recentes seja muito pequeno. Nesses casos, a

aproximação estatística, utilizada pelo [STEPD](#), pode se tornar consideravelmente imprecisa ([MEHTA; PATEL, 1996](#), p. 1). Por outro lado, o teste exato de Fisher consegue manter a precisão independente dos tamanhos das amostras ([MCDONALD, 2014](#), p. 65).

Com o objetivo de criar detectores capazes de manterem a precisão estatística, independente do tamanho das janelas dos dados analisados, este trabalho propôs 3 novos métodos que utilizam o teste exato de Fisher na detecção de mudanças de conceitos em fluxos de dados. São eles: [FPDD](#), [FSDD](#) e [FTDD](#).

Todos os 3 métodos funcionam de forma muito similar ao [STEPD](#). Eles monitoram os resultados das previsões de um classificador base, usam duas janelas de dados (recente e antiga), aplicam testes estatísticos para sinalizarem a ocorrência de *warnings* e *drifts*, e, além disso, também possuem 3 parâmetros, os quais – com seus respectivos valores padrão – são: o tamanho da janela com os dados recentes ($w = 30$) e os níveis de significância para a detecção de *drifts* ($\alpha_d = 0,003$) e *warnings* ($\alpha_w = 0,05$).

Os parâmetros foram configurados com os mesmos valores padrão do [STEPD](#), com o intuito de realizar uma comparação justa entre os métodos. Ademais, foram realizados uma série de experimentos exploratórios utilizando o algoritmo genético, proposto por [Santos, Barros e Gonçalves Jr. \(2015\)](#), para a procura de um melhor conjunto de parâmetros. Entretanto, nos testes realizados sobre um grande número de bases de dados, não foi encontrada nenhuma combinação de parâmetros significativamente melhor.

Dessa forma, as únicas diferenças entre os métodos propostos por este trabalho e o [STEPD](#) residem na adoção de diferentes testes estatísticos de acordo com o nível de balanceamento entre as amostras.

Além disso, os detectores propostos medem diferenças entre os erros de predição, enquanto o [STEPD](#) utiliza o número de acertos. Todavia, essa modificação foi adotada unicamente porque a sua implementação no *framework* [MOA](#) é mais eficiente computacionalmente. A medição dos erros, ao invés dos acertos, não produz nenhuma diferença nos resultados dos testes estatísticos utilizados.

Por fim, os três métodos novos aplicam o teste exato de Fisher, baseada na tabela de contingência dos erros e acertos nas duas janelas de dados analisadas – conforme a implementação proposta na seção [5.1](#) – para calcular o valor de p .

O algoritmo [3](#) apresenta um pseudo-código abstrato do fluxo geral de execução dos três métodos propostos por este trabalho para a detecção de mudanças de conceitos.

As entradas dos métodos são o fluxo de dados e os três parâmetros, assim como no [STEPD](#).

As linhas 1–5 demonstram, de forma sumarizada, os dados que necessitam ser inicializados juntamente com os métodos. A expressão *methodStatistics* é uma abstração

Algorithm 3: Detectores de Mudanças de Conceitos Baseados no Teste Exato de Fisher

Input: Fluxo de Dados S , Tamanho da Janela Recente w , Nível do *Drift* α_d , Nível do *Warning* α_w

```

1 reset methodStatistics
2 changeDetected  $\leftarrow$  false
3 maximum  $\leftarrow 2 \times w$ 
4 factorial [ ]  $\leftarrow$  fatorial de todos os números no intervalo  $[0, \textit{maximum}]$ 
5 constantFactorial  $\leftarrow$  factorial $[w]^2 / \textit{factorial}[\textit{maximum}]$ 
6 foreach instance in  $S$  do
7   if changeDetected then
8     reset methodStatistics
9     changeDetected  $\leftarrow$  false
10  Atualiza as predições nas janelas antiga e recente
11  isWarningZone  $\leftarrow$  false
12  if  $n_o \geq w$  then
13     $w_p \leftarrow \text{round}(w_o \times w / n_o)$ 
14     $r_p \leftarrow w - w_p$ 
15    Calcula o valor de  $p$ , de acordo com o método específico
16    if  $pValue < \alpha_d$  then
17      changeDetected  $\leftarrow$  true
18    else if  $pValue < \alpha_w$  then
19      isWarningZone  $\leftarrow$  true

```

referente as variáveis específicas que cada método utiliza para realizar a detecção de mudanças de conceitos. A variável *factorial* refere-se ao *array* utilizado para armazenar os resultados dos fatoriais necessários para o cálculo do valor de p , por meio do teste exato de Fisher. Além disso, *constantFactorial* é utilizada para capturar a parte da equação do teste exato de Fisher que permanece constante por todas as instâncias.

As linhas 6–19 referem-se a parte principal do algoritmo. Os métodos implementam os ajustes que são necessários após a detecção de uma mudança de conceito apenas quando eles recebem a primeira instância do novo conceito (linhas 7–9). A expressão *changeDetected*, na linha 7, tem o seu valor definido na análise da instância anterior.

Conforme explicado na apresentação das duas versões do EMZD, após a detecção de uma mudança de conceito, o MOA não implementa as modificações no classificador base diretamente nos códigos dos detectores. Ao contrário, os métodos apenas sinalizam a posição de tais mudanças para outras classes compartilhadas do *framework*.

A linha 10 abstrai as atualizações que precisam ser efetuadas nas duas janelas a cada momento em que uma nova instância do conjunto de dados analisado é processada. Aqui, a instância nova é incluída na janela recente e, caso essa janela já tenha alcançado o seu tamanho máximo (w), a sua última instância será movida para a janela antiga.

A linha 12 garante que as detecções ocorram apenas após a janela antiga possuir, no mínimo, w instâncias, ou seja: após $2 \times w$ de instâncias processadas. Não obstante, as

linhas 13–14 realizam o tratamento dos valores proporcionais aos dados contidos na janela antiga, os quais serão necessários para o cálculo do valor de p .

A linha 15 abstrai o cálculo do valor de p , o qual irá diferir de um método para o outro. Os detalhes desses cálculos serão apresentados nos pseudo-códigos referentes a cada um dos 3 métodos propostos.

Por fim, os *drifts* e *warnings* são sinalizados nas linhas 16–19.

5.3 FPDD1 e FPDD2

O **FPDD** foi proposto com a ideia de utilizar o teste exato de Fisher nas situações onde – em, pelo menos, uma das janelas – o número de erros ou de predições corretas for consideravelmente pequeno. Caso contrário, o método utiliza o teste de hipóteses entre proporções de duas amostras independentes, assim como o **STEPD**.

Para decidir se numa das janelas o número de erros ou acertos é consideravelmente pequeno, ou não, foi necessário escolher um valor mínimo para que o teste de hipóteses entre duas proporções fosse aplicado. O valor mínimo escolhido foi 5 (**BLUMAN, 2014**, p. 505).

Além disso, a fim de ganhar eficiência, o **FPDD** aplica o teste de hipóteses entre duas proporções utilizando o número de erros da janela recente (w_r) e o valor proporcional ao número de erros contidos na janela antiga (w_p), ao invés de todos os erros dessa janela (w_o), uma vez que, conforme mencionado anteriormente, essa decisão não implica em mudanças significantes nos resultados.

Este trabalho implementou duas versões do **FPDD** que diferem-se, uma da outra, basicamente na análise da quantidade mínima de erros e acertos na janela com os dados antigos, para decidir qual teste estatístico deverá ser aplicado. Mais especificamente, a primeira versão, o *Fisher Proportions Drift Detector 1* (**FPDD1**), baseia-se nos valores proporcionais ao número de erros e acertos da janela antiga, aplicando o teste exato de Fisher apenas se: $w_p < 5$ ou $r_p < 5$. Por outro lado, o *Fisher Proportions Drift Detector 2* (**FPDD2**), toma por base a quantidade total de erros e acertos pertencentes à janela com os dados antigos, e aplica o teste exato de Fisher apenas quando: $w_o < 5$ ou $r_o < 5$. Em ambos os casos, quando a condição para a execução do teste exato de Fisher não for satisfeita, o teste de hipóteses entre proporções de duas amostras independentes é aplicado.

Os algoritmos 4 e 5 apresentam os pseudo-códigos com a implementação do cálculo do valor de p no **FPDD1** e **FPDD2**, respectivamente. É importante notar-se que, a diferença básica entre os dois métodos, encontra-se na linha 1. Não obstante, as linhas 4–5 são os resultados de simplificações na equação original do teste de hipóteses entre duas proporções, considerando que as duas janelas de dados possuem tamanho w , conforme demonstrado

pela equação 5.3. Por fim, nas linhas 6–7, o cálculo do valor de p é feito através de uma função nativa do MOA – `normalProbability()`. Além disso, por se tratar de um teste bicaudal, o valor de p é duplicado (ARMITAGE; BERRY; MATTHEWS, 2002, p. 136).

Demonstração. Considerando que o valor de n_p e n_r é igual ao valor de w , então a equação do teste de hipóteses entre proporções de duas amostras independentes, aplicada ao FPDD, pode ser simplificada da seguinte forma:

$$\begin{aligned}
 T(w_p, w_r, n_p, n_r) &= \frac{|w_p/n_p - w_r/n_r| - 0,5 \times (1/n_p + 1/n_r)}{\sqrt{(w_p + w_r)/(n_p + n_r) \times (1 - (w_p + w_r)/(n_p + n_r)) \times (1/n_p + 1/n_r)}} \\
 &= \frac{|w_p/w - w_r/w| - 0,5 \times (1/w + 1/w)}{\sqrt{(w_p + w_r)/(w + w) \times (1 - (w_p + w_r)/(w + w)) \times (1/w + 1/w)}} \\
 &= \frac{|(w_p - w_r)/w| - 0,5 \times 2/w}{\sqrt{(w_p + w_r)/2w \times (1 - (w_p + w_r)/2w) \times 2/w}} \\
 &= \frac{|w_p - w_r| - 1)/w}{\sqrt{1/w \times (w_p + w_r)/2 \times (1 - (w_p + w_r)/2w) \times 2/w}} \\
 &= \frac{|w_p - w_r| - 1)/w}{\sqrt{1/w \times (w_p + w_r) \times (1 - (w_p + w_r)/2w) \times 1/w}} \\
 &= \frac{(|w_p - w_r| - 1)/w}{1/w \times \sqrt{(w_p + w_r) \times (2w - (w_p + w_r))/2w}} \\
 &= \frac{|w_p - w_r| - 1}{\sqrt{(w_p + w_r) \times (2w - (w_p + w_r))/2w}}
 \end{aligned} \tag{5.3}$$

□

Algorithm 4: Cálculo do Valor de p pelo FPDD1

```

1 if  $w_r < 5$  or  $r_r < 5$  or  $w_p < 5$  or  $r_p < 5$  then
2   | Calcula o valor de  $p$  de acordo com o FTDD
3 else
4   |  $z \leftarrow w_p + w_r$ 
5   |  $z \leftarrow (|w_p - w_r| - 1) / \text{sqrt}(z \times (\text{maximum} - z) / \text{maximum})$ 
6   |  $pValue \leftarrow \text{normalProbability}(|z|)$ 
7   |  $pValue \leftarrow 2 \times (1 - pValue)$ 

```

5.4 FSDD1 e FSDD2

Similarmente ao FPDD, o FSDD também aplica o teste exato de Fisher quando o número de erros ou de predições corretas em pelo menos uma das duas janelas de dados analisadas é inferior a 5 (BLUMAN, 2014, p. 613). Todavia, nos demais cenários, o FSDD adota o teste do qui-quadrado de homogeneidade das proporções (LARSON; FARBER,

Algorithm 5: Cálculo do Valor de p pelo FPDD2

```

1 if  $w_r < 5$  or  $r_r < 5$  or  $w_o < 5$  or  $r_o < 5$  then
2   | Calcula o valor de  $p$  de acordo com o FTDD
3 else
4   |  $z \leftarrow w_p + w_r$ 
5   |  $z \leftarrow (|w_p - w_r| - 1) / \text{sqrt}(z \times (\text{maximum} - z) / \text{maximum})$ 
6   |  $pValue \leftarrow \text{normalProbability}(|z|)$ 
7   |  $pValue \leftarrow 2 \times (1 - pValue)$ 

```

2010, p. 477), ao invés do teste entre proporções de duas amostras independentes, usados no STEPD e no FPDD.

Para aplicar o teste do qui-quadrado de homogeneidade das proporções, o FSDD leva em consideração o número de erros e acertos da janela recente (w_r e r_r) – com os seus respectivos valores esperados (ew_r e er_r) – e o valor proporcional a quantidade de predições corretas e incorretas contidas na janela antiga (w_p e r_p) – também com os seus respectivos valores esperados (ew_p e er_p) – ao invés do todos os erros e acertos dessa janela (w_o e r_o). Conforme já mencionado, essa decisão não implica em mudanças significantes nos resultados.

Também foram implementadas duas versões do FSDD. De forma semelhante ao FPDD, a diferença entre as duas reside na maneira de analisar a quantidade mínima de predições corretas e incorretas pertencentes à janela antiga. a fim de decidir qual teste estatístico será aplicado. O *Fisher Square Drift Detector 1* (FSDD1) toma por base os valores proporcionais das quantidades de erros e acertos na janela com os dados antigos, e executa o teste exato de Fisher apenas se: $w_p < 5$ ou $r_p < 5$. Já o *Fisher Square Drift Detector 2* (FSDD2) leva em conta o número total de predições corretas e incorretas da janela antiga, executando o teste exato de Fisher apenas quando: $w_o < 5$ ou $r_o < 5$. Nos dois casos, quando a condição para a aplicação do teste exato de Fisher não for satisfeita, o teste do qui-quadrado de homogeneidade das proporções é utilizado.

Os pseudo-códigos, apresentados nos algoritmos 6 e 7, demonstram – respectivamente – a implementação do cálculo do valor de p , no FSDD1 e FSDD2. A única diferença entre os dois métodos reside na linha 1. Não obstante, as equações nas linhas 4–9 referem-se ao cálculo do teste do qui-quadrado de homogeneidade das proporções. Além disso, na linha 10, o cálculo do valor de p é realizado por uma função nativa do MOA (*chiSquaredProbability()*), a qual possui os seguintes parâmetros: o resultado da equação anterior e os graus de liberdade (que para este caso, de acordo com a equação 2.5, sempre será 1).

Algorithm 6: Cálculo do Valor de p pelo FSDD1

```

1 if  $w_r < 5$  or  $r_r < 5$  or  $w_p < 5$  or  $r_p < 5$  then
2   | Calcula o valor de  $p$  de acordo com o FTDD
3 else
4   |  $ew_r \leftarrow (w_r + w_p) \times (w_r + r_r) / \text{maximum}$ 
5   |  $er_r \leftarrow (r_r + r_p) \times (w_r + r_r) / \text{maximum}$ 
6   |  $ew_p \leftarrow (w_r + w_p) \times (w_p + r_p) / \text{maximum}$ 
7   |  $er_p \leftarrow (r_r + r_p) \times (w_p + r_p) / \text{maximum}$ 
8   |  $x \leftarrow (|w_r - ew_r|)^2 / ew_r + (|r_r - er_r|)^2 / er_r$ 
9   |  $\quad + (|w_p - ew_p|)^2 / ew_p + (|r_p - er_p|)^2 / er_p$ 
10  |  $pValue \leftarrow \text{chiSquaredProbability}(x, 1)$ 

```

Algorithm 7: Cálculo do Valor de p pelo FSDD2

```

1 if  $w_r < 5$  or  $r_r < 5$  or  $w_o < 5$  or  $r_o < 5$  then
2   | Calcula o valor de  $p$  de acordo com o FTDD
3 else
4   |  $ew_r \leftarrow (w_r + w_p) \times (w_r + r_r) / \text{maximum}$ 
5   |  $er_r \leftarrow (r_r + r_p) \times (w_r + r_r) / \text{maximum}$ 
6   |  $ew_p \leftarrow (w_r + w_p) \times (w_p + r_p) / \text{maximum}$ 
7   |  $er_p \leftarrow (r_r + r_p) \times (w_p + r_p) / \text{maximum}$ 
8   |  $x \leftarrow (|w_r - ew_r|)^2 / ew_r + (|r_r - er_r|)^2 / er_r$ 
9   |  $\quad + (|w_p - ew_p|)^2 / ew_p + (|r_p - er_p|)^2 / er_p$ 
10  |  $pValue \leftarrow \text{chiSquaredProbability}(x, 1)$ 

```

5.5 FTDD

Sendo o mais simples dos 3 métodos propostos para funcionarem com base no teste exato de Fisher, o FTDD realiza a detecção de mudanças de conceitos, exclusivamente, através desse teste estatístico.

Apesar do teste exato de Fisher ser, geralmente, indicado apenas para a análise de pequenas amostras, a sua implementação – descrita na seção 5.1 – permite que o FTDD seja aplicado em conjuntos de dados de quaisquer tamanhos. Isso porque, para a aplicação deste trabalho, ele se utiliza de uma janela de dados antigos com tamanho reduzido para o valor proporcional ao número de erros, e acertos, contidos na janela antiga original.

O algoritmo 8 demonstra como o FTDD realiza o cálculo do valor de p . Nesse pseudo-código, é possível se verificar o método em questão realiza o teste estatístico utilizando o número de erros – e acertos – da janela recente (w_r e r_r) e os valores proporcionais ao número de predições corretas, e incorretas, contidas na janela antiga (w_p e r_p).

Algorithm 8: Cálculo do Valor de p pelo FTDD

```

1  $pValue \leftarrow \text{factorial}[w_r + w_p] / \text{factorial}[w_r] / \text{factorial}[w_p]$ 
2    $\quad \times \text{factorial}[r_r + r_p] / \text{factorial}[r_r] / \text{factorial}[r_p]$ 
3  $pValue \leftarrow pValue \times \text{constantFactorial} \times 2$ 

```

É importante ressaltar que a ordem das operações – no cálculo do valor de p

pelo FTDD – foi alterada com o objetivo de evitar valores extremos nos resultados intermediários, como medida de precaução para tornar os resultados mais precisos.

5.6 Considerações Finais

Este capítulo apresentou três novos detectores de mudanças de conceitos (formando um total de sete versões) baseados no teste exato de Fisher.

O capítulo 6 descreve os experimentos realizados com o intuito de comparar os métodos propostos com alguns dos principais detectores da atualidade.

6 EXPERIMENTOS

Este capítulo apresenta os experimentos realizados a fim de comparar os detectores propostos com alguns dos principais métodos da atualidade. Dessa forma, primeiramente é descrita a configuração dos experimentos, seguida pela apresentação das bases de dados utilizadas nos testes. Por fim, os detectores são comparados através de suas acurácias e por meio da análise de suas detecções de mudanças de conceitos.

6.1 Configuração dos Experimentos

Esta seção descreve todas as informações relevantes às configurações dos experimentos realizados a fim de avaliar os detectores propostos.

Os experimentos foram executados utilizando um computador com um processador *Core i7 4500U 1.8 GHz*, memória *RAM* de *8 GB dual channel DDR3 1600 MHz*, um disco rígido de *1 TB SATA 5400 RPM* e com o sistema operacional *Ubuntu Desktop 15.10 64 bits*.

Todos os testes foram realizados no ambiente *MOA* (BIFET et al., 2010). Cada detector foi testado junto a um classificador base. Para fins de avaliação, foram utilizados os seguintes preditores: *NB* e *HT* (ambos já implementados no *framework* utilizado).

Com o objetivo de avaliar a eficácia – e eficiência – dos métodos propostos, foram realizados testes comparativos com os seguintes detectores: *DDM*, *EDDM*, *STEPD*, *ADWIN*, *ECDD*, *EMZD1*, *EMZD2*, *FPDD1*, *FPDD2*, *FSDD1*, *FSDD2* e *FTDD*. Todos os detectores propostos foram implementados no *MOA*. Além disso, os métodos utilizados como referenciais, já foram previamente utilizados na área e/ou estão publicamente disponíveis no *MOA* através do endereço eletrônico <https://sites.google.com/site/moaextensions>.

A utilização de conjuntos de dados artificiais nos testes comparativos é muito importante devido a quantidade de mudanças de conceitos, suas posições e durações poderem ser configuradas para cobrir diferentes cenários. Com base nisso, este trabalho selecionou seis geradores de bases de dados artificiais, utilizando mudanças de conceitos abruptas e graduais, com três diferentes tamanhos. Criando um total de 36 conjuntos de dados artificiais. Em todos eles, quatro mudanças de conceitos foram distribuídas em intervalos regulares, possuindo diferentes tamanhos, de acordo com o tamanho de cada base de dados. Além disso, as mudanças abruptas foram simuladas pela junção de diferentes conceitos. Por outro lado, as mudanças graduais foram configuradas para ocorrerem sobre o intervalo de 500 instâncias, sendo geradas utilizando uma função de probabilidade, a fim de aumentar as chances de instâncias pertencentes a um novo conceito sejam selecionadas.

Em adição aos conjuntos de dados artificiais, três conhecidas bases de dados reais também foram escolhidas para complementar a avaliação dos métodos propostos. Todos os conjuntos de dados reais foram escolhidos por possuírem mudanças de conceitos ao longo de suas distribuições. Além disso, as três bases selecionadas possuem diferentes características e tamanhos, a fim de se verificar o desempenho dos detectores em diferentes cenários.

Foram utilizados os seguintes critérios nas avaliações dos detectores:

1. Comparação das acurácias;
2. Análise das detecções de mudanças de conceitos.

A fim de complementar a análise dos detectores, foi utilizada uma estatística chamada F_F (DEMSAR, 2006, p. 11–15), baseada no teste não-paramétrico de Friedman, com o intuito de verificar se existe – ou não – diferenças estatísticas entre os métodos avaliados, de acordo com os critérios anteriormente estabelecidos. Todos os testes estatísticos foram realizados utilizando 95% de intervalo de confiança. Para cada critério avaliado, essa estatística estabelece um *rank* com cada um dos detectores. Além disso, quanto maior for o resultado de um método nos critérios avaliados, melhor ele será considerado e possuirá um menor valor no *rank* estabelecido pelo teste de Friedman.

A hipótese nula do teste utilizado estabelece que todos os detectores são estatisticamente iguais. Por outro lado, a rejeição da hipótese nula indica que existe diferença estatística entre alguns métodos. Entretanto, o teste não especifica quais métodos são estatisticamente diferentes.

Dessa forma, para identificar quais métodos são estatisticamente diferentes, este trabalho utilizou o pós-teste de *Nemenyi* (DEMSAR, 2006, p. 11–12), o qual compara cada detector contra todos os outros utilizando uma **Diferença Crítica (DC)** como referência. A **DC** é graficamente representada por uma barra e os métodos que são conectados por ela não são estatisticamente diferentes. Assim, quando vários métodos são comparados, os resultados dos pós-testes são representados através de um diagrama simples, o qual possui uma linha horizontal contendo o *rank* com todos os métodos analisados, de maneira que os melhores detectores ficam localizados à direita do gráfico *Nemenyi* (DEMSAR, 2006, p. 15–16).

6.2 Bases Artificiais

Para este trabalho, foram selecionados seis geradores de bases de dados artificiais: *Agrawal*, *LED*, *Mixed*, *Random RBF*, *SEA* e *Sine*. Cada gerador foi utilizado para criar bases com três tamanhos diferentes, tendo cada uma delas: 20.000, 50.000 e 100.000

instâncias. Para cada tamanho, foram criados duas versões de bases diferentes, variando de acordo com a graduação das mudanças em seus conceitos, sendo elas: abruptas e graduais. Dessa forma, ao total, este trabalho utilizou 36 versões de bases de dados artificiais para fins de comparações.

Os geradores *Agrawal* e *Random RBF* foram escolhidos por criarem bases de dados complexas, contendo diversos atributos e classes, sendo ideais para avaliar o desempenho dos detectores em conjunto de dados mais robustos. Por outro lado, a escolha da *Mixed* e *Sine* se deu pelo fato delas apresentarem apenas dois possíveis conceitos e serem úteis na avaliação da capacidade dos métodos em lidarem com conceitos recorrentes. Por fim, os geradores *LED* e *SEA* foram selecionados por criarem conjuntos de dados com ruídos, sendo eficazes na avaliação dos detectores neste tipo de cenário.

6.2.1 Agrawal

Originalmente proposto por [Agrawal, Imielinski e Swami \(1993\)](#), este gerador combina um conjunto de atributos referentes a uma série informações pessoais de alguns indivíduos, com o intuito de realizar uma análise de crédito.

Cada indivíduo é representado por um conjunto de 10 atributos, sendo eles: salário, comissão, idade, escolaridade, possui carro, código postal, valor do imóvel, quantidade de anos que vive em casa própria, e o valor do empréstimo desejado. Desses, 3 são categóricos e 7 não.

De acordo com a combinação dos atributos, os indivíduos podem ser classificados em dois grupos: A ou B.

Uma vez que a classificação pode ser realizada com até dez funções diferentes, o rótulo de cada instância poderá variar de acordo com a função utilizada.

6.2.2 LED

Este gerador, proposto por [Bifet et al. \(2010, p. 1603\)](#), produz uma base com 24 atributos categóricos. Desses, 17 são irrelevantes. Ademais, o conjunto de dados possui 10 possíveis classes, também categóricas.

O problema representado por este gerador refere-se a tarefa de predição do dígito apresentando por um visor LED com 7 segmentos.

Cada atributo do conjunto de dados gerado, por padrão, possui 10% de chance de inversão (a fim de representar o ruído). Além disso, as mudanças de conceitos são simuladas através de modificações nas posições dos atributos.

6.2.3 Mixed

O gerador *Mixed* (GAMA et al., 2004, p. 6) cria um conjunto de dados composto por 4 atributos, sendo 2 deles booleanos (v e w) e 2 numéricos (x e y). De acordo com os valores de seus atributos, as instâncias são classificadas em positivas ou negativas. Para que uma instância seja rotulada como positiva, pelo menos duas das seguintes condições devem ser satisfeitas: os valores de v e w devem ser verdadeiros e $y < 0,5 + 0,3\sin(3\pi x)$.

Para realizar a simulação de mudanças de conceitos, este gerador efetua a inversão dos rótulos, onde, para que uma instância seja considerada positiva, pelo menos duas das seguintes condições devem ser satisfeitas: os valores de v e w deverão ser falsos e $y \geq 0,5 + 0,3\sin(3\pi x)$.

6.2.4 Random RBF

Os conjuntos de dados gerados pelo *Random RBF* Bifet et al. (2009, p. 141–142), são criados com n centroides. Cada centroide tem seu centro, rótulo e peso, definido aleatoriamente. Além disso, cada instância criada por este gerador possui m atributos que variam de acordo com:

1. A posição de cada centro;
2. Um número gerado aleatoriamente, indo de -1 a +1, o qual será multiplicado por um valor baseado no desvio padrão de cada centro e na distribuição Gaussiana.

A simulação das mudanças de conceitos é realizada através da modificação da posição dos centroides.

6.2.5 SEA

Proposto por Street e Kim (2001), este gerador cria bases de dados com 3 atributos numéricos – sendo 2 importantes e 1 utilizado apenas como ruído – e 2 classes. Os valores de cada atributo são gerados aleatoriamente com números que vão de 0 à 10.

Os conjuntos de dados, criados por este gerador, utilizam um limiar (θ) com 4 possíveis valores: 7, 8, 9 e 9,5. Uma determinada instância irá pertencer à classe 1, caso $f_1 + f_2 \leq \theta$, onde: f_1 e f_2 são o primeiro e o segundo atributo de cada exemplo.

As mudanças de conceitos, nas bases geradas pelo *SEA*, são produzidas através da atribuição de diferentes valores à θ .

6.2.6 Sine

As bases de dados criadas pelo *Sine* (GAMA et al., 2004, p. 5) possuem dois atributos relevantes (x e y). Cada um dos atributos possui valores uniformemente distribuídos no intervalo de 0 à 1.

Este gerador cria conjuntos de dados com dois possíveis conceitos. São eles: *Sine1* e *Sine2*. No primeiro contexto, todos os pontos abaixo da curva $y = \sin(x)$ são classificados como positivos. Já no segundo contexto, a inequação $y < 0,5 + 0,3\sin(3\pi x)$ deve ser verdadeira para que uma determinada instância seja considerada como positiva.

Nos dois contextos citados, as mudanças de conceitos são simuladas pela inversão das condições ou pela alternância entre os contextos *Sine1* e *Sine2*.

6.3 Bases Reais

Além dos geradores de conjuntos de dados artificiais, este trabalho também utilizou 3 bases de dados reais para fins comparativos. São elas: *Coverttype*, *Pokerhand* e *Sensor*.

A escolha da base *Coverttype* se deu, especialmente, com o objetivo de avaliar o desempenho dos detectores na presença de mudanças de conceitos graduais. Por outro lado, *Pokerhand* e *Sensor* foram escolhidas por serem bases mais complexas, possuindo diversos atributos e classes, sendo recomendadas na avaliação dos métodos em bases de dados mais robustas.

6.3.1 Coverttype

Contendo dados do Sistemas de Informações de Recursos da Região 2 do *US Forest Service (USFS)*, este conjunto de dados possui atributos com informações que correspondem a uma área florestal de 30×30 metros.

A *Coverttype* (BIFET et al., 2009, p. 142) possui 581.012 instâncias. Cada instância possui 54 atributos, podendo ser numéricos ou categóricos. Assim, a tarefa de classificação é prever o tipo de cobertura florestal com base nas diversas variáveis cartográficas.

A versão utilizada, para este trabalho, foi ordenada por elevação de atributos (IENCO et al., 2013, p. 87). Dessa forma, esta base induz mudanças de conceitos graduais nas distribuições das classes. Isso ocorre porque, dependendo da elevação, alguns tipos de vegetações desaparecem, enquanto outras começam a aparecer.

6.3.2 Pokerhand

O problema representado por esta base de dados, consiste na identificação do valor das 5 cartas numa mão de um jogo de *Poker*. Dessa forma, cada instância contém

5 atributos numéricos e 5 categóricos. Além disso, as classes são determinadas por 10 possíveis rótulos categóricos.

Para este trabalho, foi utilizada a versão original da *Pokerhand* (BIFET et al., 2009, p. 142), com 1.000.000 de instâncias e cartas não ordenadas.

6.3.3 Sensor

A base de dados *Sensor* (ZHU, 2010), tem os seus dados coletados a partir de 54 sensores do *Intel Berkeley Research Lab*. Os sensores medem 4 informações: temperatura, umidade, luz e tensão. Além disso, o sensor *ID* representa a classe e deve ser identificado com base nos dados coletados.

Este conjunto de dados possui 2.219.803 instâncias, 5 atributos e 54 classes. Os dados foram gravados por um período de 2 meses e as leituras foram realizadas a cada 1–3 minutos.

6.4 Comparação de Acurácias

A avaliação de acurácia foi feita com base na metodologia *prequential* (GAMA; SEBASTIÃO; RODRIGUES, 2013), utilizando uma janela deslizante de dados como mecanismo de esquecimento. Dessa forma, cada instância é utilizada inicialmente para o teste e, subsequentemente, para o treinamento (sem a reutilização dos dados).

As porcentagens das acurácias médias dos detectores, testados com 40 repetições nos conjuntos de dados artificiais, foram computadas utilizando com 95% de intervalo de confiança e são demonstradas nas tabelas 5 e 6. Não obstante, as tabelas 7 e 8 referem-se as porcentagens das acurácias dos detectores nos conjuntos de dados reais.

Além disso, os gráficos 1 e 2 apresentam os resultados das comparações das acurácias dos detectores, nas bases artificiais e reais, com o teste de *Friedman* e o pós-teste *Nemenyi*.

A tabela 5 apresenta as acurácias médias dos detectores – em 40 Repetições – utilizando NB nas bases de dados artificiais. Nela, pode-se observar que o detector EMZD2 obteve as melhores acurácias médias em 16 dos 36 casos. Vale salientar, também, que esse método foi o melhor – no quesito acurácia – em todos os testes com o gerador SEA e em todas as bases, com mudanças de conceitos graduais, do *Mixed*. Em segundo lugar, o FSDD2 obteve as maiores médias, nas acurácias, em todos os conjuntos de dados – com mudanças abruptas – gerados pelos geradores *Mixed* e *Sine*, sendo o melhor em 8 das 36 bases de dados. Ademais, o FSDD1, alcançou as maiores acurácias médias em 3 conjuntos, e o FPDD2, EMZD1 e DDM, conseguiram as maiores médias nas acurácias dos preditores em apenas um dos casos. Finalmente, o ADWIN e o ECDD obtiveram as

Tabela 5 – Médias – em Porcentagem (%) – das Acurácias dos Detectores em 40 Repetições, Utilizando NB nos Conjuntos de Dados Artificiais, com 95% de Intervalo de Confiança.

Tamanho	Gradação	Métodos	Agrawal	LED	Mixed	Random RBF	SEA	Sine	
20K	Abruptas	ADWIN	64,02 (+-0,17)	63,33 (+-0,49)	90,48 (+-0,10)	30,34 (+-0,31)	85,65 (+-0,22)	86,69 (+-0,14)	
		DDM	63,20 (+-0,47)	71,21 (+-0,23)	90,38 (+-0,51)	30,76 (+-0,41)	86,04 (+-0,31)	83,58 (+-1,46)	
		EDDM	61,61 (+-0,28)	69,03 (+-0,26)	89,52 (+-0,46)	30,80 (+-0,43)	85,63 (+-0,20)	85,73 (+-0,46)	
		ECDD	62,32 (+-0,14)	68,02 (+-0,36)	89,43 (+-0,19)	31,17 (+-0,52)	84,85 (+-0,23)	86,39 (+-0,13)	
		STEPD	64,24 (+-0,19)	65,46 (+-1,32)	90,96 (+-0,15)	29,60 (+-0,33)	86,58 (+-0,22)	87,19 (+-0,13)	
		EMZD1	60,07 (+-0,81)	71,26 (+-0,21)	90,87 (+-0,15)	30,77 (+-0,49)	86,34 (+-0,25)	86,75 (+-0,15)	
		EMZD2	63,96 (+-0,31)	71,26 (+-0,19)	90,97 (+-0,13)	29,86 (+-0,30)	86,59 (+-0,21)	86,81 (+-0,21)	
		FPDD1	61,90 (+-0,31)	70,39 (+-0,42)	91,18 (+-0,11)	31,09 (+-0,40)	84,48 (+-0,17)	87,23 (+-0,15)	
		FPDD2	61,90 (+-0,31)	70,39 (+-0,42)	91,18 (+-0,11)	31,09 (+-0,40)	84,48 (+-0,17)	87,23 (+-0,15)	
		FSDD1	62,87 (+-0,35)	70,77 (+-0,29)	91,18 (+-0,11)	31,06 (+-0,40)	84,48 (+-0,17)	87,24 (+-0,15)	
		FSDD2	62,87 (+-0,35)	70,77 (+-0,29)	91,19 (+-0,12)	31,06 (+-0,40)	84,63 (+-0,21)	87,25 (+-0,15)	
		FTDD	61,94 (+-0,29)	70,33 (+-0,45)	91,18 (+-0,11)	31,07 (+-0,40)	84,48 (+-0,17)	87,24 (+-0,15)	
		ADWIN	63,02 (+-0,19)	62,38 (+-0,45)	87,12 (+-0,12)	30,40 (+-0,34)	85,46 (+-0,21)	84,04 (+-0,12)	
		DDM	62,53 (+-0,47)	70,46 (+-0,19)	87,82 (+-0,15)	30,81 (+-0,44)	86,07 (+-0,29)	84,70 (+-0,16)	
	EDDM	61,69 (+-0,31)	69,19 (+-0,24)	87,99 (+-0,15)	30,69 (+-0,40)	85,64 (+-0,22)	84,74 (+-0,14)		
	ECDD	61,79 (+-0,14)	66,98 (+-0,38)	86,87 (+-0,15)	31,17 (+-0,52)	84,69 (+-0,22)	84,19 (+-0,12)		
	STEPD	63,25 (+-0,20)	64,04 (+-1,15)	87,47 (+-0,13)	29,68 (+-0,32)	86,34 (+-0,22)	84,42 (+-0,16)		
	EMZD1	59,74 (+-0,73)	70,20 (+-0,20)	87,82 (+-0,14)	30,78 (+-0,50)	86,24 (+-0,25)	84,69 (+-0,16)		
	EMZD2	63,73 (+-0,22)	70,19 (+-0,18)	88,00 (+-0,13)	29,83 (+-0,31)	86,47 (+-0,22)	84,91 (+-0,17)		
	FPDD1	60,97 (+-0,41)	67,28 (+-0,83)	87,63 (+-0,13)	31,13 (+-0,40)	84,44 (+-0,16)	84,80 (+-0,16)		
	FPDD2	60,97 (+-0,41)	67,28 (+-0,83)	87,65 (+-0,13)	31,13 (+-0,40)	84,44 (+-0,16)	84,82 (+-0,16)		
	FSDD1	61,60 (+-0,31)	68,66 (+-0,60)	87,66 (+-0,14)	31,10 (+-0,40)	84,44 (+-0,16)	84,74 (+-0,15)		
	FSDD2	61,60 (+-0,31)	68,65 (+-0,60)	87,55 (+-0,13)	31,10 (+-0,40)	84,57 (+-0,20)	84,65 (+-0,17)		
	FTDD	61,06 (+-0,31)	67,50 (+-0,78)	87,63 (+-0,13)	31,12 (+-0,40)	84,44 (+-0,16)	84,77 (+-0,16)		
	50K	Abruptas	ADWIN	65,48 (+-0,11)	64,66 (+-0,44)	91,44 (+-0,09)	30,54 (+-0,29)	87,06 (+-0,14)	87,12 (+-0,10)
			DDM	63,70 (+-0,54)	71,93 (+-0,55)	90,91 (+-0,73)	30,95 (+-0,43)	86,58 (+-0,31)	83,57 (+-1,31)
			EDDM	62,75 (+-0,22)	70,07 (+-0,16)	90,13 (+-0,45)	31,08 (+-0,37)	86,31 (+-0,12)	85,58 (+-0,51)
			ECDD	62,73 (+-0,11)	68,75 (+-0,29)	89,80 (+-0,12)	31,22 (+-0,56)	84,98 (+-0,15)	86,42 (+-0,09)
STEPD			65,06 (+-0,12)	68,55 (+-0,70)	91,40 (+-0,12)	29,21 (+-0,25)	87,11 (+-0,14)	87,26 (+-0,10)	
EMZD1			61,17 (+-0,76)	72,68 (+-0,16)	91,43 (+-0,10)	31,03 (+-0,49)	86,94 (+-0,27)	86,74 (+-0,22)	
EMZD2			65,25 (+-0,20)	72,72 (+-0,15)	91,53 (+-0,09)	30,16 (+-0,26)	87,33 (+-0,14)	86,90 (+-0,28)	
FPDD1			63,47 (+-0,43)	72,25 (+-0,18)	91,71 (+-0,08)	31,07 (+-0,41)	84,72 (+-0,14)	87,39 (+-0,10)	
FPDD2			63,47 (+-0,43)	72,25 (+-0,18)	91,71 (+-0,08)	31,07 (+-0,41)	84,72 (+-0,14)	87,39 (+-0,10)	
FSDD1			65,17 (+-0,24)	72,26 (+-0,18)	91,71 (+-0,08)	31,01 (+-0,40)	84,72 (+-0,14)	87,39 (+-0,10)	
FSDD2			65,17 (+-0,24)	72,26 (+-0,18)	91,72 (+-0,08)	31,01 (+-0,40)	85,14 (+-0,28)	87,40 (+-0,10)	
FTDD			63,54 (+-0,42)	72,23 (+-0,19)	91,71 (+-0,08)	31,03 (+-0,42)	84,72 (+-0,14)	87,39 (+-0,10)	
ADWIN			65,15 (+-0,11)	64,44 (+-0,42)	90,06 (+-0,08)	30,62 (+-0,31)	86,92 (+-0,15)	86,06 (+-0,08)	
DDM			63,88 (+-0,48)	72,33 (+-0,22)	90,39 (+-0,10)	30,90 (+-0,41)	86,76 (+-0,26)	86,30 (+-0,20)	
EDDM		62,78 (+-0,22)	70,25 (+-0,17)	90,19 (+-0,09)	31,08 (+-0,37)	86,35 (+-0,16)	86,02 (+-0,14)		
ECDD		62,47 (+-0,09)	68,38 (+-0,29)	88,80 (+-0,12)	31,21 (+-0,57)	84,90 (+-0,15)	85,60 (+-0,11)		
STEPD		64,70 (+-0,12)	68,01 (+-0,77)	90,08 (+-0,10)	29,25 (+-0,25)	87,03 (+-0,15)	86,23 (+-0,10)		
EMZD1		61,26 (+-0,81)	72,41 (+-0,15)	90,39 (+-0,08)	30,92 (+-0,50)	87,01 (+-0,22)	86,39 (+-0,16)		
EMZD2		65,06 (+-0,26)	72,43 (+-0,15)	90,45 (+-0,08)	30,17 (+-0,27)	87,27 (+-0,18)	86,53 (+-0,16)		
FPDD1		62,89 (+-0,36)	71,64 (+-0,16)	90,41 (+-0,09)	31,01 (+-0,40)	84,68 (+-0,12)	86,56 (+-0,09)		
FPDD2		62,89 (+-0,36)	71,64 (+-0,16)	90,41 (+-0,09)	31,01 (+-0,40)	84,68 (+-0,13)	86,61 (+-0,10)		
FSDD1		64,26 (+-0,33)	71,66 (+-0,16)	90,42 (+-0,09)	30,96 (+-0,40)	84,68 (+-0,13)	86,58 (+-0,09)		
FSDD2		64,26 (+-0,33)	71,66 (+-0,16)	90,35 (+-0,09)	30,96 (+-0,40)	85,13 (+-0,28)	86,58 (+-0,09)		
FTDD		62,89 (+-0,37)	71,63 (+-0,16)	90,41 (+-0,09)	30,98 (+-0,41)	84,68 (+-0,12)	86,56 (+-0,09)		
100K		Abruptas	ADWIN	65,97 (+-0,08)	65,15 (+-0,45)	91,75 (+-0,05)	30,61 (+-0,24)	87,58 (+-0,09)	87,29 (+-0,07)
			DDM	64,02 (+-0,57)	72,58 (+-0,31)	90,90 (+-0,88)	31,32 (+-0,38)	86,56 (+-0,31)	82,82 (+-1,64)
			EDDM	63,30 (+-0,16)	70,46 (+-0,15)	90,16 (+-0,77)	31,56 (+-0,31)	86,61 (+-0,10)	85,85 (+-0,40)
			ECDD	62,87 (+-0,07)	69,00 (+-0,19)	89,83 (+-0,08)	31,43 (+-0,48)	85,06 (+-0,10)	86,43 (+-0,08)
	STEPD		65,40 (+-0,07)	69,32 (+-0,50)	91,53 (+-0,07)	29,12 (+-0,16)	87,27 (+-0,10)	87,28 (+-0,07)	
	EMZD1		61,60 (+-0,73)	73,17 (+-0,17)	91,68 (+-0,06)	31,49 (+-0,38)	87,41 (+-0,20)	86,52 (+-0,31)	
	EMZD2		65,33 (+-0,39)	73,27 (+-0,11)	91,74 (+-0,06)	30,73 (+-0,26)	87,67 (+-0,10)	87,03 (+-0,14)	
	FPDD1		65,03 (+-0,39)	72,87 (+-0,15)	91,90 (+-0,05)	31,54 (+-0,37)	85,01 (+-0,22)	87,43 (+-0,08)	
	FPDD2		65,03 (+-0,39)	72,87 (+-0,15)	91,90 (+-0,05)	31,54 (+-0,37)	85,01 (+-0,22)	87,43 (+-0,08)	
	FSDD1		65,87 (+-0,13)	72,88 (+-0,17)	91,90 (+-0,05)	31,63 (+-0,36)	85,01 (+-0,22)	87,43 (+-0,08)	
	FSDD2		65,87 (+-0,13)	72,88 (+-0,17)	91,91 (+-0,05)	31,63 (+-0,36)	85,78 (+-0,28)	87,44 (+-0,08)	
	FTDD		65,03 (+-0,39)	72,88 (+-0,16)	91,90 (+-0,05)	31,54 (+-0,37)	85,01 (+-0,22)	87,43 (+-0,08)	
	ADWIN		65,81 (+-0,08)	65,46 (+-0,37)	91,05 (+-0,05)	30,62 (+-0,23)	87,52 (+-0,09)	86,74 (+-0,07)	
	DDM		63,96 (+-0,56)	72,47 (+-0,39)	91,23 (+-0,05)	31,30 (+-0,37)	86,77 (+-0,32)	86,38 (+-0,56)	
	EDDM	63,34 (+-0,18)	70,53 (+-0,13)	90,61 (+-0,08)	31,57 (+-0,28)	86,66 (+-0,10)	86,16 (+-0,15)		
	ECDD	62,75 (+-0,07)	68,81 (+-0,20)	89,32 (+-0,08)	31,44 (+-0,48)	85,03 (+-0,10)	86,00 (+-0,08)		
	STEPD	65,16 (+-0,08)	69,05 (+-0,52)	90,87 (+-0,07)	29,13 (+-0,17)	87,23 (+-0,10)	86,73 (+-0,06)		
	EMZD1	61,42 (+-0,74)	72,95 (+-0,24)	91,22 (+-0,06)	31,38 (+-0,40)	87,35 (+-0,23)	86,64 (+-0,24)		
	EMZD2	65,49 (+-0,27)	73,15 (+-0,10)	91,27 (+-0,05)	30,73 (+-0,24)	87,58 (+-0,12)	86,90 (+-0,15)		
	FPDD1	64,42 (+-0,41)	72,41 (+-0,14)	91,23 (+-0,06)	31,60 (+-0,36)	84,95 (+-0,21)	87,03 (+-0,07)		
	FPDD2	64,42 (+-0,41)	72,41 (+-0,14)	91,23 (+-0,06)	31,60 (+-0,36)	84,95 (+-0,21)	87,04 (+-0,07)		
	FSDD1	65,61 (+-0,11)	72,56 (+-0,15)	91,24 (+-0,05)	31,66 (+-0,35)	84,96 (+-0,21)	87,05 (+-0,08)		
	FSDD2	65,61 (+-0,11)	72,56 (+-0,15)	91,20 (+-0,05)	31,66 (+-0,35)	85,64 (+-0,28)	87,03 (+-0,07)		
	FTDD	64,42 (+-0,41)	72,43 (+-0,13)	91,23 (+-0,06)	31,57 (+-0,36)	84,95 (+-0,21)	87,04 (+-0,07)		

Tabela 6 – Médias – em Porcentagem (%) – das Acurácias dos Detectores em 40 Repetições, Utilizando HT nos Conjuntos de Dados Artificiais, com 95% de Intervalo de Confiança.

Tamanho	Gradação	Métodos	Agrawal	LED	Mixed	Random RBF	SEA	Sine		
20K	Abruptas	ADWIN	64,28 (+-0,18)	62,99 (+-0,48)	90,14 (+-0,10)	31,97 (+-0,32)	85,48 (+-0,19)	88,68 (+-0,12)		
		DDM	65,09 (+-1,04)	71,20 (+-0,23)	89,22 (+-0,45)	31,77 (+-0,39)	85,55 (+-0,23)	89,32 (+-0,12)		
		EDDM	64,81 (+-0,49)	69,01 (+-0,25)	89,26 (+-0,33)	32,16 (+-0,38)	85,46 (+-0,22)	87,22 (+-0,17)		
		ECDD	63,69 (+-0,35)	67,98 (+-0,36)	89,37 (+-0,18)	31,15 (+-0,52)	84,81 (+-0,22)	86,85 (+-0,19)		
		STEPD	64,91 (+-0,27)	65,07 (+-1,31)	90,66 (+-0,13)	31,28 (+-0,33)	86,13 (+-0,18)	89,25 (+-0,18)		
		EMZD1	63,18 (+-1,10)	71,26 (+-0,21)	90,46 (+-0,14)	30,95 (+-0,50)	85,99 (+-0,23)	88,55 (+-0,18)		
		EMZD2	67,21 (+-0,40)	71,25 (+-0,19)	90,55 (+-0,12)	31,39 (+-0,32)	86,18 (+-0,17)	89,09 (+-0,22)		
		FPDD1	64,00 (+-0,73)	70,33 (+-0,40)	90,66 (+-0,13)	32,54 (+-0,39)	84,31 (+-0,15)	89,90 (+-0,11)		
		FPDD2	64,00 (+-0,73)	70,33 (+-0,40)	90,66 (+-0,13)	32,54 (+-0,39)	84,31 (+-0,15)	89,91 (+-0,11)		
		FSDD1	64,90 (+-0,60)	70,66 (+-0,29)	90,66 (+-0,13)	32,50 (+-0,40)	84,31 (+-0,15)	89,90 (+-0,11)		
		FSDD2	64,90 (+-0,60)	70,65 (+-0,29)	90,68 (+-0,13)	32,50 (+-0,40)	84,46 (+-0,16)	89,89 (+-0,11)		
		FTDD	64,12 (+-0,69)	70,30 (+-0,41)	90,66 (+-0,13)	32,53 (+-0,39)	84,31 (+-0,15)	89,90 (+-0,11)		
		Graduais	ADWIN	63,35 (+-0,19)	62,30 (+-0,57)	86,91 (+-0,13)	31,96 (+-0,30)	85,28 (+-0,21)	85,45 (+-0,11)	
			DDM	64,33 (+-0,91)	70,45 (+-0,19)	87,30 (+-0,15)	31,79 (+-0,46)	85,59 (+-0,21)	86,69 (+-0,12)	
	EDDM		63,88 (+-0,62)	69,17 (+-0,24)	87,58 (+-0,13)	32,21 (+-0,36)	85,49 (+-0,21)	86,53 (+-0,13)		
	ECDD		62,88 (+-0,26)	66,94 (+-0,38)	86,70 (+-0,14)	31,15 (+-0,52)	84,65 (+-0,22)	84,95 (+-0,17)		
	STEPD		63,94 (+-0,20)	64,36 (+-1,11)	87,26 (+-0,13)	31,09 (+-0,29)	85,88 (+-0,17)	85,96 (+-0,12)		
	EMZD1		62,14 (+-1,07)	70,17 (+-0,19)	87,28 (+-0,14)	30,94 (+-0,50)	85,83 (+-0,24)	86,51 (+-0,13)		
	EMZD2		65,91 (+-0,52)	70,16 (+-0,18)	87,43 (+-0,15)	31,50 (+-0,28)	86,03 (+-0,20)	86,87 (+-0,14)		
	FPDD1		60,94 (+-0,80)	67,43 (+-0,77)	87,14 (+-0,13)	32,55 (+-0,41)	84,32 (+-0,15)	86,80 (+-0,11)		
	FPDD2		60,94 (+-0,80)	67,43 (+-0,77)	87,14 (+-0,14)	32,55 (+-0,41)	84,32 (+-0,15)	86,84 (+-0,11)		
	FSDD1		63,21 (+-0,65)	68,43 (+-0,60)	87,16 (+-0,12)	32,53 (+-0,40)	84,32 (+-0,15)	86,80 (+-0,11)		
	FSDD2		63,21 (+-0,65)	68,44 (+-0,60)	87,06 (+-0,12)	32,53 (+-0,40)	84,47 (+-0,20)	86,66 (+-0,10)		
	FTDD		61,46 (+-0,72)	67,60 (+-0,71)	87,13 (+-0,13)	32,54 (+-0,41)	84,32 (+-0,15)	86,80 (+-0,11)		
	50K		Abruptas	ADWIN	65,78 (+-0,13)	64,00 (+-0,44)	91,48 (+-0,10)	32,19 (+-0,24)	85,96 (+-0,12)	89,84 (+-0,09)
				DDM	68,26 (+-1,62)	72,12 (+-0,39)	91,45 (+-0,29)	32,46 (+-0,31)	85,39 (+-0,22)	91,05 (+-0,12)
		EDDM		67,28 (+-0,78)	69,98 (+-0,21)	90,34 (+-0,14)	32,46 (+-0,29)	85,61 (+-0,13)	88,91 (+-0,19)	
		ECDD		64,68 (+-0,55)	68,71 (+-0,29)	89,76 (+-0,12)	33,17 (+-0,29)	84,96 (+-0,16)	87,09 (+-0,12)	
STEPD		66,22 (+-0,23)		67,89 (+-0,93)	91,11 (+-0,09)	31,07 (+-0,19)	86,47 (+-0,13)	90,40 (+-0,17)		
EMZD1		64,90 (+-1,62)		72,69 (+-0,16)	91,05 (+-0,15)	32,54 (+-0,31)	86,02 (+-0,20)	90,23 (+-0,20)		
EMZD2		70,98 (+-0,73)		72,72 (+-0,15)	91,22 (+-0,14)	31,76 (+-0,21)	86,29 (+-0,14)	90,92 (+-0,15)		
FPDD1		66,72 (+-0,72)		72,23 (+-0,18)	92,09 (+-0,11)	32,61 (+-0,34)	84,71 (+-0,11)	91,52 (+-0,12)		
FPDD2		66,72 (+-0,72)		72,23 (+-0,18)	92,09 (+-0,11)	32,61 (+-0,34)	84,72 (+-0,11)	91,52 (+-0,12)		
FSDD1		68,10 (+-0,67)		72,21 (+-0,18)	92,09 (+-0,11)	32,58 (+-0,33)	84,71 (+-0,11)	91,52 (+-0,12)		
FSDD2		68,09 (+-0,67)		72,21 (+-0,18)	92,02 (+-0,13)	32,58 (+-0,33)	85,10 (+-0,17)	91,52 (+-0,12)		
FTDD		66,94 (+-0,82)		72,20 (+-0,19)	92,09 (+-0,11)	32,61 (+-0,35)	84,71 (+-0,11)	91,52 (+-0,12)		
Graduais		ADWIN		65,31 (+-0,14)	63,91 (+-0,51)	89,97 (+-0,09)	32,09 (+-0,24)	85,87 (+-0,13)	88,50 (+-0,08)	
		DDM		68,60 (+-1,37)	72,35 (+-0,20)	90,88 (+-0,08)	32,49 (+-0,30)	85,48 (+-0,21)	90,27 (+-0,09)	
		EDDM	67,12 (+-0,80)	70,22 (+-0,17)	90,35 (+-0,10)	32,54 (+-0,31)	85,54 (+-0,12)	88,96 (+-0,21)		
		ECDD	64,54 (+-0,60)	68,36 (+-0,30)	88,69 (+-0,12)	33,15 (+-0,28)	84,86 (+-0,16)	86,37 (+-0,13)		
		STEPD	65,75 (+-0,21)	67,89 (+-0,80)	89,79 (+-0,09)	31,06 (+-0,22)	86,40 (+-0,14)	89,17 (+-0,16)		
		EMZD1	65,42 (+-1,66)	72,40 (+-0,15)	90,43 (+-0,11)	32,52 (+-0,32)	85,94 (+-0,20)	90,02 (+-0,15)		
		EMZD2	70,87 (+-0,31)	72,41 (+-0,15)	90,63 (+-0,09)	31,73 (+-0,23)	86,20 (+-0,15)	90,31 (+-0,09)		
		FPDD1	65,99 (+-0,74)	71,62 (+-0,16)	90,74 (+-0,07)	32,63 (+-0,36)	84,71 (+-0,12)	90,26 (+-0,09)		
		FPDD2	65,99 (+-0,74)	71,62 (+-0,16)	90,74 (+-0,07)	32,63 (+-0,36)	84,71 (+-0,12)	90,31 (+-0,09)		
		FSDD1	67,18 (+-0,54)	71,63 (+-0,15)	90,77 (+-0,07)	32,62 (+-0,34)	84,71 (+-0,12)	90,27 (+-0,09)		
		FSDD2	67,19 (+-0,55)	71,64 (+-0,15)	90,64 (+-0,08)	32,62 (+-0,34)	85,06 (+-0,16)	90,21 (+-0,10)		
		FTDD	65,83 (+-0,73)	71,61 (+-0,15)	90,75 (+-0,07)	32,63 (+-0,36)	84,71 (+-0,12)	90,26 (+-0,09)		
		100K	Abruptas	ADWIN	66,41 (+-0,11)	64,67 (+-0,41)	91,79 (+-0,08)	32,43 (+-0,21)	86,48 (+-0,08)	90,28 (+-0,06)
				DDM	71,28 (+-1,73)	72,68 (+-0,24)	92,79 (+-0,13)	33,60 (+-0,21)	85,58 (+-0,23)	92,30 (+-0,08)
EDDM				69,53 (+-0,89)	70,34 (+-0,15)	91,47 (+-0,10)	33,28 (+-0,25)	85,83 (+-0,11)	90,49 (+-0,16)	
ECDD				66,26 (+-0,59)	68,97 (+-0,19)	89,78 (+-0,08)	34,83 (+-0,19)	85,03 (+-0,11)	87,17 (+-0,09)	
STEPD	66,97 (+-0,23)			68,84 (+-0,64)	91,26 (+-0,08)	31,18 (+-0,15)	86,60 (+-0,07)	90,97 (+-0,13)		
EMZD1	66,06 (+-1,93)			73,07 (+-0,34)	91,56 (+-0,18)	33,83 (+-0,23)	86,08 (+-0,20)	91,85 (+-0,13)		
EMZD2	74,44 (+-0,24)			73,26 (+-0,11)	91,83 (+-0,22)	32,41 (+-0,24)	86,27 (+-0,16)	92,24 (+-0,08)		
FPDD1	69,71 (+-0,77)			72,87 (+-0,15)	93,17 (+-0,06)	33,35 (+-0,27)	85,00 (+-0,15)	92,59 (+-0,08)		
FPDD2	69,71 (+-0,77)			72,87 (+-0,15)	93,15 (+-0,06)	33,35 (+-0,27)	85,00 (+-0,15)	92,59 (+-0,08)		
FSDD1	71,24 (+-0,81)			72,85 (+-0,16)	93,17 (+-0,06)	33,40 (+-0,25)	85,00 (+-0,15)	92,59 (+-0,08)		
FSDD2	71,24 (+-0,81)			72,84 (+-0,16)	93,10 (+-0,07)	33,40 (+-0,25)	85,65 (+-0,23)	92,57 (+-0,09)		
FTDD	69,93 (+-0,91)			72,88 (+-0,15)	93,17 (+-0,06)	33,33 (+-0,27)	85,00 (+-0,15)	92,59 (+-0,08)		
Graduais	ADWIN			66,26 (+-0,10)	64,49 (+-0,42)	91,02 (+-0,08)	32,49 (+-0,19)	86,41 (+-0,08)	89,59 (+-0,06)	
	DDM			71,85 (+-1,51)	72,64 (+-0,27)	92,46 (+-0,07)	33,63 (+-0,19)	85,63 (+-0,23)	91,97 (+-0,07)	
	EDDM		69,08 (+-1,06)	70,41 (+-0,14)	91,54 (+-0,09)	33,30 (+-0,24)	85,83 (+-0,11)	90,63 (+-0,16)		
	ECDD		65,78 (+-0,53)	68,78 (+-0,20)	89,24 (+-0,08)	34,82 (+-0,17)	85,00 (+-0,10)	86,82 (+-0,11)		
	STEPD		66,56 (+-0,22)	68,54 (+-0,62)	90,66 (+-0,08)	31,16 (+-0,15)	86,56 (+-0,08)	90,44 (+-0,13)		
	EMZD1		64,60 (+-1,60)	72,99 (+-0,31)	92,01 (+-0,12)	33,91 (+-0,20)	86,09 (+-0,20)	91,83 (+-0,08)		
	EMZD2		74,16 (+-0,29)	73,15 (+-0,10)	92,19 (+-0,10)	32,33 (+-0,19)	86,32 (+-0,14)	91,98 (+-0,08)		
	FPDD1		69,40 (+-0,89)	72,45 (+-0,13)	92,45 (+-0,05)	33,32 (+-0,26)	84,89 (+-0,12)	91,92 (+-0,07)		
	FPDD2		69,41 (+-0,89)	72,45 (+-0,13)	92,45 (+-0,05)	33,32 (+-0,26)	84,88 (+-0,12)	91,97 (+-0,07)		
	FSDD1		70,18 (+-0,70)	72,55 (+-0,15)	92,46 (+-0,05)	33,43 (+-0,23)	84,89 (+-0,12)	91,93 (+-0,07)		
	FSDD2		70,22 (+-0,71)	72,55 (+-0,15)	92,34 (+-0,06)	33,43 (+-0,23)	85,60 (+-0,20)	91,88 (+-0,08)		
	FTDD		69,53 (+-0,86)	72,47 (+-0,12)	92,46 (+-0,05)	33,32 (+-0,26)	84,89 (+-0,12)	91,92 (+-0,07)		

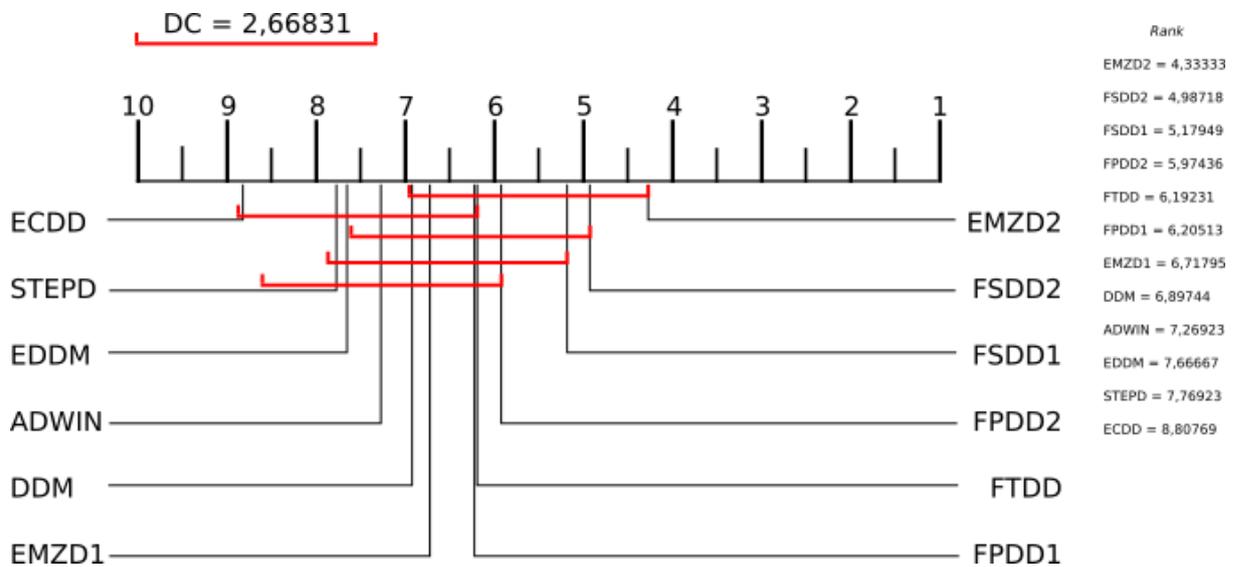


Figura 1 – Comparação das Acurácias dos Detectores, Utilizando o NB, através do Teste de *Friedman* e o Pós-Teste *Nemenyi*, com 95% de intervalo de confiança.

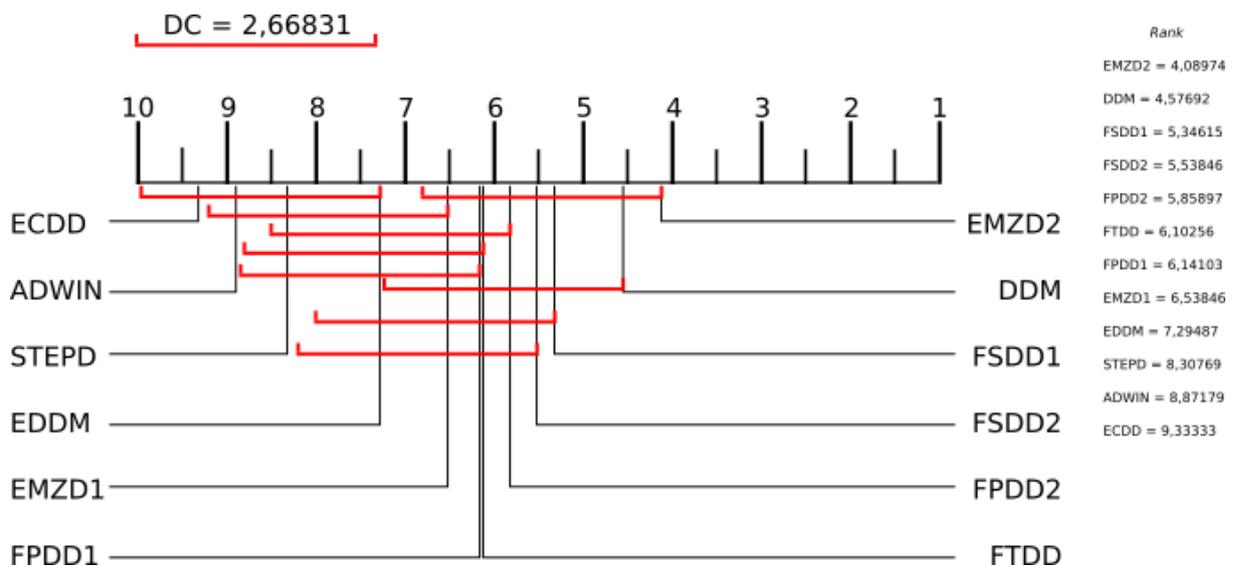


Figura 2 – Comparação das Acurácias dos Detectores, Utilizando o HT, através do Teste de *Friedman* e o Pós-Teste *Nemenyi*, com 95% de intervalo de confiança.

Tabela 7 – Acurácias – em Porcentagem (%) – dos Detectores, Utilizando NB nos Conjuntos de Dados Reais.

Métodos	Coverttype	Pokerhand	Sensor
ADWIN	67,73	50,11	34,40
DDM	67,14	50,11	80,85
EDDM	66,41	50,12	84,71
ECDD	67,39	50,11	14,90
STEPD	67,62	48,84	18,71
EMZD1	64,69	49,96	86,24
EMZD2	66,86	50,11	85,58
FPDD1	68,29	50,05	32,79
FPDD2	68,29	50,05	32,86
FSDD1	68,17	50,00	27,29
FSDD2	68,14	50,00	27,79
FTDD	68,38	50,05	30,41

Tabela 8 – Acurácias – em Porcentagem (%) – dos Detectores, Utilizando HT nos Conjuntos de Dados Reais.

Métodos	Coverttype	Pokerhand	Sensor
ADWIN	71,26	50,45	34,33
DDM	75,64	51,81	86,50
EDDM	76,39	51,13	84,76
ECDD	70,05	51,00	76,34
STEPD	70,75	49,76	18,63
EMZD1	74,43	50,87	86,29
EMZD2	75,36	53,83	85,86
FPDD1	72,04	51,44	31,25
FPDD2	72,04	51,44	31,51
FSDD1	71,36	51,82	27,43
FSDD2	71,35	51,82	27,61
FTDD	72,78	51,43	30,43

melhores acurácias médias em 4 bases, entretanto, especialmente nos conjuntos do gerador *LED*, alcançaram resultados consideravelmente inferiores aos demais.

Na tabela 6, que se refere às acurácias médias dos detectores – em 40 Repetições – utilizando *HT* nos conjuntos de dados artificiais, observa-se que o método *EMZD2* alcançou os melhores resultados, nas médias das acurácias, em 15 dos 36 conjuntos de dados. Além disso, esse método obteve as maiores acurácias médias em todos os testes com o gerador *Agrawal* e em todas as bases graduais geradas pelo *Sine*. Em seguida, é importante notar que, embora o *DDM* tenha obtido as melhores acurácias médias em apenas 2 bases, ele apresentou resultados competitivos na maioria dos casos, especialmente nos geradores *LED* e *Sine*. Semelhantemente ao *DDM*, as duas versões do *FSDD*, apesar de terem alcançado as maiores médias nas acurácias em apenas alguns poucos casos, obtiveram bons resultados nos conjuntos gerados pelo *Mixed*, *Random RBF* e *Sine*. Não obstante,

o **FPDD2**, **FPDD1** e o **FTDD**, foram os melhores detectores, em relação as acurácias dos classificadores, em – respectivamente – 9, 8 e 7 bases de dados. Também vale salientar que, em somente uma das bases, o **EMZD1** e **EDDM** alcançaram as melhores médias das acurácias. Por fim, apesar do **STEPD** ter conseguido as maiores acurácias médias em 4 bases, tal detector obteve resultados consideravelmente inferiores aos demais, especialmente nas bases geradas pelo *LED*.

A tabela 7 demonstra as acurácias dos detectores, utilizando **NB** nos conjuntos de dados reais. Neste cenário, o método **FTDD** obteve a melhor acurácia na *Coverttype*. Não obstante, todos os demais métodos baseados no teste exato de Fisher alcançaram resultados bem próximos ao **FTDD**. Já na base *Pokerhand*, todos os detectores – com exceção do **STEPD** – obtiveram acurácias semelhantes. Nesse conjunto, o **EDDM** foi o melhor. Finalmente, os resultados mais discrepantes surgiram no conjunto de dados *Sensor*. Nessa base, o detector **EMZD1** alcançou o melhor resultado, com 86,24% de acurácia, sendo acompanhado pelo **EMZD2** e o **EDDM**, com 85,58% e 84,71% de acurácias, respectivamente. Por outro lado, o pior resultado foi 14,90%, com o **ECDD**.

As acurácias dos detectores utilizando **HT** nos conjuntos de dados reais são apresentadas na tabela 8. Neste contexto, os resultados dos métodos nas bases *Coverttype* e *Pokerhand* sofreram maiores variações do que com o **NB**. Na base de dados *Coverttype*, o **EDDM** alcançou a melhor acurácia, com uma taxa de 76,39%. Após ele, o **DDM** e **EMZD2** obtiveram 75,64% e 75,36% de acurácias, respectivamente. Não obstante, nessa mesma base, o **ECDD** obteve apenas 70,05%. Na *Coverttype*, o valor mais discrepante refere-se a acurácia alcançada pelo **EMZD2**, sendo o melhor resultado, com 53,83%. Por fim, na base *Sensor*, o **DDM** e o **STEPD** alcançaram a maior – e pior – acurácia, com 86,50% e 18,63% de acurácias, respectivamente.

A figura 1 apresenta o *rank*, e o seu respectivo gráfico, referentes ao teste de *Friedman* – aplicado juntamente com pós-teste *Nemenyi* – a fim de compararem as acurácias dos detectores funcionando em conjunto com o classificador **NB**. Ao analisar as posições de cada método no *rank*, observa-se que todos os métodos propostos obtiveram melhores resultados do que os demais analisados. Não obstante, o **EMZD2** foi o melhor posicionado no *rank* e superou, estatisticamente, o **ADWIN**, **EDDM**, **STEPD**, **ECDD**. Em segundo lugar, o **FSDD2** foi estatisticamente melhor do que o **EDDM**, **STEPD** e o **ECDD**. Além deles, o **FSDD1** e o **FPDD2** foram estatisticamente superiores ao **ECDD**.

A figura 2 contém o *rank* – e o gráfico – relativos ao teste de *Friedman*, com o pós-teste *Nemenyi*, aplicados na comparação das acurácias dos detectores com o classificador **HT**. Neste cenário, os métodos propostos também ficaram nas melhores posições do *rank*, com a exceção do **DDM**, que obteve o segundo lugar. Estando nas primeiras colocações, o **EMZD2** e o **DDM** alcançaram superioridade estatística sobre o **EDDM**, **STEPD**, **ADWIN** e **ECDD**. Em seguida, o **FSDD1** e **FSDD2** foram estatisticamente

superiores ao [STEPD](#), [ADWIN](#) e [ECDD](#). Não obstante, o [FPDD2](#), [FTDD](#) e [FPDD1](#) foram melhores, estatisticamente, do que o [ADWIN](#) e [ECDD](#). Por fim, o [EMZD1](#) superou estatisticamente apenas o [ECDD](#).

Numa análise mais profunda, é possível observar que o [EMZD2](#), quando comparado com o [EDDM](#) (método no qual foi inspirado), obteve melhores acurácias em 59 dos 72 testes com bases de dados artificiais, sendo superior em 30 casos com o [NB](#) e em 29 casos com o [HT](#). Ademais, apesar de ter obtido a melhor acurácia geral em apenas uma das bases de dados reais, o [EMZD2](#) superou o [EDDM](#) em quatro dos seis testes com os conjuntos de dados reais. Não obstante, é importante salientar que o [EMZD2](#) foi estatisticamente superior ao [EDDM](#) em ambos os classificadores.

Por outro lado, o [EMZD1](#), comparado com o detector no qual foi inspirado, obteve maiores acurácias em 24 dos 36 conjuntos de dados artificiais, com o classificador [NB](#). Já com o [HT](#), o [EMZD1](#) foi superior ao [DDM](#) em apenas 16 das 36 bases de dados artificiais. Além disso, quanto à comparação das acurácias alcançadas nas bases de dados reais, o [EMZD1](#) foi superior ao [DDM](#) em apenas um dos casos. No mais, o teste de Friedman não constatou diferenças estatísticas entre o [EMZD1](#) e o [DDM](#), com nenhum dos dois classificadores testados.

Já os detectores baseados no teste exato de Fisher ([FPDD](#), [FSDD](#) e [FTDD](#)), no que diz respeito a acurácia, alcançaram resultados semelhantes com ambos os classificadores testados (sendo as duas versões do [FSDD](#) levemente superiores na maioria dos casos). Todas as versões dos métodos propostos que utilizam o teste exato de Fisher na detecção das mudanças de conceitos superaram o [STEPD](#), método no qual foram inspirados, em 24 das 36 bases de dados artificiais com o [NB](#). Nos testes com esse classificador, o [STEPD](#) conseguiu melhores acurácias apenas para as bases geradas pelos geradores *Agrawal* e *SEA*. Já para os testes realizados com o [HT](#), os métodos baseados no teste exato de Fisher obtiveram maiores acurácias do que o [STEPD](#) em 27 dos 36 conjuntos de dados artificiais. Além disso, o [STEPD](#) não conseguiu superar os detectores propostos em nenhum dos testes com as base de dados reais. Não obstante, quanto à comparação estatística realizada através do teste de Friedman, as duas versões do [FSDD](#) superaram estatisticamente o [STEPD](#) nos testes com o [HT](#). Já nos testes como classificador [NB](#), apenas o [FSDD2](#) foi estatisticamente superior ao [STEPD](#).

6.5 Análise das Detecções de Mudanças de Conceitos

Uma perspectiva diferente da avaliação do desempenho dos detectores pode ser obtida através da análise das suas detecções de mudanças de conceitos. Assim, para cada configuração das bases de dados artificiais com mudanças de conceitos abruptas, cada um

dos detectores foi testado com 40 repetições, afim de se mensurar as seguintes métricas: ¹

1. Distância Média entre os Reais Pontos de Mudanças de Conceitos e as Detecções (DM);
2. Quantidade de Falsos Positivos (FP);
3. Quantidade de Falsos Negativos (FN);
4. Quantidade de Verdadeiros Positivos (VP);
5. Quantidade de Verdadeiros Negativos (VN);
6. *Precision*;
7. *Recall*;
8. *Matthews Correlation Coefficient* (MCC).

A DM mensura o número médio de instâncias referentes ao atraso dos métodos em detectarem modificações nas distribuições dos dados analisados (PESARANGHADER; VIKTOR, 2016, p. 102). Além disso, FP e FN representam, respectivamente, o número de pontos incorretamente considerados como mudanças de conceitos e o número de mudanças de conceitos não identificadas (DU; SONG; JIA, 2014, p. 348). Por outro lado, VP e VN, respectivamente, indicam a quantidade de mudanças de conceitos corretamente reportadas e o número de pontos, corretamente, não identificados como mudanças de conceitos (WANG; ABRAHAM, 2015, p. 2).

As métricas *Precision* e *Recall*, definidas por Perry, Kent e Berry (1955), são representadas, respectivamente, pelas equações 6.1 e 6.2. Tais medidas, têm o objetivo de medir a relevância dos falsos positivos, e falsos negativos, relacionados com as detecções de mudanças de conceitos realizadas corretamente. Assim, a *Precision* retorna a proporção das mudanças detectadas que são, de fato, mudanças existentes. Já a *Recall* retorna a proporção das mudanças de conceitos existentes que foram corretamente detectadas por cada método.

$$Precision = VP / (VP + FP) \quad (6.1)$$

$$Recall = VP / (VP + FN) \quad (6.2)$$

A ultima métrica, MCC (MATTHEWS, 1975), foi utilizada porque as medidas *Precision* e *Recall* são influenciadas pelo desequilíbrio entre falsos positivos e falsos

¹ Um conjunto de tabelas com o detalhamento de tais métricas é apresentado no apêndice deste trabalho.

negativos, sendo a *Precision* mais afetada pela quantidade de falsos positivos e a *Recall* pelos falsos negativos. Por outro lado, a métrica *MCC* é mais balanceada e toma por base a matriz de confusão com as quantidades de *VP*, *VN*, *FP*, e *FN*, retornando, assim, valores no intervalo entre -1 e 1, conforme o definido pela equação 6.3.

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP) \times (VP + FN) \times (VN + FP) \times (VN + FN)}} \quad (6.3)$$

Para efeitos da avaliação do desempenho dos detectores, através das métricas anteriormente listadas, é importante citar que são considerados os melhores métodos, aqueles que possuem os menores valores em *DM*, *FP* e *FN*, e os maiores valores nas demais medidas.

Tabela 9 – Análise das Detecções de Mudanças de Conceitos Abruptas, Em 40 Repetições, Utilizando NB nos Conjuntos de Dados Artificiais.

Tamanho	Métodos	DM	FP	FN	VP	VN	<i>Precision</i>	<i>Recall</i>	MCC
20K	ADWIN	49,24	533,17	97,17	62,83	799306,83	0,105425	0,392708	0,203183
	DDM	54,46	120,00	114,33	45,67	799720,00	0,275654	0,285417	0,280346
	EDDM	48,74	619,83	122,33	37,67	799220,17	0,057288	0,235417	0,115785
	ECDD	21,94	861,00	67,67	92,33	798979,00	0,096853	0,577083	0,236092
	STEPD	33,01	434,00	65,33	94,67	799406,00	0,179067	0,591667	0,325273
	EMZD1	41,84	325,50	91,33	68,67	799514,50	0,174207	0,429167	0,273211
	EMZD2	40,13	333,00	93,67	66,33	799507,00	0,166110	0,414583	0,262200
	FPDD1	34,17	33,67	90,17	69,83	799806,33	0,674718	0,436458	0,542594
	FPDD2	34,37	33,67	90,17	69,83	799806,33	0,674718	0,436458	0,542594
	FSDD1	33,90	50,50	84,67	75,33	799789,50	0,598675	0,470833	0,530838
	FSDD2	39,35	53,17	84,50	75,50	799786,83	0,586788	0,471875	0,526119
	FTDD	34,07	37,00	90,00	70,00	799803,00	0,654206	0,437500	0,534916
50K	ADWIN	82,77	1022,83	74,83	85,17	1998817,17	0,076865	0,532292	0,202127
	DDM	114,16	122,83	105,50	54,50	1999717,17	0,307331	0,340625	0,323493
	EDDM	77,53	799,50	126,33	33,67	1999040,50	0,040408	0,210417	0,092049
	ECDD	34,55	2169,17	53,17	106,83	1997670,83	0,046939	0,667708	0,176842
	STEPD	51,27	876,83	49,83	110,17	1998963,17	0,111618	0,688542	0,277105
	EMZD1	86,00	489,33	74,50	85,50	1999350,67	0,148739	0,534375	0,281826
	EMZD2	87,10	519,00	60,00	100,00	1999321,00	0,161551	0,625000	0,317661
	FPDD1	49,41	36,67	80,33	79,67	1999803,33	0,684814	0,497917	0,583907
	FPDD2	49,51	36,67	80,33	79,67	1999803,33	0,684814	0,497917	0,583907
	FSDD1	48,82	56,17	75,00	85,00	1999783,83	0,602125	0,531250	0,565546
	FSDD2	58,89	61,00	74,50	85,50	1999779,00	0,583618	0,534375	0,558420
	FTDD	49,18	39,67	80,50	79,50	1999800,33	0,667133	0,496875	0,575716
100K	ADWIN	155,51	1729,33	47,83	112,17	3998110,67	0,060910	0,701042	0,206558
	DDM	224,79	124,83	99,83	60,17	3999715,17	0,325225	0,376042	0,349684
	EDDM	208,41	879,50	127,83	32,17	3998960,50	0,035283	0,201042	0,084138
	ECDD	47,60	4429,00	50,17	109,83	3995411,00	0,024199	0,686458	0,128747
	STEPD	86,20	1675,00	41,17	118,83	3998165,00	0,066245	0,742708	0,221733
	EMZD1	140,15	751,83	68,00	92,00	3999088,17	0,109026	0,575000	0,250319
	EMZD2	154,36	697,00	47,33	112,67	3999143,00	0,139152	0,704167	0,312975
	FPDD1	138,94	45,83	72,67	87,33	3999794,17	0,655820	0,545833	0,598290
	FPDD2	139,09	45,67	72,67	87,33	3999794,33	0,656642	0,545833	0,598665
	FSDD1	139,63	72,17	66,50	93,50	3999767,83	0,564386	0,584375	0,574276
	FSDD2	124,07	78,50	65,00	95,00	3999761,50	0,547550	0,593750	0,570165
	FTDD	140,66	47,17	72,67	87,33	3999792,83	0,649318	0,545833	0,595317

Tabela 10 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT nos Conjuntos de Dados Artificiais.

Tamanho	Métodos	DM	FP	FN	VP	VN	<i>Precision</i>	<i>Recall</i>	MCC
20K	ADWIN	36,14	559,00	96,17	63,83	799281,00	0,102490	0,398958	0,201914
	DDM	44,58	111,50	104,00	56,00	799728,50	0,334330	0,350000	0,341940
	EDDM	50,16	559,83	121,83	38,17	799280,17	0,063820	0,238542	0,123060
	ECDD	21,83	865,67	64,83	95,17	798974,33	0,099050	0,594792	0,242397
	STEPD	31,51	503,00	67,50	92,50	799337,00	0,155330	0,578125	0,299424
	EMZD1	36,21	389,00	84,83	75,17	799451,00	0,161940	0,469792	0,275588
	EMZD2	40,68	290,50	85,83	74,17	799549,50	0,203380	0,463542	0,306843
	FPDD1	33,91	44,33	89,17	70,83	799795,67	0,615050	0,442708	0,521732
	FPDD2	34,22	44,17	89,17	70,83	799795,83	0,61594	0,442708	0,522110
	FSDD1	33,66	60,33	84,33	75,67	799779,67	0,556370	0,472917	0,512861
	FSDD2	33,12	67,33	84,33	75,67	799772,67	0,529140	0,472917	0,500144
FTDD	34,20	46,17	89,00	71,00	799793,83	0,605970	0,443750	0,518475	
50K	ADWIN	101,99	1161,00	75,33	84,67	1998679,00	0,067970	0,529167	0,189493
	DDM	106,43	135,17	93,83	66,17	1999704,83	0,328640	0,413542	0,368600
	EDDM	98,62	759,33	128,33	31,67	1999080,67	0,040030	0,197917	0,088856
	ECDD	33,62	2165,83	53,83	106,17	1997674,17	0,046730	0,663542	0,175891
	STEPD	49,51	1029,17	52,83	107,17	1998810,83	0,094310	0,669792	0,251200
	EMZD1	88,71	640,67	66,33	93,67	1999199,33	0,127550	0,585417	0,273151
	EMZD2	91,79	464,83	58,83	101,17	1999375,17	0,178740	0,632292	0,336088
	FPDD1	47,49	54,67	77,67	82,33	1999785,33	0,600970	0,514583	0,556070
	FPDD2	47,71	54,33	77,67	82,33	1999785,67	0,60244	0,514583	0,556748
	FSDD1	39,15	76,50	71,00	89,00	1999763,50	0,537760	0,556250	0,546892
	FSDD2	54,64	86,50	70,50	89,50	1999753,50	0,508520	0,559375	0,533304
FTDD	46,14	56,00	77,17	82,83	1999784,00	0,596640	0,517708	0,555741	
100K	ADWIN	155,24	2149,50	54,17	105,83	3997690,50	0,046930	0,661458	0,176083
	DDM	174,92	131,00	85,67	74,33	3999709,00	0,362010	0,464583	0,410077
	EDDM	225,44	839,67	129,00	31,00	3999000,33	0,035600	0,193750	0,082974
	ECDD	46,73	4370,33	48,67	111,33	3995469,67	0,024840	0,695833	0,131340
	STEPD	80,53	1954,33	47,50	112,50	3997885,67	0,054430	0,703125	0,195543
	EMZD1	148,66	963,00	58,67	101,33	3998877,00	0,095210	0,633333	0,245492
	EMZD2	148,72	620,50	42,67	117,33	3999219,50	0,159020	0,733333	0,341446
	FPDD1	124,12	88,67	66,67	93,33	3999751,33	0,512820	0,583333	0,546922
	FPDD2	124,41	89,83	66,67	93,33	3999750,17	0,509550	0,583333	0,545178
	FSDD1	119,78	111,67	61,83	98,17	3999728,33	0,467830	0,613542	0,535735
	FSDD2	95,87	124,67	59,50	100,50	3999715,33	0,446340	0,628125	0,529463
FTDD	119,94	87,33	66,00	94,00	3999752,67	0,51838	0,587500	0,551841	

As tabelas 9 e 10, apresentam as médias alcançadas por cada detector em **DM**, **FP**, **FN**, **VP** e **VN**, sobre os testes com 40 repetições no conjunto das 7 bases de dados artificiais utilizadas, para cada tamanho de base gerado (20.000, 50.000 e 100.000 instâncias) com mudanças de conceitos abruptas. Além disso, também são adicionadas as métricas *Precision*, *Recall* e **MCC**, calculadas sobre as médias apresentadas.

Uma vez que, dentre as métricas listadas, a **MCC** é a mais balanceada no que diz respeito a combinação de **FP**, **FN**, **VP** e **VN**, este trabalho optou por utilizar esta métrica como referência para os testes estatísticos que objetivam verificar quais foram os melhores detectores no contexto da análise das detecções de mudanças de conceitos.

As figuras 3 e 4 demonstram os resultados das comparações estatísticas entre os detectores sobre os conjuntos de dados artificiais com mudanças de conceitos abruptas, tomando por base a métrica **MCC** e utilizando o teste de *Friedman* com o pós-teste

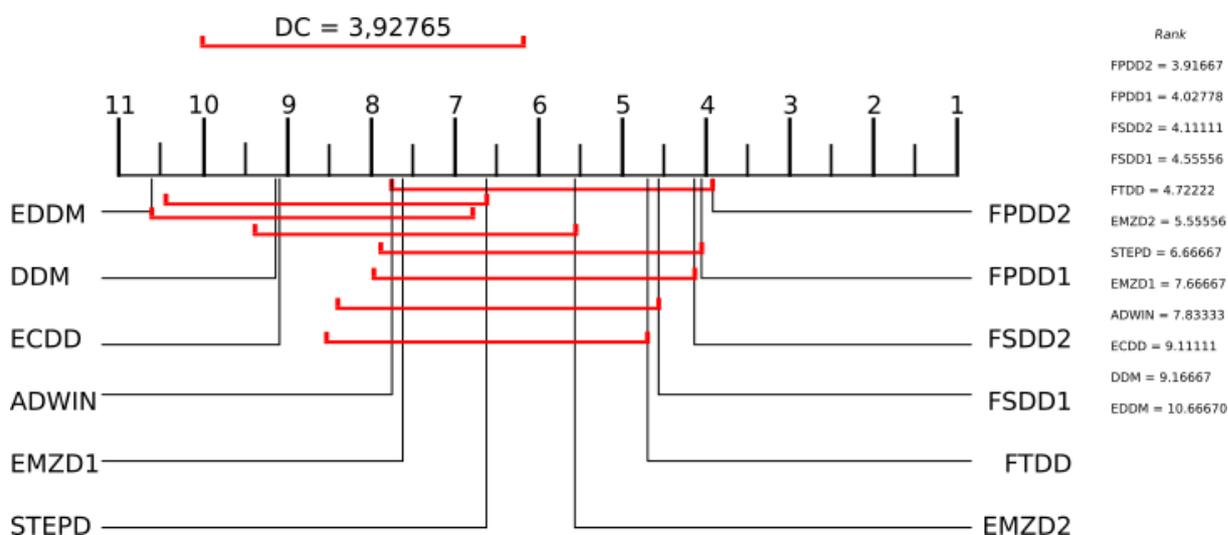


Figura 3 – Comparação do MCC dos Detectores, Utilizando o NB, através do Teste de *Friedman* e o Pós-Teste *Nemenyi*, com 95% de intervalo de confiança.

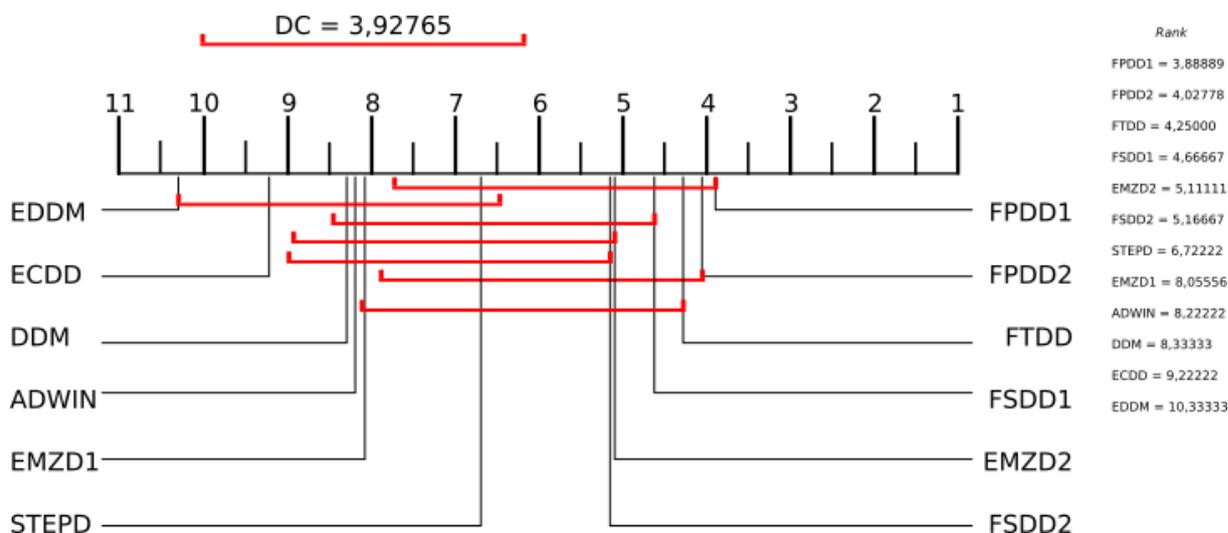


Figura 4 – Comparação do MCC dos Detectores, Utilizando o HT, através do Teste de *Friedman* e o Pós-Teste *Nemenyi*, com 95% de intervalo de confiança.

Nemenyi.

A tabela 9 apresenta a análise das detecções de mudanças de conceitos abruptas, em 40 repetições, utilizando NB nos conjuntos de dados artificiais. Nela, é possível observar que o método FPDD obteve os melhores resultados nas métricas FP, VN, Precision e MCC, sendo – nesse contexto – o FPDD2 melhor em todos os casos, seguido pelo FPDD1, que foi o melhor apenas nas bases com 20.000 e 50.000 instâncias. Não obstante, o STEP alcançou resultados superiores aos demais, para todos os casos, nas seguintes métricas: FN, VP e Recall. Por fim, o ECDD superou os demais métodos apenas na métrica DM.

Na tabela 10, que demonstra a análise das detecções de mudanças de conceitos abruptas, em 40 repetições, utilizando HT nos conjuntos de dados artificiais, observa-se

que, nas métricas **FP**, **VN**, *Precision* e **MCC**, o **FPDD2** e **FTDD** alcançaram os melhores resultados. Sendo o **FTDD** melhor nas bases com 100.000 instâncias e o **FPDD2** melhor nos demais tamanhos. Já nas métricas **FN**, **VP** e *Recall*, o **STEPD** e o **EMZD2** foram superiores aos demais. Nesse contexto, o **EMZD2** obteve resultados superiores nos conjuntos de dados com 100.000 exemplos, sendo o **STEPD** superior nas bases 20.000 e 50.000 instâncias. Finalmente, apenas na métrica **DM**, o **ECDD** foi melhor do que os demais.

A figura 3 contém o *rank* – com o seu respectivo gráfico – referentes ao teste de *Friedman*, aplicado em conjunto com o pós-teste *Nemenyi*, com o objetivo de, considerando a métrica **MCC**, realizar uma comparação estatística entre os detectores funcionando em conjunto com o classificador **NB**. Analisando as posições de cada detector no *rank*, pode se observar que, com exceção do **EMZD1**, todos os métodos propostos obtiveram os melhores resultados. Não obstante, o **FPDD2**, **FPDD1**, **FSDD2**, **FSDD1** e **FTDD** obtiveram, nessa ordem, as primeiras posições, sendo estatisticamente superiores aos métodos **ECDD**, **DDM** e **EDDM**. Em seguida, o **EMZD2** e o **STEPD**, foram estatisticamente melhores do que o **EDDM** apenas. Além disso, não houve diferenças estatísticas entre os demais detectores.

A figura 4 apresenta o *rank* – e o gráfico – relativos ao teste de *Friedman*, com o pós-teste *Nemenyi*, aplicados na comparação estatística entre os detectores, quanto aos seus desempenhos na métrica **MCC**, utilizando o classificador **HT**. Neste cenário, semelhantemente ao que ocorreu na análise utilizando o classificador **NB**, os métodos propostos também ficaram nas melhores posições do *rank*, excetuando-se novamente **EMZD1**. Estando nas primeiras colocações, o **FPDD1** e **FPDD2** obtiveram superioridade estatística sobre o **EMZD1**, **ADWIN**, **DDM**, **ECDD** e **EDDM**. Após eles, o **FTDD** foi estatisticamente melhor do que o **ADWIN**, **DDM**, **ECDD** e **EDDM**. Não obstante, os métodos **FSDD1**, **EMZD2** e **FSDD2** foram estatisticamente superiores ao **ECDD** e **EDDM**. Ademais, não houve outras diferenças estatísticas entre os métodos.

Numa análise mais detalhada, é possível verificar que o **EMZD2** superou o método no qual foi inspirado (**EDDM**) em todas as médias das métricas utilizadas (com a única exceção sendo a **DM**, nas bases de 50.000 instâncias, com o **NB**). Além disso, na comparação estatística baseada na métrica **MCC**, o **EMZD2** foi estatisticamente superior ao **EDDM** com os dois classificadores testados.

Por outro lado, o **EMZD1** superou o **DDM** (método que lhe serviu de inspiração) apenas nas médias de **DM**, **FN**, **VP** e *Recall*, sendo inferior nas demais. Entretanto, uma vez que a comparação estatística com base na **MCC** leva em conta cada resultado alcançado nessa métrica em todas as bases artificiais abruptas testadas, no teste de *Friedman*, o **EMZD1** foi melhor avaliado do que o **DDM** (com os dois classificadores testados), visto que obteve melhor **MCC** na maioria dos casos. Não obstante, quanto ao **MCC**, não houve diferença estatística entre os dois métodos com nenhum dos classificadores testados.

Além do que já foi mencionado, é importante citar que as duas versões do **EMZD** foram configuradas com altos níveis de significância (para a sinalização de *warning* e *drift*), uma vez que nos testes exploratórios realizados com o algoritmo genético proposto por Santos, Barros e Gonçalves Jr. (2015), tal configuração elevou os índices de acurácia do método. Entretanto, a utilização de níveis elevados de significância penalizou o **EMZD** com altos índices de **FP**, comprometendo também os seus desempenhos em **VN**, *Precision* e **MCC**. Não obstante, o **EMZD1** e **EMZD2** foram melhores nos índices médios de **FN**, **VP** e *Recall* do que o **DDM** e **EDDM**, comprovando um aumento significativo do poder do teste estatístico – o qual é inversamente proporcional à quantidade de falsos negativos (BANERJEE et al., 2009) – em relação às estratégias utilizadas pelos métodos nos quais foram inspirados.

Por fim, os métodos baseados no teste exato de Fisher (**FPDD**, **FSDD** e **FTDD**) superaram todos os demais detectores, incluindo o método no qual foram inspirados (**STEPD**), em todas as médias das seguintes métricas: **FP**, **VN**, *Precision* e **MCC**. Os baixos índices de **FP** – e consequentemente bons resultados em **VN**, *Precision* e **MCC** – tem relação direta com o fato do teste exato de Fisher ser conservador (CRANS; SHUSTER, 2008) e, por isso, tender a controlar o erro do tipo I (falsos positivos) com um nível inferior ao nível de significância adotado. Apesar disso, no que diz respeito a comparação estatística com base no **MCC**, o teste de Friedman não detectou diferenças entre o **STEPD** e os detectores propostos com base no teste exato de Fisher.

6.6 Gráficos de Acurácia x MCC

A visualização do desempenho dos métodos propostos, pode ser realizada através da elaboração de gráficos que contemplem as diferentes métricas analisadas. Nesse sentido, os gráficos com curvas *Receiver Operator Characteristic* (**ROC**) apresentam um balanceamento entre os benefícios (taxa de verdadeiros positivos) e custos (taxa de falsos positivos) envolvidos na tarefa de classificação (FAWCETT, 2006), além de possibilitarem a aplicação do teste estatístico *Kolmogorov – Smirnov* (**KS**) para se verificar possíveis diferenças estatísticas entre os preditores analisados (BRADLEY, 2013). Não obstante, curvas *Precision – Recall* são utilizadas com o intuito de apresentarem o balanceamento entre a precisão e sensibilidade de determinado algoritmo (DAVIS; GOADRICH, 2006).

Uma vez que a utilização de curvas **ROC** (ou *Precision – Recall*) permitiriam apenas uma análise do ponto de vista das acurácias dos classificadores ou das detecções dos métodos avaliados e, por outro lado, os detectores propostos nesta dissertação foram avaliados sob as duas óticas; este trabalho optou pela apresentação de uma série de gráficos com duas dimensões, contendo os índices de **MCC** no eixo das abcissas e as taxas de acurácia nas coordenadas, ambos seguindo o intervalo (0, 1). É importante salientar que

apesar do **MCC** possuir um intervalo $(-1, 1)$ (MATTHEWS, 1975, p. 445), os únicos valores negativos alcançados nessa métrica, pelos detectores testados neste trabalho, podem ser arredondados para zero com até três casas decimais. Dessa forma, objetivando uma melhor visualização, optou-se pela apresentação da métrica **MCC** com o intervalo de 0 a 1.

A escolha desse tipo de gráfico se deu com o objetivo de verificar o balanceamento no desempenho dos detectores quanto às acurácias e detecções. Nesse sentido, cada método é apresentado com uma série de segmentos de linha ligando os pontos de conexão entre os índices de acurácias e **MCC** alcançados sobre todos os conjuntos de dados artificiais com mudanças de conceitos abruptas testados neste trabalho. Além disso, os detectores com os pontos de conexão mais concentrados na área superior direita do gráfico apresentam resultados melhores e mais balanceados. Por fim, não foi adicionado nenhum teste estatístico utilizando as distribuições apresentadas nos gráficos, com base no fato de que todos os métodos já foram estatisticamente comparados no que diz respeito às suas acurácias e a análise de suas detecções.

As figuras 5 e 6 apresentam os gráficos de acurácia e **MCC** referentes ao desempenho dos detectores em conjunto com o **NB**. Nelas, pode-se observar que todos os métodos propostos com base no teste exato de Fisher, seguidos pelo **DDM** e **EMZD2**, obtiveram os maiores níveis de acurácia e **MCC**. Além deles, **ADWIN**, **STEPD** e **EMZD1** apresentaram gráficos com altos índices de acurácias, acompanhados de níveis medianos em **MCC**. Por fim, **EDDM** e **ECDD** conseguiram boas acurácias, porém apenas com baixos índices de **MCC**.

Nas figuras 7 e 8 que demonstram os gráficos referentes ao desempenho dos detectores com **HT**, quanto a acurácia e **MCC**, é possível verificar que – de forma semelhante aos gráficos com **NB** – os métodos inspirados no teste exato de Fisher, nesse caso seguidos apenas pelo **DDM**, alcançaram os melhores índices de balanceamento entre acurácia e **MCC**. Não obstante, **ADWIN**, **STEPD**, **EMZD1** e **EMZD2** obtiveram gráficos semelhantes, com altos níveis de acurácia, porém com o **MCC** mediano na maioria dos casos. Por último, **EDDM** e **ECDD** alcançaram bons índices de acurácia, apenas ao custo de um baixo desempenho em **MCC**.

Com base nas análises dos gráficos de acurácia e **MCC** apresentados nesta seção, é importante salientar que, dentre todos os métodos avaliados, o **FPDD**, **FSDD** e **FTDD** obtiveram os gráficos com os melhores níveis de balanceamento nas duas métricas, com ambos classificadores testados (**NB** e **HT**).

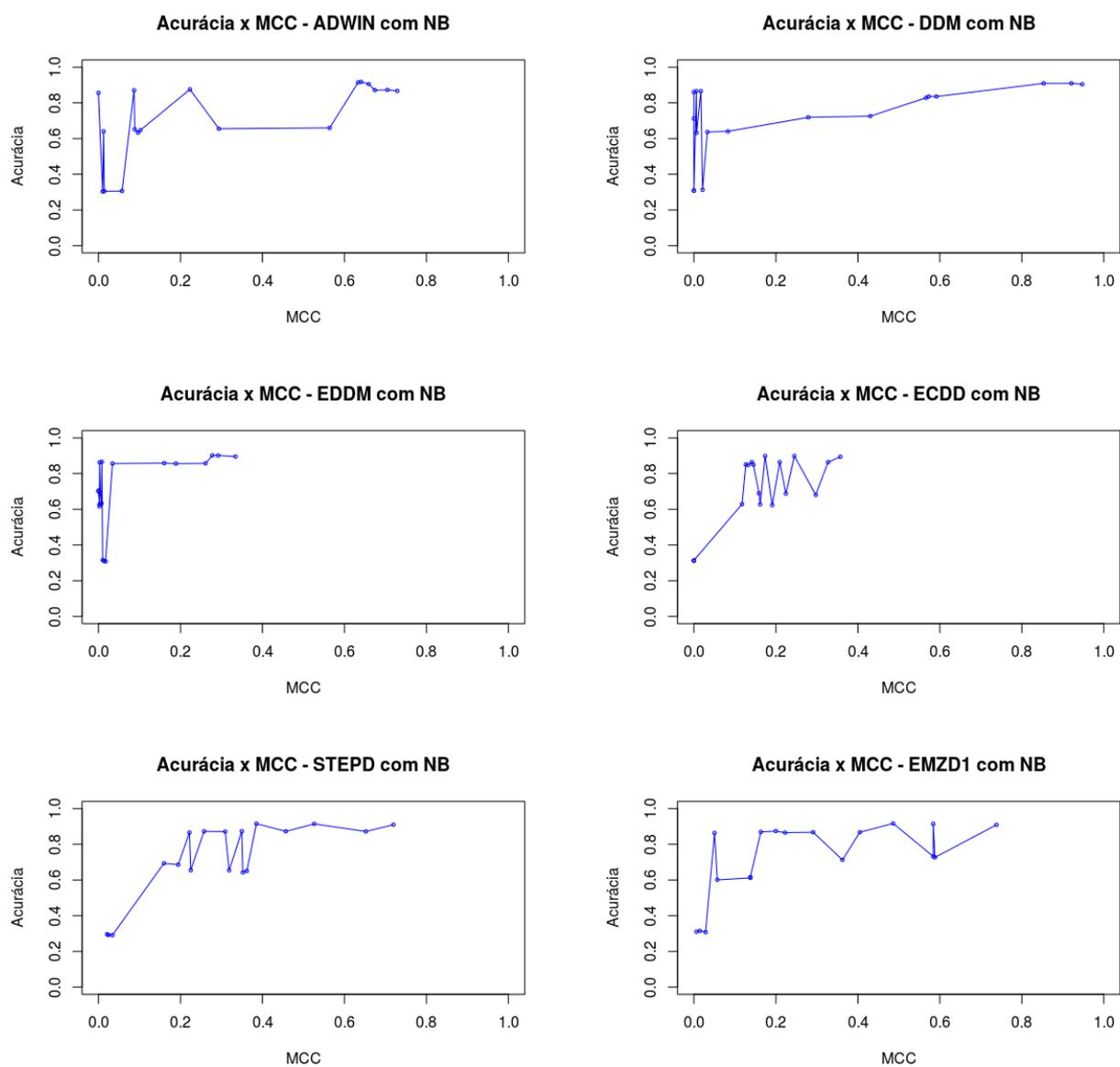


Figura 5 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando NB - Parte 1.

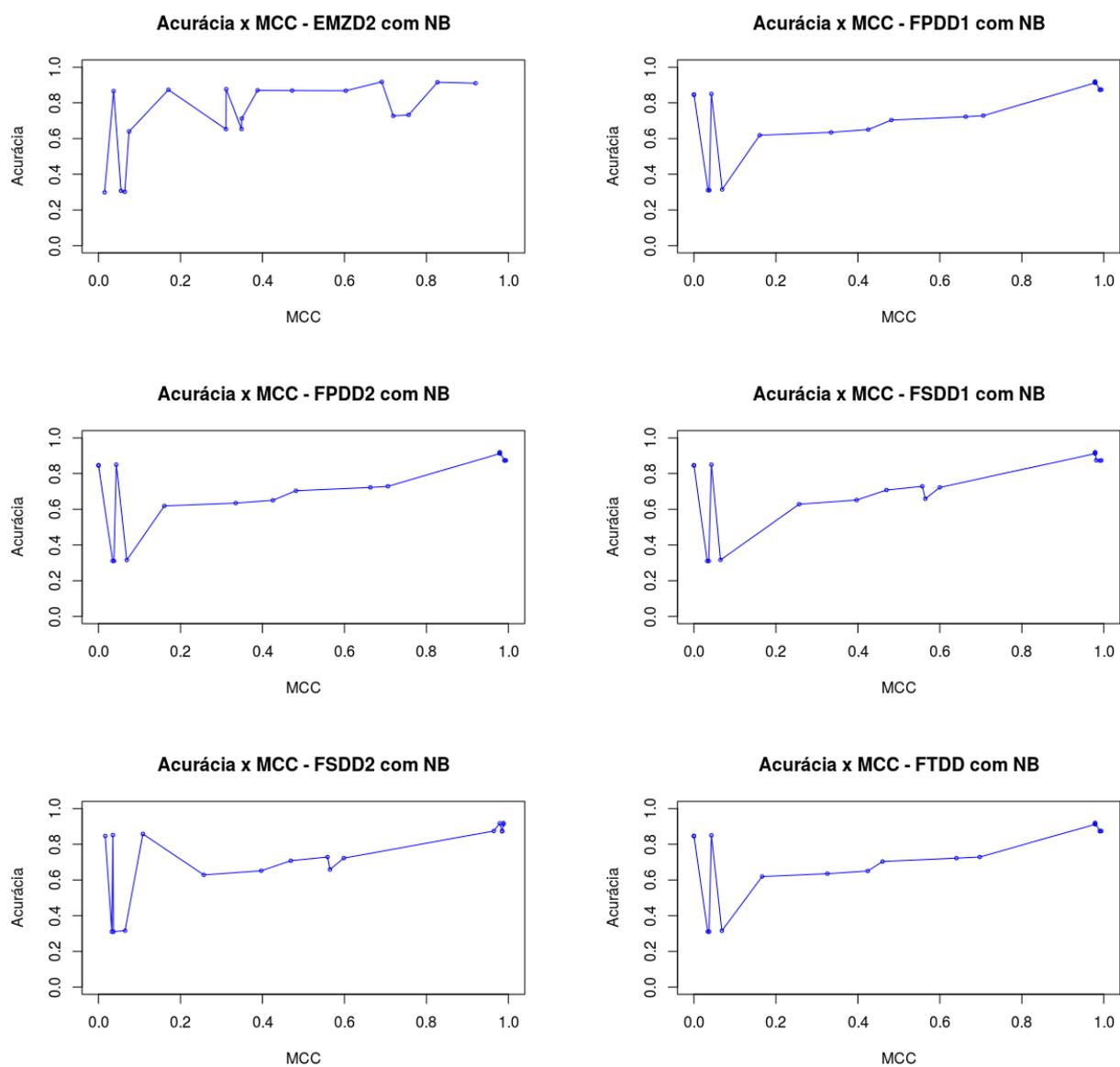


Figura 6 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando NB - Parte 2.

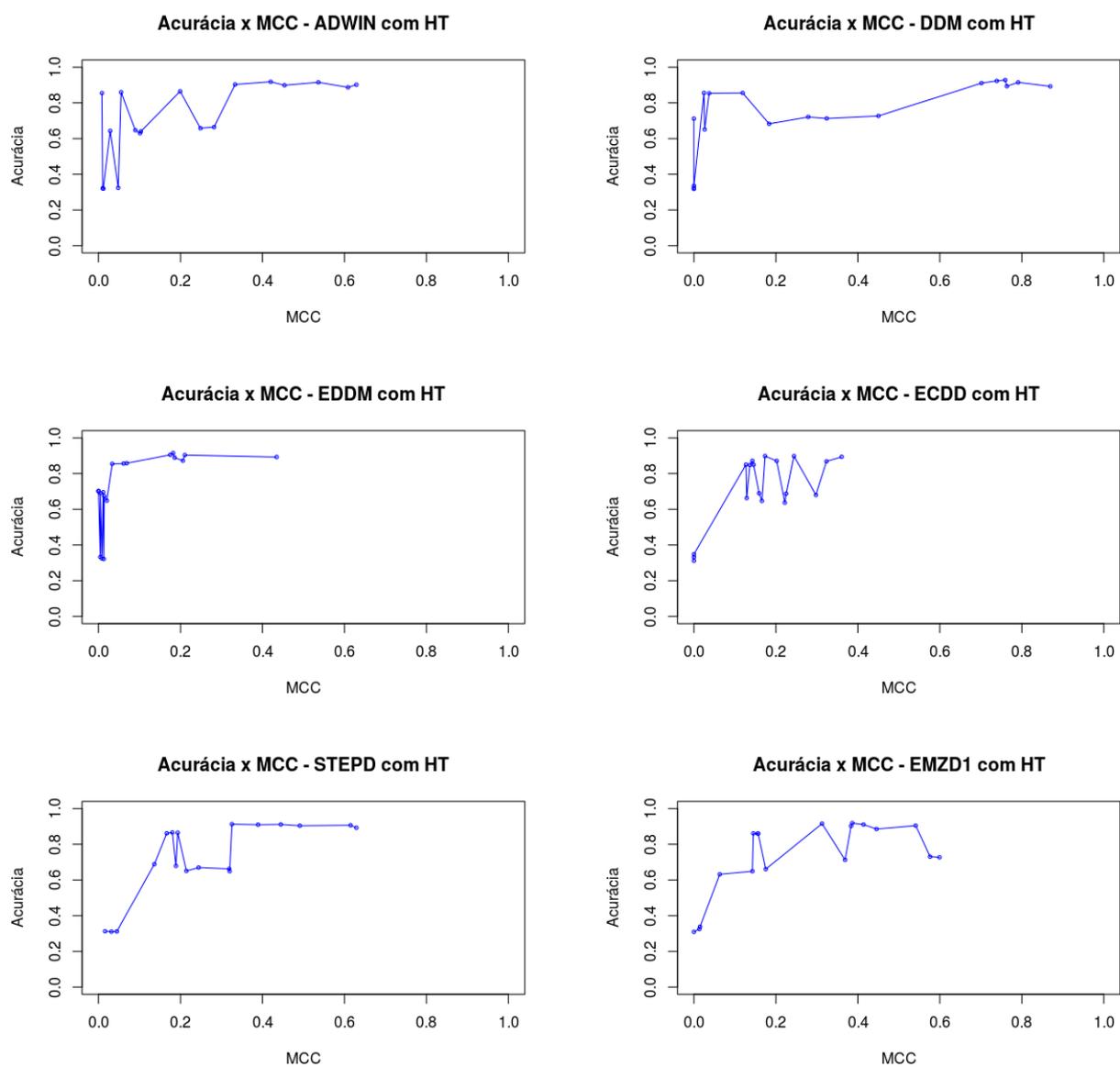


Figura 7 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando HT - Parte 1.

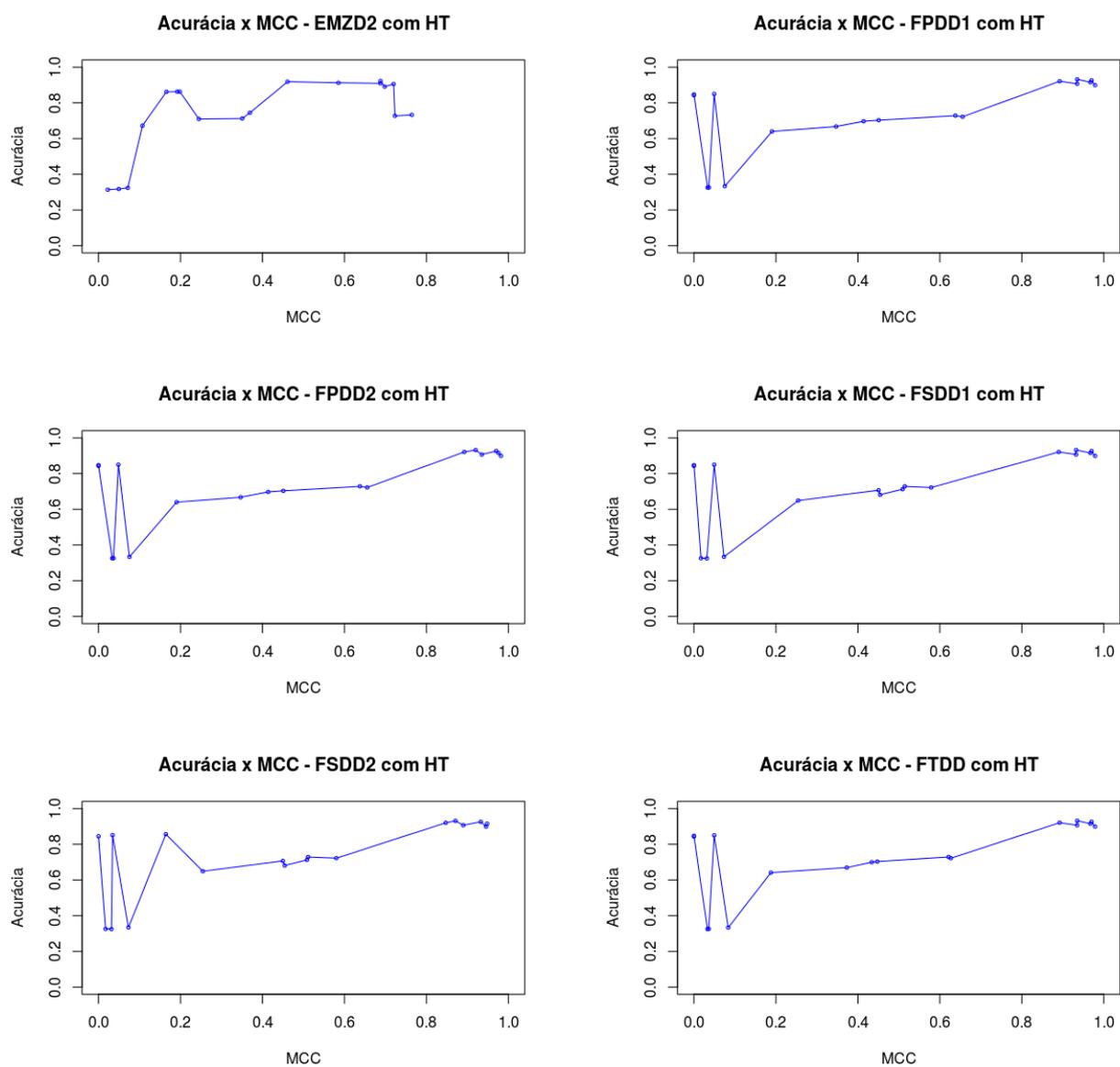


Figura 8 – Gráficos de Acurácia x MCC em Testes com Bases de Dados Artificiais com Mudanças de Conceitos Abruptas e Utilizando HT - Parte 2.

7 CONCLUSÕES

Este trabalho propôs quatro novos algoritmos de detecção de mudanças de conceitos em fluxos de dados, que formam um total de sete versões, baseadas em alguns dos principais detectores da atualidade. Os métodos que serviram de inspiração para os algoritmos propostos são: [DDM](#), [EDDM](#) e [STEPD](#).

O capítulo 4 apresentou o método [EMZD](#), o qual utiliza o teste de hipóteses entre médias de duas amostras independentes para sinalizar as mudanças de conceitos, com o intuito de melhorar a acurácia dos métodos originais, em bases de dados com mudanças abruptas e graduais, independentemente dos tamanhos de seus conceitos. Esse método foi proposto em duas versões, sendo a primeira ([EMZD1](#)) inspirada no [DDM](#) e a segunda ([EMZD2](#)) no [EDDM](#).

Não obstante, no capítulo 5, este trabalho também propôs uma eficiente implementação do teste exato de Fisher para a criação de três novos detectores de mudanças de conceitos baseados no [STEPD](#). Esses métodos foram propostos com o objetivo de melhorar a capacidade de detecção do [STEPD](#) em cenários com amostras desbalanceadas e, dessa forma, alcançarem melhores acurácias. O primeiro deles, [FPDD](#), foi proposto em duas versões ([FPDD1](#) e [FPDD2](#)) e ambas são variações do [STEPD](#) que utilizam o teste exato de Fisher apenas quando as amostras são pequenas. Em seguida, o [FSDD](#) também foi implementado em duas versões ([FSDD1](#) e [FSDD2](#)) que funcionam de forma similar ao [FPDD](#), porém para os casos das amostras suficientemente grandes e/ou balanceadas, cada uma delas utiliza o teste do qui-quadrado para homogeneidade de proporções. Por último, o [FTDD](#) foi proposto para utilizar apenas o teste exato de Fisher, independente do tamanho das amostras.

No capítulo 6, foram apresentados os experimentos comparativos envolvendo os métodos propostos e os seguintes detectores: [DDM](#), [EDDM](#), [STEPD](#), [ADWIN](#) e [ECDD](#). Todos os testes foram realizados no [MOA](#) e cada detector foi testado junto a dois classificadores ([NB](#) e [HT](#)). Para a realização desses experimentos, foram utilizados seis geradores de bases de dados artificiais: *Agrawal*, *LED*, *Mixed*, *Random RBF*, *SEA* e *Sine*. Cada gerador foi utilizado para criar bases com três tamanhos diferentes, tendo cada uma delas: 20.000, 50.000 e 100.000 instâncias. Assim, para cada tamanho, foram criadas duas versões de bases diferentes, variando de acordo com a graduação das mudanças em seus conceitos (número de instâncias necessárias para que uma mudança de conceito seja completa), sendo: abruptas (graduação 1) e graduais (graduação 500). Dessa forma, ao total, foram utilizadas 36 versões de bases de dados artificiais para fins de comparações. Além disso, este trabalho também utilizou 3 bases de dados reais para fins comparativos. Foram elas:

Coverttype, Pokerhand e Sensor.

O primeiro critério utilizado para avaliar o desempenho dos detectores testados, foi a acurácia. Nesse quesito, o [EMZD2](#) alcançou os melhores resultados na grande maioria das bases de dados artificiais. Além disso, o [EMZD2](#) obteve a melhor colocação no *rank* do teste de Friedman e foi estatisticamente superior ao [EDDM](#), [STEPD](#), [ADWIN](#) e [ECDD](#) em ambos os classificadores. Por outro lado, o [EMZD1](#) não conseguiu superar estatisticamente o detector no qual foi inspirado ([DDM](#)) com nenhum dos classificadores testados. Já os detectores baseados no teste exato de Fisher ([FPDD](#), [FSDD](#) e [FTDD](#)), quando comparados com o método que lhe serviu como base ([STEPD](#)), alcançaram melhores acurácias na maioria dos casos. Porém, quanto à comparação estatística realizada pelo teste de Friedman, apenas o [FSDD](#) foi estatisticamente melhor do que o [STEPD](#), sendo as duas versões superiores com o [HT](#) e apenas o [FSDD2](#) superior com o classificador [NB](#).

Uma outra abordagem utilizada na avaliação do desempenho dos métodos testados, foi a análise das detecções de mudanças de conceitos. Nesse contexto, foram avaliadas oito métricas ([DM](#), [FP](#), [FN](#), [VP](#), [VN](#), *Precision*, *Recall* e [MCC](#)). Dentre elas, uma vez que a [MCC](#) é a mais balanceada no que diz respeito a combinação de [FP](#), [FN](#), [VP](#) e [VN](#), ela foi utilizada como referência para identificar os métodos melhores avaliados quanto à capacidade de detecção de mudanças de conceitos. Nesse contexto, os métodos baseados no teste exato de Fisher ([FPDD](#), [FSDD](#) e [FTDD](#)), apesar de não conseguirem superar estatisticamente o método no qual foram inspirados ([STEPD](#)), obtiveram os melhores resultados em todas as médias da métrica [MCC](#), assim como em todas as médias de [FP](#), [VN](#), *Precision* e [MCC](#) (sendo o [FPDD](#) considerado o melhor dentre eles). Não obstante, o [EMZD2](#) superou o método no qual foi inspirado ([EDDM](#)) em todas as médias das métricas utilizadas (com a única exceção sendo a [DM](#), nas bases de 50.000 instâncias, com o [NB](#)), sendo também estatisticamente superior ao [EDDM](#), quanto a comparação estatística com base no [MCC](#), com os dois classificadores testados. Por outro lado, o [EMZD1](#) superou o [DDM](#) (método que lhe serviu de inspiração) apenas nas médias de [DM](#), [FN](#), [VP](#) e *Recall*, não sendo detectadas diferenças estatísticas (quanto a comparação entre [MCC](#)) entre os dois métodos.

Além dos dois critérios utilizados nas avaliações dos detectores testados, este trabalho também apresentou uma série de gráficos contendo os índices de acurácia e [MCC](#) alcançados por cada método avaliado. Com base neles, foi possível verificar que, dentre todos os detectores avaliados, o [FPDD](#), [FSDD](#) e [FTDD](#) obtiveram os gráficos com os melhores níveis de balanceamento nas duas métricas, com ambos classificadores testados ([NB](#) e [HT](#)).

Por fim, com base nas análises e testes comparativos realizados por este trabalho, verificou-se que todos os métodos propostos nesta dissertação (com exceção do [EMZD1](#))

superaram, na maioria dos casos, todos os demais métodos testados (incluindo os detectores nos quais foram inspirados), tanto nas acurácias, como nas precisões das detecções de mudanças de conceitos. Mais especificamente, dentre todos os métodos analisados, o **EMZD2** foi o melhor avaliado quanto a comparação das acurácias e as duas versões do **FPDD**, seguidas pelo **FSDD** e **FTDD**, obtiveram os melhores resultados na análise das detecções de mudanças de conceitos. Dessa forma, este trabalho indica o uso do **EMZD2** em cenários onde o principal objetivo seja apenas maximizar as acurácias dos classificadores. Por outro lado, em cenários mais conservadores, onde a precisão nas detecções das mudanças de conceitos for mais importante do que as acurácias dos preditores, esta pesquisa indica a utilização dos métodos propostos com base no teste exato de Fisher, em especial, o **FPDD**.

7.1 Publicações

Os métodos propostos nesta dissertação serviram de base para a escrita de dois artigos. O primeiro deles foi submetido para um periódico e o segundo para uma conferência, ambos internacionais e bem classificados no qualis da CAPES. São eles:

- **Drift Detection using Fisher’s Exact Test** (Danilo Rafael de Lima Cabral e Roberto Souto Maior de Barros);
- **EMZD: Equal Means Z-Test Concept Drift Detector** (Danilo Rafael de Lima Cabral e Roberto Souto Maior de Barros).

Não obstante, outros artigos motivados pelo tema da dissertação, mas que não estão diretamente ligados a este trabalho, também foram submetidos para periódicos internacionais de boa qualidade. São eles:

- **RDDM: Reactive Drift Detection Method** (Roberto S. M. Barros, Danilo R. L. Cabral, Paulo M. Gonçalves Jr. e Silas G. T. C. Santos);
- **Wilcoxon Rank Sum Test Drift Detector** (Roberto Souto Maior de Barros, Juan Isidro González Hidalgo e Danilo Rafael de Lima Cabral).

7.2 Limitações

Seguem abaixo as principais limitações deste trabalho:

- Além dos classificadores referenciados neste trabalho, existem outros preditores bem conceituados que também seriam úteis na análise experimental dos métodos

propostos. Dentre os principais, pode-se citar: *Random Forests* (SAFFARI et al., 2009), *Support Vector Machines* (LASKOV et al., 2006), *Radial Basis Function Networks* (FREEMAN; SAAD, 1997), e outros;

- Todos os métodos de detecção de mudanças de conceitos referenciados neste trabalho, foram testados apenas com a configuração padrão de parâmetros sugerida pelos seus respectivos autores.

7.3 Trabalhos Futuros

Além dos métodos propostos nesta dissertação, existem os seguintes trabalhos futuros que podem ser desenvolvidos como continuação deste trabalho.

- Propor uma nova versão do classificador NB capaz de lidar com mudanças de conceitos, visando o aprimoramento na classificação em fluxos de dados com distribuições dinâmicas;
- Propor uma máquina de comitê, sensível a mudanças de conceitos, que funcione através da combinação de testes estatísticos aplicados às tarefas de adição de novos classificadores, eliminação dos classificadores mais fracos, verificação de diversidade dentre os preditores, determinação dos votos válidos para a ponderação dos resultados, dentre outras que possam trazer melhorias na tarefa de classificação em fluxos de dados com distribuições dinâmicas;
- Propor um novo método que combine as estratégias de detecção das duas versões do EMZD e um novo detector capaz de combinar os testes estatísticos aplicados no FPDD, FSDD e FTDD, com o objetivo de aprimorar a capacidade de detecção de mudanças de conceitos em diferentes cenários.

ACRÔNIMOS

- ADWIN** *Adaptative Windowing*. 15, 29, 32, 49, 54, 59, 60, 65, 67, 72, 73
- AM** Aprendizagem de Máquina. 26, 27
- DC** Diferença Crítica. 50
- DDM** *Drift Detection Method*. 15–17, 29, 30, 35, 36, 49, 54, 58–60, 65–67, 72, 73
- DM** Distância Média entre os Reais Pontos de Mudanças de Conceitos e as Detecções. 61–65, 73
- ECDD** *EWMA for Concept Drift Detection*. 15, 29, 32, 33, 49, 54, 59, 60, 64, 65, 67, 72, 73
- EDDM** *Early Drift Detection Method*. 15–17, 29, 30, 35, 38, 49, 59, 60, 65–67, 72, 73
- EMZD** *Equal Means Z – Test Concept Drift Detector*. 16, 17, 35, 36, 38, 43, 66, 72, 75
- EMZD1** *Equal Means Z – Test Concept Drift Detector 1*. 35, 36, 49, 54, 59, 60, 65–67, 72, 73
- EMZD2** *Equal Means Z – Test Concept Drift Detector 2*. 35, 36, 38, 49, 54, 58–60, 65–67, 72–74
- EWMA** *Exponentially Weighted Moving Average*. 32
- FN** Quantidade de Falsos Negativos. 61–66, 73
- FP** Quantidade de Falsos Positivos. 61–66, 73
- FPDD** *Fisher Proportions Drift Detector*. 17, 42, 44–46, 60, 64, 66, 67, 72–75
- FPDD1** *Fisher Proportions Drift Detector 1*. 44, 49, 59, 60, 64, 65, 72
- FPDD2** *Fisher Proportions Drift Detector 2*. 44, 49, 54, 59, 60, 64, 65, 72
- FSDD** *Fisher Square Drift Detector*. 17, 42, 45, 46, 58, 60, 66, 67, 72–75
- FSDD1** *Fisher Square Drift Detector 1*. 46, 49, 54, 59, 65, 72
- FSDD2** *Fisher Square Drift Detector 2*. 46, 49, 54, 59, 60, 65, 72, 73
- FTDD** *Fisher Test Drift Detector*. 17, 42, 45–49, 59, 60, 65–67, 72–75

HT *Hoefding Tree*. 34, 49, 58–60, 64, 65, 67, 72, 73

KS *Kolmogorov – Smirnov*. 66

MCC *Matthews Correlation Coefficient*. 61–67, 73

MOA *Massive Online Analysis*. 17, 36–39, 42, 43, 45, 46, 49, 72

NB *Naive Bayes*. 34, 49, 54, 59, 60, 64, 65, 67, 72, 73, 75

ROC *Receiver Operator Characteristic*. 66

STEPD *Statistical Test of Equal Proportions*. 15–17, 29, 31, 40–42, 44, 46, 49, 59, 60, 64–67, 72, 73

VN *Quantidade de Verdadeiros Negativos*. 61–66, 73

VP *Quantidade de Verdadeiros Positivos*. 61–66, 73

REFERÊNCIAS

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. Database mining: A performance perspective. *IEEE Trans. on Knowl. and Data Eng.*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 5, n. 6, p. 914–925, dez. 1993. ISSN 1041-4347. Disponível em: <<http://dx.doi.org/10.1109/69.250074>>. Citado na página 51.
- AGRESTI, A. A survey of exact inference for contingency tables. *Statistical Science*, v. 7, n. 1, p. 131–153, 1992. Citado 2 vezes nas páginas 16 e 22.
- ALLUA, S.; THOMPSON, C. B. Hypothesis testing. *Air Medical Journal*, Elsevier, v. 28, p. 108–110, 2009. Citado 4 vezes nas páginas 19, 20, 22 e 25.
- ARMITAGE, P.; BERRY, G.; MATTHEWS, J. *Statistical Methods in Medical Research (Fourth Edition)*. Oxford, UK: Blacwell Science, 2002. ISBN 0-632-05257-0. Citado 2 vezes nas páginas 24 e 45.
- ATTAR, V. et al. An instance-window based classification algorithm for handling gradual concept drifts. In: CAO, L. et al. (Ed.). *Agents and Data Mining Interaction: 7th International Workshop on Agents and Data Mining Interation, ADMI 2011, Taipei, Taiwan, May 2-6, 2011, Revised Selected Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 156–172. ISBN 978-3-642-27609-5. Disponível em: <http://dx.doi.org/10.1007/978-3-642-27609-5_11>. Citado na página 29.
- BAENA-GARCIA, M. et al. Early drift detection method. In: *International Workshop on Knowledge Discovery from Data Streams*. [S.l.: s.n.], 2006. p. 77–86. Citado 3 vezes nas páginas 15, 16 e 29.
- BANERJEE, A. et al. Hypothesis testing, type i and type ii errors. In: *Industrial Psychiatry Journal*. [S.l.: s.n.], 2009. v. 18, n. 2, p. 127–131. Citado na página 66.
- 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. Citado 2 vezes nas páginas 15 e 29.
- BIFET, A. et al. MOA: Massive online analysis. *Journal of Machine Learning Research*, MIT Press, v. 11, p. 1601–1604, 2010. Citado 4 vezes nas páginas 17, 28, 49 e 51.
- BIFET, A. et al. 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: ACM, 2009. (KDD '09), p. 139–148. ISBN 978-1-60558-495-9. Disponível em: <<http://dx.doi.org/10.1145/1557019.1557041>>. Citado 3 vezes nas páginas 52, 53 e 54.
- BLUMAN, A. *Elementary Statistics: A Step by Step Approach*. Ninth. New York, USA: McGraw-Hill, 2014. Citado 8 vezes nas páginas 19, 20, 21, 22, 24, 25, 44 e 45.
- BRADLEY, A. P. {ROC} curve equivalence using the kolmogorov–smirnov test. *Pattern Recognition Letters*, v. 34, n. 5, p. 470 – 475, 2013. ISSN 0167-8655. Citado na página 66.

BRZEZINSKI, D.; STEAFNOWSKI, J. Stream classification. In: *Encyclopedia of Machine Learning*. [S.l.]: Springer, 2016. Citado 5 vezes nas páginas 15, 27, 28, 29 e 34.

CRANS, G. G.; SHUSTER, J. J. How conservative is fisher's exact test? a quantitative evaluation of the two-sample comparative binomial trial. *Statistics in Medicine*, John Wiley & Sons, Ltd., v. 27, n. 18, p. 3598–3611, 2008. ISSN 1097-0258. Disponível em: <<http://dx.doi.org/10.1002/sim.3221>>. Citado na página 66.

DAVIS, J.; GOADRICH, M. The relationship between precision-recall and roc curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY, USA: ACM, 2006. (ICML '06), p. 233–240. ISBN 1-59593-383-2. Disponível em: <<http://doi.acm.org/10.1145/1143844.1143874>>. Citado na página 66.

DAVIS, R. B.; MUKAMAL, K. J. Hypothesis testing: Means. *Circulation*, v. 114, n. 10, p. 1078–1082, 2006. Citado na página 19.

DAYAN, P. Unsupervised learning. In: *The MIT Encyclopedia of the Cognitive Sciences*. New York, USA: The MIT Press, 1999. p. 857–859. Citado na página 27.

DELANY, S. J. et al. A case-based technique for tracking concept drift in spam filtering. *Know.-Based Syst.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 18, n. 4-5, p. 187–195, ago. 2005. ISSN 0950-7051. Disponível em: <<http://dx.doi.org/10.1016/j.knosys.2004.10.002>>. Citado na página 28.

DEMSAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, JMLR.org, v. 7, p. 1–30, 2006. Citado na página 50.

DU, L.; SONG, Q.; JIA, X. Detecting concept drift: An information entropy based method using an adaptive sliding window. *Intell. Data Anal.*, IOS Press, Amsterdam, The Netherlands, The Netherlands, v. 18, n. 3, p. 337–364, maio 2014. ISSN 1088-467X. Disponível em: <<http://dx.doi.org/10.3233/IDA-140645>>. Citado na página 61.

DU, L. et al. A selective detector ensemble for concept drift detection. *The Computer Journal*, 2014. Disponível em: <<http://comjnl.oxfordjournals.org/content/early/2014/06/19comjnl.bxu050.abstract>>. Citado 2 vezes nas páginas 16 e 35.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification (Second Edition)*. New York, NY, USA: JOHN WILEY & SONS, INC., 2001. Citado na página 26.

FACELI, K. et al. *Inteligência Artificial: Uma Abordagem de Aprendizagem de Máquina*. Rio de Janeiro, RJ, Brasil: LTC, 2011. Citado 2 vezes nas páginas 26 e 27.

FAWCETT, T. An introduction to roc analysis. *Pattern Recogn. Lett.*, Elsevier Science Inc., New York, NY, USA, v. 27, n. 8, p. 861–874, jun. 2006. ISSN 0167-8655. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2005.10.010>>. Citado na página 66.

FISHER, R. *Statistical Methods for Research Workers*. London, England: Oliver and Boyd, 1934. (Biological Monographs and Manuals). ISBN 0-05-002170-2. Citado na página 23.

FREEMAN, J. A. S.; SAAD, D. Online learning in radial basis function networks. *Neural Computation*, v. 9, n. 7, p. 1601–1622, 1997. ISSN 0899-7667. Citado na página 75.

- GAMA, J. et al. Learning with drift detection. In: *Advances in Artificial Intelligence: SBIA 2004*. [S.l.]: Springer, 2004, (LNCS, v. 3171). p. 286–295. Citado 6 vezes nas páginas 15, 16, 27, 29, 52 e 53.
- GAMA, J.; SEBASTIÃO, R.; RODRIGUES, P. P. On evaluating stream learning algorithms. *Machine Learning*, v. 90, n. 3, p. 317–346, 2013. Citado na página 54.
- GONÇALVES JR., P. M.; BARROS, R. S. M. RCD: A recurring concept drift framework. *Pattern Recognition Letters*, Elsevier, v. 34, n. 9, p. 1018–1025, 2013. Citado 2 vezes nas páginas 15 e 28.
- HULTEN, G.; SPENCER, L.; DOMINGOS, P. Mining time-changing data streams. In: *ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*. [S.l.]: ACM Press, 2001. p. 97–106. Citado na página 34.
- IENCO, D. et al. Clustering based active learning for evolving data streams. In: FÜRNKRANZ, J.; HÜLLERMEIER, E.; HIGUCHI, T. (Ed.). *Discovery Science (DS'13)*. [S.l.]: Springer, 2013, (LNCS, v. 8140). p. 79–93. Citado na página 53.
- JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. In: *Eleventh Conference on Uncertainty in Artificial Intelligence*. San Mateo: Morgan Kaufmann, 1995. p. 338–345. Citado na página 34.
- KATAKIS, I.; TSOUMAKAS, G.; VLAHAVAS, I. Tracking recurring contexts using ensemble classifiers: an application to email filtering. *Knowledge and Information Systems*, v. 22, n. 3, p. 371–391, 2010. Citado 2 vezes nas páginas 15 e 29.
- KOLTER, J. Z.; MALOOF, M. A. Dynamic weighted majority: An ensemble method for drifting concepts. *Journal of Machine Learning Research*, JMLR.org, v. 8, p. 2755–2790, 2007. Citado na página 28.
- LANE, T.; BRODLEY, C. E. Approaches to online learning and concept drift for user identification in computer security. In: *Proceedings of the 4th Internat. Conference on Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press, 1998. p. 259–263. Citado 2 vezes nas páginas 15 e 29.
- LARSON, R.; FARBER, B. *Estatística Aplicada*. 4. ed. São Paulo, Brasil: Pearson Prentice Hall, 2010. Citado 8 vezes nas páginas 16, 20, 21, 22, 24, 25, 35 e 46.
- LASKOV, P. et al. Incremental support vector learning: Analysis, implementation and applications. *J. Mach. Learn. Res.*, JMLR.org, v. 7, p. 1909–1936, 2006. ISSN 1532-4435. Citado na página 75.
- LEE, Y.; WANG, L.; RYU, K. A system architecture for monitoring sensor data stream. In: *Proceedings of The 7th IEEE International Conf. on Computer and Information Technology*. [S.l.: s.n.], 2007. p. 1026–1031. Citado 2 vezes nas páginas 15 e 29.
- MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, v. 405, n. 2, p. 442–451, 1975. Citado 2 vezes nas páginas 61 e 67.
- MCDONALD, J. *Handbook of Biological Statistics*. 3. ed. Maryland, USA: Sparky House Publishing, 2014. Citado 3 vezes nas páginas 16, 23 e 42.

- MEHTA, C.; PATEL, N. A network algorithm for performing fisher's exact test in $r \times c$ contingency tables. *Journal of the American Statistical Association*, v. 78, n. 382, p. 427–434, 1983. Citado na página 23.
- MEHTA, C. R.; PATEL, N. R. *SPSS Exact Tests 7.0™ for Windows*®. Chicago, USA: SPSS Inc, 1996. Citado 3 vezes nas páginas 16, 23 e 42.
- MINKU, L. L.; WHITE, A. P.; YAO, X. The impact of diversity on online ensemble learning in the presence of concept drift. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 5, p. 730–742, May 2010. ISSN 1041-4347. Citado 2 vezes nas páginas 15 e 28.
- MITCHELL, T. *Machine Learning*. New York, NY, USA: McGraw-Hill Science/Engineering/Math, 1997. Citado 2 vezes nas páginas 26 e 30.
- NISHIDA, K.; YAMAUCHI, K. Detecting concept drift using statistical testing. In: CORRUBLE, V.; TAKEDA, M.; SUZUKI, E. (Ed.). *Proceedings of 10th International Conference on Discovery Science (DS'07)*. [S.l.]: Springer, 2007. (LNCS, v. 4755), p. 264–269. ISBN 978-3-540-75487-9. Citado 4 vezes nas páginas 15, 29, 40 e 41.
- PERRY, J.; KENT, A.; BERRY, M. Machine literature searching X. machine language; factors underlying its design and development. *American Documentation*, v. 6, n. 4, p. 242–254, 1955. Citado na página 61.
- PESARANGHADER, A.; VIKTOR, H. L. Fast hoeffding drift detection method for evolving data streams. In: _____. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part II*. Cham: Springer International Publishing, 2016. p. 96–111. ISBN 978-3-319-46227-1. Disponível em: <http://dx.doi.org/10.1007/978-3-319-46227-1_7>. Citado na página 61.
- ROBERTS, S. W. Control chart tests based on geometric moving averages. *Technometrics*, v. 1, n. 3, p. 239–250, 1959. Citado na página 32.
- ROSS, G. J. et al. Exponentially weighted moving average charts for detecting concept drift. *Pattern Recognition Letters*, v. 33, n. 2, p. 191–198, 2012. Citado 2 vezes nas páginas 15 e 29.
- SAFFARI, A. et al. On-line random forests. In: *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. [S.l.: s.n.], 2009. p. 1393–1400. Citado na página 75.
- SALPERWYCK, C.; BOULLÉ, M.; LEMAIRE, V. Concept drift detection using supervised bivariate grids. In: *Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN)*. Killarney, Ireland: [s.n.], 2015. p. 1–9. Citado 2 vezes nas páginas 16 e 35.
- SANTOS, S. G. T. C.; BARROS, R. S. M.; GONÇALVES JR., P. M. Optimizing the parameters of drift detection methods using a genetic algorithm. In: *Proc. of The 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'15)*. Italy: [s.n.], 2015. p. 1077–1084. ISSN 1082-3409. Citado 3 vezes nas páginas 36, 42 e 66.

- SMAILOVIC, J. et al. Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences*, v. 285, p. 181–203, 2014. Citado 2 vezes nas páginas 15 e 29.
- STREET, W. N.; KIM, Y. A streaming ensemble algorithm (SEA) for large-scale classification. In: *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2001. (KDD '01), p. 377–382. ISBN 1-58113-391-X. Disponível em: <<http://dx.doi.org/10.1145/502512.502568>>. Citado na página 52.
- WANG, H.; ABRAHAM, Z. Concept drift detection for streaming data. In: *Proceedings of IEEE International Joint Conference of Neural Networks (IJCNN)*. Killarney, Ireland: [s.n.], 2015. p. 1–9. Citado na página 61.
- ZHU, X. *Stream Data Mining Repository*. 2010. Online. Disponível em: <<http://www.cse.fau.edu/~xqzhu/stream.html>>. Citado na página 54.
- ŽLIOBAITĖ, I.; PECHENIZKIY, M.; GAMA, J. An overview of concept drift applications. In: JAPKOWICZ, N.; STEFANOWSKI, J. (Ed.). *Big Data Analysis: New Alg. for a New Society*. [S.l.]: Springer, 2016. p. 91–114. Citado na página 15.

APÊNDICES

APÊNDICE A – TABELAS DETALHADAS, POR MÉTRICAS, COM AS ANÁLISES DAS DETECÇÕES DE MUDANÇAS DE CONCEITOS

Este apêndice apresenta os resultados alcançados por cada um dos detectores testados, no que diz respeito as análises das detecções de mudanças conceitos, de acordo com cada uma das métricas estabelecidas.

Tabela 11 – Médias das Distâncias entre as Mudanças de Conceitos Abruptas e as Detecções dos Métodos, Utilizando NB nos Conjuntos de Dados Artificiais.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média Geral
20K	ADWIN	70,00	45,88	40,00	50,00	NaN	40,31	49,24
	DDM	70,00	NaN	43,21	NaN	NaN	50,18	54,46
	EDDM	40,00	80,00	51,25	38,00	44,55	38,66	48,74
	ECDD	32,95	21,25	10,00	NaN	35,54	9,94	21,94
	STEPD	45,80	28,73	10,25	54,00	45,56	13,69	33,01
	EMZD1	50,71	58,43	20,44	47,50	53,57	20,40	41,84
	EMZD2	53,81	60,18	24,13	30,00	47,14	25,51	40,13
	FPDD1	52,35	35,06	17,56	46,67	NaN	19,19	34,17
	FPDD2	52,35	35,06	17,56	46,67	NaN	20,19	34,37
	FSDD1	50,29	36,49	17,56	46,67	NaN	18,50	33,90
	FSDD2	50,29	36,49	15,13	46,67	70,00	17,50	39,35
	FTDD	52,22	34,94	17,56	46,67	NaN	18,94	34,07
50K	ADWIN	145,13	79,40	33,38	48,00	149,26	41,44	82,77
	DDM	144,00	148,04	70,51	NaN	120,00	88,26	114,16
	EDDM	70,00	NaN	138,65	30,00	70,00	79,00	77,53
	ECDD	53,48	27,32	10,00	NaN	72,08	9,87	34,55
	STEPD	70,57	47,07	10,44	83,33	81,57	14,62	51,27
	EMZD1	118,89	104,59	29,38	100,00	133,85	29,30	86,00
	EMZD2	120,50	110,00	31,94	87,35	138,16	34,63	87,10
	FPDD1	85,91	56,36	17,13	67,50	NaN	20,13	49,41
	FPDD2	85,91	56,36	17,13	67,50	NaN	20,63	49,51
	FSDD1	82,39	47,65	17,13	77,50	NaN	19,44	48,82
	FSDD2	82,39	47,65	14,38	77,50	113,33	18,06	58,89
	FTDD	85,91	55,69	17,13	67,50	NaN	19,69	49,18
100K	ADWIN	208,37	120,00	40,00	286,96	237,37	40,38	155,51
	DDM	318,46	247,78	101,76	203,33	340,00	137,39	224,79
	EDDM	293,33	NaN	285,44	106,67	183,33	173,30	208,41
	ECDD	67,37	40,98	9,88	NaN	109,58	10,20	47,60
	STEPD	92,97	61,11	10,25	213,33	126,37	13,19	86,20
	EMZD1	191,40	136,25	33,00	213,33	216,32	50,60	140,15
	EMZD2	217,84	145,37	38,50	236,06	238,57	49,80	154,36
	FPDD1	138,96	65,63	18,00	218,75	373,33	18,94	138,94
	FPDD2	138,96	65,63	18,00	218,75	373,33	19,88	139,09
	FSDD1	141,09	58,29	18,00	228,75	373,33	18,31	139,63
	FSDD2	141,09	58,29	15,94	228,75	283,33	17,00	124,07
	FTDD	138,96	66,35	18,00	228,75	373,33	18,56	140,66

Tabela 12 – Médias das Distâncias entre as Mudanças de Conceitos Abruptas e as Detecções dos Métodos, Utilizando HT nos Conjuntos de Dados Artificiais.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média Geral
20K	ADWIN	45,71	40,91	40,00	50,00	0,00	40,19	36,14
	DDM	50,00	NaN	33,90	NaN	46,32	48,11	44,58
	EDDM	47,50	80,00	44,24	46,67	50,00	32,56	50,16
	ECDD	31,62	21,25	9,87	NaN	36,14	10,25	21,83
	STEPD	43,72	28,07	11,00	47,50	45,95	12,83	31,51
	EMZD1	40,00	58,71	13,18	NaN	45,11	24,05	36,21
	EMZD2	43,75	60,35	16,84	42,86	48,82	31,46	40,68
	FPDD1	53,33	34,10	17,06	46,67	NaN	18,38	33,91
	FPDD2	53,33	34,10	17,06	46,67	NaN	19,94	34,22
	FSDD1	50,00	36,17	17,06	46,67	NaN	18,38	33,66
	FSDD2	50,00	36,17	15,75	46,67	NaN	17,00	33,12
FTDD	53,33	35,57	17,06	46,67	NaN	18,38	34,20	
50K	ADWIN	135,12	78,16	32,63	170,00	155,00	41,00	101,99
	DDM	150,34	148,26	63,48	NaN	100,00	70,06	106,43
	EDDM	111,67	NaN	106,97	80,00	114,58	79,88	98,62
	ECDD	48,11	28,48	10,06	NaN	71,16	10,27	33,62
	STEPD	62,27	50,70	10,63	79,17	82,42	11,84	49,51
	EMZD1	116,60	105,16	23,13	140,00	112,00	35,38	88,71
	EMZD2	137,41	110,50	28,81	104,00	126,13	43,90	91,79
	FPDD1	82,67	56,36	18,38	62,50	NaN	17,56	47,49
	FPDD2	82,67	56,36	18,38	62,50	NaN	18,63	47,71
	FSDD1	82,15	47,65	18,38	30,00	NaN	17,56	39,15
	FSDD2	81,94	47,65	15,88	30,00	136,67	15,69	54,64
FTDD	76,56	55,69	18,38	62,50	NaN	17,56	46,14	
100K	ADWIN	173,57	117,43	40,00	306,88	253,24	40,31	155,24
	DDM	235,58	244,31	83,38	NaN	220,00	91,31	174,92
	EDDM	240,00	NaN	287,54	300,00	160,67	139,01	225,44
	ECDD	56,18	41,43	9,81	NaN	111,54	14,70	46,73
	STEPD	71,46	64,10	11,88	197,50	126,44	11,82	80,53
	EMZD1	194,18	136,69	28,62	316,67	167,33	48,44	148,66
	EMZD2	213,28	147,19	38,62	242,50	195,43	55,31	148,72
	FPDD1	112,23	65,20	16,88	196,67	336,67	17,06	124,12
	FPDD2	112,33	65,20	17,94	196,67	336,67	17,63	124,41
	FSDD1	88,67	58,37	16,81	201,11	336,67	17,06	119,78
	FSDD2	88,52	58,37	14,69	201,11	197,06	15,44	95,87
FTDD	98,67	66,35	16,88	184,00	336,67	17,06	119,94	

Tabela 13 – Taxas de Falsos Positivos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	348	1687	208	523	292	141	533,17
	DDM	142	156	17	101	186	118	120,00
	EDDM	863	271	419	501	631	1034	619,83
	ECDD	1230	555	1034	0	1050	1297	861,00
	STEPD	302	1379	149	346	212	216	434,00
	EMZD1	359	163	131	118	464	718	325,50
	EMZD2	471	109	29	948	217	224	333,00
	FPDD1	53	89	7	44	6	3	33,67
	FPDD2	53	89	7	44	6	3	33,67
	FSDD1	81	156	7	50	6	3	50,50
	FSDD2	81	157	4	50	22	5	53,17
FTDD	55	105	7	46	6	3	37,00	
50K	ADWIN	342	4034	239	752	578	192	1022,83
	DDM	139	124	26	119	192	137	122,83
	EDDM	957	288	579	561	762	1650	799,50
	ECDD	3046	1445	2498	0	2630	3396	2169,17
	STEPD	588	2373	418	864	429	589	876,83
	EMZD1	390	147	309	161	582	1347	489,33
	EMZD2	546	56	74	1708	272	458	519,00
	FPDD1	64	62	7	67	18	2	36,67
	FPDD2	64	62	7	67	18	2	36,67
	FSDD1	111	127	7	72	18	2	56,17
	FSDD2	111	128	7	72	43	5	61,00
FTDD	70	72	7	69	18	2	39,67	
100K	ADWIN	307	8049	230	978	650	162	1729,33
	DDM	140	103	48	123	178	157	124,83
	EDDM	1027	292	567	525	812	2054	879,50
	ECDD	6182	2999	5141	0	5297	6955	4429,00
	STEPD	1034	4336	919	1735	878	1148	1675,00
	EMZD1	565	203	517	262	828	2136	751,83
	EMZD2	677	62	175	2213	305	750	697,00
	FPDD1	88	73	7	76	28	3	45,83
	FPDD2	88	73	7	76	27	3	45,67
	FSDD1	99	206	7	87	28	6	72,17
	FSDD2	99	204	4	87	65	12	78,50
FTDD	89	78	7	78	28	3	47,17	

Tabela 14 – Taxas de Falsos Positivos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	364	1776	236	385	330	263	559,00
	DDM	138	157	42	80	140	112	111,50
	EDDM	925	269	392	324	542	907	559,83
	ECDD	1291	553	1013	0	1023	1314	865,67
	STEPD	365	1501	263	378	271	240	503,00
	EMZD1	543	155	369	120	519	628	389,00
	EMZD2	523	108	143	581	229	159	290,50
	FPDD1	75	109	23	47	5	7	44,33
	FPDD2	75	109	23	47	5	6	44,17
	FSDD1	95	178	24	53	5	7	60,33
	FSDD2	95	179	42	53	16	19	67,33
FTDD	78	116	23	48	5	7	46,17	
50K	ADWIN	597	4321	396	527	508	617	1161,00
	DDM	127	124	85	96	214	165	135,17
	EDDM	1119	291	737	345	931	1133	759,33
	ECDD	3249	1437	2456	0	2560	3293	2165,83
	STEPD	751	2734	649	890	662	489	1029,17
	EMZD1	715	137	774	138	1177	903	640,67
	EMZD2	667	54	307	999	587	175	464,83
	FPDD1	127	66	41	71	12	11	54,67
	FPDD2	127	66	41	71	13	8	54,33
	FSDD1	169	145	42	80	12	11	76,50
	FSDD2	169	144	63	80	45	18	86,50
FTDD	120	80	41	72	12	11	56,00	
100K	ADWIN	1122	8409	747	677	661	1281	2149,50
	DDM	109	88	110	94	252	133	131,00
	EDDM	1179	293	828	329	1151	1258	839,67
	ECDD	6321	2987	5085	0	5171	6658	4370,33
	STEPD	1461	4885	1350	1776	1370	884	1954,33
	EMZD1	847	200	1458	249	2112	912	963,00
	EMZD2	723	60	584	1217	961	178	620,50
	FPDD1	284	115	23	80	20	10	88,67
	FPDD2	284	115	29	80	21	10	89,83
	FSDD1	267	264	24	85	20	10	111,67
	FSDD2	268	269	51	86	50	24	124,67
FTDD	261	131	23	79	20	10	87,33	

Tabela 15 – Taxas de Falsos Negativos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	157	109	0	157	160	0	97,17
	DDM	159	160	1	160	160	46	114,33
	EDDM	159	159	64	155	149	48	122,33
	ECDD	72	64	4	160	104	2	67,67
	STEPD	72	50	0	155	115	0	65,33
	EMZD1	146	90	1	156	146	9	91,33
	EMZD2	139	103	0	154	153	13	93,67
	FPDD1	143	81	0	157	160	0	90,17
	FPDD2	143	81	0	157	160	0	90,17
	FSDD1	125	66	0	157	160	0	84,67
	FSDD2	125	66	0	157	159	0	84,50
	FTDD	142	81	0	157	160	0	90,00
50K	ADWIN	84	77	0	155	133	0	74,83
	DDM	155	114	2	160	157	45	105,50
	EDDM	159	160	64	156	159	60	126,33
	ECDD	45	48	0	160	64	2	53,17
	STEPD	38	37	0	151	71	2	49,83
	EMZD1	124	38	0	159	108	18	74,50
	EMZD2	60	39	0	126	122	13	60,00
	FPDD1	116	50	0	156	160	0	80,33
	FPDD2	116	50	0	156	160	0	80,33
	FSDD1	93	41	0	156	160	0	75,00
	FSDD2	93	41	0	156	157	0	74,50
	FTDD	116	51	0	156	160	0	80,50
100K	ADWIN	7	59	0	137	84	0	47,83
	DDM	147	88	7	157	159	41	99,83
	EDDM	157	160	70	157	157	66	127,83
	ECDD	42	48	0	160	42	9	50,17
	STEPD	22	25	0	142	58	0	41,17
	EMZD1	117	24	0	157	84	26	68,00
	EMZD2	35	26	0	127	83	13	47,33
	FPDD1	93	34	0	152	157	0	72,67
	FPDD2	93	34	0	152	157	0	72,67
	FSDD1	59	31	0	152	157	0	66,50
	FSDD2	59	31	0	152	148	0	65,00
	FTDD	93	34	0	152	157	0	72,67

Tabela 16 – Taxas de Falsos Negativos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	153	105	2	157	158	2	96,17
	DDM	156	160	6	160	141	1	104,00
	EDDM	152	159	35	157	150	78	121,83
	ECDD	55	64	4	160	103	3	64,83
	STEPD	74	51	0	156	123	1	67,50
	EMZD1	141	90	3	160	113	2	84,83
	EMZD2	128	103	2	153	126	3	85,83
	FPDD1	136	82	0	157	160	0	89,17
	FPDD2	136	82	0	157	160	0	89,17
	FSDD1	123	66	0	157	160	0	84,33
	FSDD2	123	66	0	157	160	0	84,33
	FTDD	136	81	0	157	160	0	89,00
50K	ADWIN	78	73	0	157	144	0	75,33
	DDM	131	114	5	160	153	0	93,83
	EDDM	154	160	84	158	136	78	128,33
	ECDD	38	48	2	160	65	10	53,83
	STEPD	41	32	0	148	94	2	52,83
	EMZD1	110	38	0	158	90	2	66,33
	EMZD2	75	39	0	140	98	1	58,83
	FPDD1	100	50	0	156	160	0	77,67
	FPDD2	100	50	0	156	160	0	77,67
	FSDD1	67	41	0	158	160	0	71,00
	FSDD2	67	41	0	158	157	0	70,50
	FTDD	96	51	0	156	160	0	77,17
100K	ADWIN	34	55	0	144	92	0	54,17
	DDM	108	88	3	160	155	0	85,67
	EDDM	155	160	91	159	130	79	129,00
	ECDD	29	48	1	160	43	11	48,67
	STEPD	37	38	0	136	73	1	47,50
	EMZD1	93	27	1	157	74	0	58,67
	EMZD2	23	25	1	128	79	0	42,67
	FPDD1	57	35	0	151	157	0	66,67
	FPDD2	57	35	0	151	157	0	66,67
	FSDD1	32	31	0	151	157	0	61,83
	FSDD2	32	31	0	151	143	0	59,50
	FTDD	55	34	0	150	157	0	66,00

Tabela 17 – Taxas de Verdadeiros Positivos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	3	51	160	3	0	160	62,83
	DDM	1	0	159	0	0	114	45,67
	EDDM	1	1	96	5	11	112	37,67
	ECDD	88	96	156	0	56	158	92,33
	STEPD	88	110	160	5	45	160	94,67
	EMZD1	14	70	159	4	14	151	68,67
	EMZD2	21	57	160	6	7	147	66,33
	FPDD1	17	79	160	3	0	160	69,83
	FPDD2	17	79	160	3	0	160	69,83
	FSDD1	35	94	160	3	0	160	75,33
	FSDD2	35	94	160	3	1	160	75,50
FTDD	18	79	160	3	0	160	70,00	
50K	ADWIN	76	83	160	5	27	160	85,17
	DDM	5	46	158	0	3	115	54,50
	EDDM	1	0	96	4	1	100	33,67
	ECDD	115	112	160	0	96	158	106,83
	STEPD	122	123	160	9	89	158	110,17
	EMZD1	36	122	160	1	52	142	85,50
	EMZD2	100	121	160	34	38	147	100,00
	FPDD1	44	110	160	4	0	160	79,67
	FPDD2	44	110	160	4	0	160	79,67
	FSDD1	67	119	160	4	0	160	85,00
	FSDD2	67	119	160	4	3	160	85,50
FTDD	44	109	160	4	0	160	79,50	
100K	ADWIN	153	101	160	23	76	160	112,17
	DDM	13	72	153	3	1	119	60,17
	EDDM	3	0	90	3	3	94	32,17
	ECDD	118	112	160	0	118	151	109,83
	STEPD	138	135	160	18	102	160	118,83
	EMZD1	43	136	160	3	76	134	92,00
	EMZD2	125	134	160	33	77	147	112,67
	FPDD1	67	126	160	8	3	160	87,33
	FPDD2	67	126	160	8	3	160	87,33
	FSDD1	101	129	160	8	3	160	93,50
	FSDD2	101	129	160	8	12	160	95,00
FTDD	67	126	160	8	3	160	87,33	

Tabela 18 – Taxas de Verdadeiros Positivos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	7	55	158	3	2	158	63,83
	DDM	4	0	154	0	19	159	56,00
	EDDM	8	1	125	3	10	82	38,17
	ECDD	105	96	156	0	57	157	95,17
	STEPD	86	109	160	4	37	159	92,50
	EMZD1	19	70	157	0	47	158	75,17
	EMZD2	32	57	158	7	34	157	74,17
	FPDD1	24	78	160	3	0	160	70,83
	FPDD2	24	78	160	3	0	160	70,83
	FSDD1	37	94	160	3	0	160	75,67
	FSDD2	37	94	160	3	0	160	75,67
FTDD	24	79	160	3	0	160	71,00	
50K	ADWIN	82	87	160	3	16	160	84,67
	DDM	29	46	155	0	7	160	66,17
	EDDM	6	0	76	2	24	82	31,67
	ECDD	122	112	158	0	95	150	106,17
	STEPD	119	128	160	12	66	158	107,17
	EMZD1	50	122	160	2	70	158	93,67
	EMZD2	85	121	160	20	62	159	101,17
	FPDD1	60	110	160	4	0	160	82,33
	FPDD2	60	110	160	4	0	160	82,33
	FSDD1	93	119	160	2	0	160	89,00
	FSDD2	93	119	160	2	3	160	89,50
FTDD	64	109	160	4	0	160	82,83	
100K	ADWIN	126	105	160	16	68	160	105,83
	DDM	52	72	157	0	5	160	74,33
	EDDM	5	0	69	1	30	81	31,00
	ECDD	131	112	159	0	117	149	111,33
	STEPD	123	122	160	24	87	159	112,50
	EMZD1	67	133	159	3	86	160	101,33
	EMZD2	137	135	159	32	81	160	117,33
	FPDD1	103	125	160	9	3	160	93,33
	FPDD2	103	125	160	9	3	160	93,33
	FSDD1	128	129	160	9	3	160	98,17
	FSDD2	128	129	160	9	17	160	100,50
FTDD	105	126	160	10	3	160	94,00	

Tabela 19 – Taxas de Verdadeiros Negativos dos Detectores, Utilizando NB nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	799492	798153	799632	799317	799548	799699	799306,83
	DDM	799698	799684	799823	799739	799654	799722	799720,00
	EDDM	798977	799569	799421	799339	799209	798806	799220,17
	ECDD	798610	799285	798806	799840	798790	798543	798979,00
	STEPD	799538	798461	799691	799494	799628	799624	799406,00
	EMZD1	799481	799677	799709	799722	799376	799122	799514,50
	EMZD2	799369	799731	799811	798892	799623	799616	799507,00
	FPDD1	799787	799751	799833	799796	799834	799837	799806,33
	FPDD2	799787	799751	799833	799796	799834	799837	799806,33
	FSDD1	799759	799684	799833	799790	799834	799837	799789,50
	FSDD2	799759	799683	799836	799790	799818	799835	799786,83
FTDD	799785	799735	799833	799794	799834	799837	799803,00	
50K	ADWIN	1999498	1995806	1999601	1999088	1999262	1999648	1998817,17
	DDM	1999701	1999716	1999814	1999721	1999648	1999703	1999717,17
	EDDM	1998883	1999552	1999261	1999279	1999078	1998190	1999040,50
	ECDD	1996794	1998395	1997342	1999840	1997210	1996444	1997670,83
	STEPD	1999252	1997467	1999422	1998976	1999411	1999251	1998963,17
	EMZD1	1999450	1999693	1999531	1999679	1999258	1998493	1999350,67
	EMZD2	1999294	1999784	1999766	1998132	1999568	1999382	1999321,00
	FPDD1	1999776	1999778	1999833	1999773	1999822	1999838	1999803,33
	FPDD2	1999776	1999778	1999833	1999773	1999822	1999838	1999803,33
	FSDD1	1999729	1999713	1999833	1999768	1999822	1999838	1999783,83
	FSDD2	1999729	1999712	1999833	1999768	1999797	1999835	1999779,00
FTDD	1999770	1999768	1999833	1999771	1999822	1999838	1999800,33	
100K	ADWIN	3999533	3991791	3999610	3998862	3999190	3999678	3998110,67
	DDM	3999700	3999737	3999792	3999717	3999662	3999683	3999715,17
	EDDM	3998813	3999548	3999273	3999315	3999028	3997786	3998960,50
	ECDD	3993658	3996841	3994699	3999840	3994543	3992885	3995411,00
	STEPD	3998806	3995504	3998921	3998105	3998962	3998692	3998165,00
	EMZD1	3999275	3999637	3999323	3999578	3999012	3997704	3999088,17
	EMZD2	3999163	3999778	3999665	3997627	3999535	3999090	3999143,00
	FPDD1	3999752	3999767	3999833	3999764	3999812	3999837	3999794,17
	FPDD2	3999752	3999767	3999833	3999764	3999813	3999837	3999794,33
	FSDD1	3999741	3999634	3999833	3999753	3999812	3999834	3999767,83
	FSDD2	3999741	3999636	3999836	3999753	3999775	3999828	3999761,50
FTDD	3999751	3999762	3999833	3999762	3999812	3999837	3999792,83	

Tabela 20 – Taxas de Verdadeiros Negativos dos Detectores, Utilizando HT nos Conjuntos de Dados Artificiais com Mudanças de Conceitos Abruptas.

Tamanho	Métodos	Agrawal	LED	Mixed	RBF	SEA	Sine	Média
20K	ADWIN	799476	798064	799604	799455	799510	799577	799281,00
	DDM	799702	799683	799798	799760	799700	799728	799728,50
	EDDM	798915	799571	799448	799516	799298	798933	799280,17
	ECDD	798549	799287	798827	799840	798817	798526	798974,33
	STEPD	799475	798339	799577	799462	799569	799600	799337,00
	EMZD1	799297	799685	799471	799720	799321	799212	799451,00
	EMZD2	799317	799732	799697	799259	799611	799681	799549,50
	FPDD1	799765	799731	799817	799793	799835	799833	799795,67
	FPDD2	799765	799731	799817	799793	799835	799834	799795,83
	FSDD1	799745	799662	799816	799787	799835	799833	799779,67
	FSDD2	799745	799661	799798	799787	799824	799821	799772,67
FTDD	799762	799724	799817	799792	799835	799833	799793,83	
50K	ADWIN	1999243	1995519	1999444	1999313	1999332	1999223	1998679,00
	DDM	1999713	1999716	1999755	1999744	1999626	1999675	1999704,83
	EDDM	1998721	1999549	1999103	1999495	1998909	1998707	1999080,67
	ECDD	1996591	1998403	1997384	1999840	1997280	1996547	1997674,17
	STEPD	1999089	1997106	1999191	1998950	1999178	1999351	1998810,83
	EMZD1	1999125	1999703	1999066	1999702	1998663	1998937	1999199,33
	EMZD2	1999173	1999786	1999533	1998841	1999253	1999665	1999375,17
	FPDD1	1999713	1999774	1999799	1999769	1999828	1999829	1999785,33
	FPDD2	1999713	1999774	1999799	1999769	1999827	1999832	1999785,67
	FSDD1	1999671	1999695	1999798	1999760	1999828	1999829	1999763,50
	FSDD2	1999671	1999696	1999777	1999760	1999795	1999822	1999753,50
FTDD	1999720	1999760	1999799	1999768	1999828	1999829	1999784,00	
100K	ADWIN	3998718	3991431	3999093	3999163	3999179	3998559	3997690,50
	DDM	3999731	3999752	3999730	3999746	3999588	3999707	3999709,00
	EDDM	3998661	3999547	3999012	3999511	3998689	3998582	3999000,33
	ECDD	3993519	3996853	3994755	3999840	3994669	3993182	3995469,67
	STEPD	3998379	3994955	3998490	3998064	3998470	3998956	3997885,67
	EMZD1	3998993	3999640	3998382	3999591	3997728	3998928	3998877,00
	EMZD2	3999117	3999780	3999256	3998623	3998879	3999662	3999219,50
	FPDD1	3999556	3999725	3999817	3999760	3999820	3999830	3999751,33
	FPDD2	3999556	3999725	3999811	3999760	3999819	3999830	3999750,17
	FSDD1	3999573	3999576	3999816	3999755	3999820	3999830	3999728,33
	FSDD2	3999572	3999571	3999789	3999754	3999790	3999816	3999715,33
FTDD	3999579	3999709	3999817	3999761	3999820	3999830	3999752,67	

APÊNDICE B – TABELAS DETALHADAS, POR BASES DE DADOS, COM AS ANÁLISES DAS DETECÇÕES DE MUDANÇAS DE CONCEITOS

Neste apêndice são apresentados as acurácias obtidas por cada um dos detectores testados, de acordo com cada classificador utilizado e base dados analisada.

Tabela 21 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados *Agrawal*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	70,00	348	157	3	799492	0,008547	0,018750	0,012367
	DDM	70,00	142	159	1	799698	0,006993	0,006250	0,006423
	EDDM	40,00	863	159	1	798977	0,001157	0,006250	0,002226
	ECDD	32,95	1230	72	88	798610	0,066768	0,550000	0,191233
	STEPD	45,80	302	72	88	799538	0,225641	0,550000	0,352091
	EMZD1	50,71	359	146	14	799481	0,037534	0,087500	0,057021
	EMZD2	53,81	471	139	21	799369	0,042683	0,131250	0,074527
	FPDD1	52,35	53	143	17	799787	0,242857	0,106250	0,160526
	FPDD2	52,35	53	143	17	799787	0,242857	0,106250	0,160526
	FSDD1	50,29	81	125	35	799759	0,301724	0,218750	0,256783
	FSDD2	50,29	81	125	35	799759	0,301724	0,218750	0,256783
FTDD	52,22	55	142	18	799785	0,246575	0,112500	0,166442	
50K	ADWIN	145,13	342	84	76	1999498	0,181818	0,475000	0,293790
	DDM	144,00	139	155	5	1999701	0,034722	0,031250	0,032867
	EDDM	70,00	957	159	1	1998883	0,001044	0,006250	0,002359
	ECDD	53,48	3046	45	115	1996794	0,036381	0,718750	0,161484
	STEPD	70,57	588	38	122	1999252	0,171831	0,762500	0,361879
	EMZD1	118,89	390	124	36	1999450	0,084507	0,225000	0,137781
	EMZD2	120,50	546	60	100	1999294	0,154799	0,625000	0,310947
	FPDD1	85,91	64	116	44	1999776	0,407407	0,275000	0,334676
	FPDD2	85,91	64	116	44	1999776	0,407407	0,275000	0,334676
	FSDD1	82,39	111	93	67	1999729	0,376404	0,418750	0,396962
	FSDD2	82,39	111	93	67	1999729	0,376404	0,418750	0,396962
FTDD	85,91	70	116	44	1999770	0,385965	0,275000	0,325747	
100K	ADWIN	208,37	307	7	153	3999533	0,332609	0,956250	0,563941
	DDM	318,46	140	147	13	3999700	0,084967	0,081250	0,083052
	EDDM	293,33	1027	157	3	3998813	0,002913	0,018750	0,007290
	ECDD	67,37	6182	42	118	3993658	0,018730	0,737500	0,117375
	STEPD	92,97	1034	22	138	3998806	0,117747	0,862500	0,318625
	EMZD1	191,40	565	117	43	3999275	0,070724	0,268750	0,137801
	EMZD2	217,84	677	35	125	3999163	0,155860	0,781250	0,348902
	FPDD1	138,96	88	93	67	3999752	0,432258	0,418750	0,425428
	FPDD2	138,96	88	93	67	3999752	0,432258	0,418750	0,425428
	FSDD1	141,09	99	59	101	3999741	0,505000	0,631250	0,564588
	FSDD2	141,09	99	59	101	3999741	0,505000	0,631250	0,564588
FTDD	138,96	89	93	67	3999751	0,429487	0,418750	0,424062	

Tabela 22 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados *Agrawal*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	45,71	364	153	7	799476	0,018868	0,043750	0,028436
	DDM	50,00	138	156	4	799702	0,028169	0,025000	0,026354
	EDDM	47,50	925	152	8	798915	0,008574	0,050000	0,020237
	ECDD	31,62	1291	55	105	798549	0,075215	0,656250	0,221796
	STEPD	43,72	365	74	86	799475	0,190687	0,537500	0,319934
	EMZD1	40,00	543	141	19	799297	0,033808	0,118750	0,063015
	EMZD2	43,75	523	128	32	799317	0,057658	0,200000	0,107060
	FPDD1	53,33	75	136	24	799765	0,242424	0,150000	0,190566
	FPDD2	53,33	75	136	24	799765	0,242424	0,150000	0,190566
	FSDD1	50,00	95	123	37	799745	0,280303	0,231250	0,254463
	FSDD2	50,00	95	123	37	799745	0,280303	0,231250	0,254463
FTDD	53,33	78	136	24	799762	0,235294	0,150000	0,187738	
50K	ADWIN	135,12	597	78	82	1999243	0,120766	0,512500	0,248669
	DDM	150,34	127	131	29	1999713	0,185897	0,181250	0,183495
	EDDM	111,67	1119	154	6	1998721	0,005333	0,037500	0,013934
	ECDD	48,11	3249	38	122	1996591	0,036191	0,762500	0,165899
	STEPD	62,27	751	41	119	1999089	0,136782	0,743750	0,318849
	EMZD1	116,60	715	110	50	1999125	0,065359	0,312500	0,142774
	EMZD2	137,41	667	75	85	1999173	0,113032	0,531250	0,244930
	FPDD1	82,67	127	100	60	1999713	0,320856	0,375000	0,346817
	FPDD2	82,67	127	100	60	1999713	0,320856	0,375000	0,346817
	FSDD1	82,15	169	67	93	1999671	0,354962	0,581250	0,454172
	FSDD2	81,94	169	67	93	1999671	0,354962	0,581250	0,454172
FTDD	76,56	120	96	64	1999720	0,347826	0,400000	0,372948	
100K	ADWIN	173,57	1122	34	126	3998718	0,100962	0,787500	0,281908
	DDM	235,58	109	108	52	3999731	0,322981	0,325000	0,323962
	EDDM	240,00	1179	155	5	3998661	0,004223	0,031250	0,011381
	ECDD	56,18	6321	29	131	3993519	0,020304	0,818750	0,128785
	STEPD	71,46	1461	37	123	3998379	0,077652	0,768750	0,244252
	EMZD1	194,18	847	93	67	3998993	0,073304	0,418750	0,175131
	EMZD2	213,28	723	23	137	3999117	0,159302	0,856250	0,369282
	FPDD1	112,23	284	57	103	3999556	0,266150	0,643750	0,413891
	FPDD2	112,33	284	57	103	3999556	0,266150	0,643750	0,413891
	FSDD1	88,67	267	32	128	3999573	0,324051	0,800000	0,509129
	FSDD2	88,52	268	32	128	3999572	0,323232	0,800000	0,508486
FTDD	98,67	261	55	105	3999579	0,286885	0,656250	0,433867	

Tabela 23 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados *LED*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	45,88	1687	109	51	798153	0,029344	0,318750	0,096168
	DDM	NaN	156	160	0	799684	0,000000	0,000000	-0,000198
	EDDM	80,00	271	159	1	799569	0,003676	0,006250	0,004534
	ECDD	21,25	555	64	96	799285	0,147465	0,600000	0,297202
	STEPD	28,73	1379	50	110	798461	0,073875	0,687500	0,224986
	EMZD1	58,43	163	90	70	799677	0,300429	0,437500	0,362391
	EMZD2	60,18	109	103	57	799731	0,343373	0,356250	0,349620
	FPDD1	35,06	89	81	79	799751	0,470238	0,493750	0,481744
	FPDD2	35,06	89	81	79	799751	0,470238	0,493750	0,481744
	FSDD1	36,49	156	66	94	799684	0,376000	0,587500	0,469870
	FSDD2	36,49	157	66	94	799683	0,374502	0,587500	0,468933
FTDD	34,94	105	81	79	799735	0,429348	0,493750	0,460309	
50K	ADWIN	79,40	4034	77	83	1995806	0,020160	0,518750	0,101968
	DDM	148,04	124	114	46	1999716	0,270588	0,287500	0,278857
	EDDM	NaN	288	160	0	1999552	0,000000	0,000000	-0,000107
	ECDD	27,32	1445	48	112	1998395	0,071933	0,700000	0,224242
	STEPD	47,07	2373	37	123	1997467	0,049279	0,768750	0,194449
	EMZD1	104,59	147	38	122	1999693	0,453532	0,762500	0,588022
	EMZD2	110,00	56	39	121	1999784	0,683616	0,756250	0,718993
	FPDD1	56,36	62	50	110	1999778	0,639535	0,687500	0,663056
	FPDD2	56,36	62	50	110	1999778	0,639535	0,687500	0,663056
	FSDD1	47,65	127	41	119	1999713	0,483740	0,743750	0,599780
	FSDD2	47,65	128	41	119	1999712	0,481781	0,743750	0,598564
FTDD	55,69	72	51	109	1999768	0,602210	0,681250	0,640481	
100K	ADWIN	120,00	8049	59	101	3991791	0,012393	0,631250	0,088253
	DDM	247,78	103	88	72	3999737	0,411429	0,450000	0,430258
	EDDM	NaN	292	160	0	3999548	0,000000	0,000000	-0,000054
	ECDD	40,98	2999	48	112	3996841	0,036001	0,700000	0,158636
	STEPD	61,11	4336	25	135	3995504	0,030195	0,843750	0,159495
	EMZD1	136,25	203	24	136	3999637	0,401180	0,850000	0,583933
	EMZD2	145,37	62	26	134	3999778	0,683673	0,837500	0,756677
	FPDD1	65,63	73	34	126	3999767	0,633166	0,787500	0,706116
	FPDD2	65,63	73	34	126	3999767	0,633166	0,787500	0,706116
	FSDD1	58,29	206	31	129	3999634	0,385075	0,806250	0,557172
	FSDD2	58,29	204	31	129	3999636	0,387387	0,806250	0,558843
FTDD	66,35	78	34	126	3999762	0,617647	0,787500	0,697408	

Tabela 24 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados *LED*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	40,91	1776	105	55	798064	0,030038	0,343750	0,101064
	DDM	NaN	157	160	0	799683	0,000000	0,000000	-0,000198
	EDDM	80,00	269	159	1	799571	0,003704	0,006250	0,004553
	ECDD	21,25	553	64	96	799287	0,147920	0,600000	0,297660
	STEPD	28,07	1501	51	109	798339	0,067702	0,681250	0,214363
	EMZD1	58,71	155	90	70	799685	0,311111	0,437500	0,368784
	EMZD2	60,35	108	103	57	799732	0,345455	0,356250	0,350679
	FPDD1	34,10	109	82	78	799731	0,417112	0,487500	0,450816
	FPDD2	34,10	109	82	78	799731	0,417112	0,487500	0,450816
	FSDD1	36,17	178	66	94	799662	0,345588	0,587500	0,450453
	FSDD2	36,17	179	66	94	799661	0,344322	0,587500	0,449626
FTDD	35,57	116	81	79	799724	0,405128	0,493750	0,447128	
50K	ADWIN	78,16	4321	73	87	1995519	0,019737	0,543750	0,103293
	DDM	148,26	124	114	46	1999716	0,270588	0,287500	0,278857
	EDDM	NaN	291	160	0	1999549	0,000000	0,000000	-0,000108
	ECDD	28,48	1437	48	112	1998403	0,072305	0,700000	0,224821
	STEPD	50,70	2734	32	128	1997106	0,044724	0,800000	0,188958
	EMZD1	105,16	137	38	122	1999703	0,471042	0,762500	0,599269
	EMZD2	110,50	54	39	121	1999786	0,691429	0,756250	0,723090
	FPDD1	56,36	66	50	110	1999774	0,625000	0,687500	0,655477
	FPDD2	56,36	66	50	110	1999774	0,625000	0,687500	0,655477
	FSDD1	47,65	145	41	119	1999695	0,450758	0,743750	0,578967
	FSDD2	47,65	144	41	119	1999696	0,452471	0,743750	0,580067
FTDD	55,69	80	51	109	1999760	0,576720	0,681250	0,626777	
100K	ADWIN	117,43	8409	55	105	3991431	0,012333	0,656250	0,089768
	DDM	244,31	88	88	72	3999752	0,450000	0,450000	0,449978
	EDDM	NaN	293	160	0	3999547	0,000000	0,000000	-0,000054
	ECDD	41,43	2987	48	112	3996853	0,036141	0,700000	0,158944
	STEPD	64,10	4885	38	122	3994955	0,024366	0,762500	0,136169
	EMZD1	136,69	200	27	133	3999640	0,399399	0,831250	0,576173
	EMZD2	147,19	60	25	135	3999780	0,692308	0,843750	0,764277
	FPDD1	65,20	115	35	125	3999725	0,520833	0,781250	0,637871
	FPDD2	65,20	115	35	125	3999725	0,520833	0,781250	0,637871
	FSDD1	58,37	264	31	129	3999576	0,328244	0,806250	0,514411
	FSDD2	58,37	269	31	129	3999571	0,324121	0,806250	0,511169
FTDD	66,35	131	34	126	3999709	0,490272	0,787500	0,621343	

Tabela 25 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados *Mixed*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	40,00	208	0	160	799632	0,434783	1,000000	0,659295
	DDM	43,21	17	1	159	799823	0,903409	0,993750	0,947493
	EDDM	51,25	419	64	96	799421	0,186408	0,600000	0,334214
	ECDD	10,00	1034	4	156	798806	0,131092	0,975000	0,357268
	STEPD	10,25	149	0	160	799691	0,517799	1,000000	0,719516
	EMZD1	20,44	131	1	159	799709	0,548276	0,993750	0,738077
	EMZD2	24,13	29	0	160	799811	0,846561	1,000000	0,920071
	FPDD1	17,56	7	0	160	799833	0,958084	1,000000	0,978813
	FPDD2	17,56	7	0	160	799833	0,958084	1,000000	0,978813
	FSDD1	17,56	7	0	160	799833	0,958084	1,000000	0,978813
	FSDD2	15,13	4	0	160	799836	0,975610	1,000000	0,987727
FTDD	17,56	7	0	160	799833	0,958084	1,000000	0,978813	
50K	ADWIN	33,38	239	0	160	1999601	0,401003	1,000000	0,633210
	DDM	70,51	26	2	158	1999814	0,858696	0,987500	0,920842
	EDDM	138,65	579	64	96	1999261	0,142222	0,600000	0,292015
	ECDD	10,00	2498	0	160	1997342	0,060196	1,000000	0,245195
	STEPD	10,44	418	0	160	1999422	0,276817	1,000000	0,526079
	EMZD1	29,38	309	0	160	1999531	0,341151	1,000000	0,584037
	EMZD2	31,94	74	0	160	1999766	0,683761	1,000000	0,826883
	FPDD1	17,13	7	0	160	1999833	0,958084	1,000000	0,978816
	FPDD2	17,13	7	0	160	1999833	0,958084	1,000000	0,978816
	FSDD1	17,13	7	0	160	1999833	0,958084	1,000000	0,978816
	FSDD2	14,38	7	0	160	1999833	0,958084	1,000000	0,978816
FTDD	17,13	7	0	160	1999833	0,958084	1,000000	0,978816	
100K	ADWIN	40,00	230	0	160	3999610	0,410256	1,000000	0,640494
	DDM	101,76	48	7	153	3999792	0,761194	0,956250	0,853159
	EDDM	285,44	567	70	90	3999273	0,136986	0,562500	0,277535
	ECDD	9,88	5141	0	160	3994699	0,030183	1,000000	0,173621
	STEPD	10,25	919	0	160	3998921	0,148285	1,000000	0,385034
	EMZD1	33,00	517	0	160	3999323	0,236337	1,000000	0,486113
	EMZD2	38,50	175	0	160	3999665	0,477612	1,000000	0,691080
	FPDD1	18,00	7	0	160	3999833	0,958084	1,000000	0,978817
	FPDD2	18,00	7	0	160	3999833	0,958084	1,000000	0,978817
	FSDD1	18,00	7	0	160	3999833	0,958084	1,000000	0,978817
	FSDD2	15,94	4	0	160	3999836	0,975610	1,000000	0,987729
FTDD	18,00	7	0	160	3999833	0,958084	1,000000	0,978817	

Tabela 26 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados *Mixed*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	40,00	236	2	158	799604	0,401015	0,987500	0,629191
	DDM	33,90	42	6	154	799798	0,785714	0,962500	0,869598
	EDDM	44,24	392	35	125	799448	0,241779	0,781250	0,434439
	ECDD	9,87	1013	4	156	798827	0,133447	0,975000	0,360468
	STEPD	11,00	263	0	160	799577	0,378251	1,000000	0,614920
	EMZD1	13,18	369	3	157	799471	0,298479	0,981250	0,541056
	EMZD2	16,84	143	2	158	799697	0,524917	0,987500	0,719902
	FPDD1	17,06	23	0	160	799817	0,874317	1,000000	0,935036
	FPDD2	17,06	23	0	160	799817	0,874317	1,000000	0,935036
	FSDD1	17,06	24	0	160	799816	0,869565	1,000000	0,932491
	FSDD2	15,75	42	0	160	799798	0,792079	1,000000	0,889965
FTDD	17,06	23	0	160	799817	0,874317	1,000000	0,935036	
50K	ADWIN	32,63	396	0	160	1999444	0,287770	1,000000	0,536389
	DDM	63,48	85	5	155	1999755	0,645833	0,968750	0,790962
	EDDM	106,97	737	84	76	1999103	0,093481	0,475000	0,210592
	ECDD	10,06	2456	2	158	1997384	0,060444	0,987500	0,244158
	STEPD	10,63	649	0	160	1999191	0,197775	1,000000	0,444647
	EMZD1	23,13	774	0	160	1999066	0,171306	1,000000	0,413811
	EMZD2	28,81	307	0	160	1999533	0,342612	1,000000	0,585286
	FPDD1	18,38	41	0	160	1999799	0,796020	1,000000	0,892190
	FPDD2	18,38	41	0	160	1999799	0,796020	1,000000	0,892190
	FSDD1	18,38	42	0	160	1999798	0,792079	1,000000	0,889979
	FSDD2	15,88	63	0	160	1999777	0,717489	1,000000	0,847034
FTDD	18,38	41	0	160	1999799	0,796020	1,000000	0,892190	
100K	ADWIN	40,00	747	0	160	3999093	0,176406	1,000000	0,419968
	DDM	83,38	110	3	157	3999730	0,588015	0,981250	0,759587
	EDDM	287,54	828	91	69	3999012	0,076923	0,431250	0,182064
	ECDD	9,81	5085	1	159	3994755	0,030320	0,993750	0,173471
	STEPD	11,88	1350	0	160	3998490	0,105960	1,000000	0,325460
	EMZD1	28,62	1458	1	159	3998382	0,098330	0,993750	0,312537
	EMZD2	38,62	584	1	159	3999256	0,213997	0,993750	0,461116
	FPDD1	16,88	23	0	160	3999817	0,874317	1,000000	0,935046
	FPDD2	17,94	29	0	160	3999811	0,846561	1,000000	0,920084
	FSDD1	16,81	24	0	160	3999816	0,869565	1,000000	0,932502
	FSDD2	14,69	51	0	160	3999789	0,758294	1,000000	0,870795
FTDD	16,88	23	0	160	3999817	0,874317	1,000000	0,935046	

Tabela 27 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados *Random RBF*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	50,00	523	157	3	799317	0,005703	0,018750	0,009983
	DDM	NaN	101	160	0	799739	0,000000	0,000000	-0,000159
	EDDM	38,00	501	155	5	799339	0,009881	0,031250	0,017224
	ECDD	NaN	0	160	0	799840	0,000000	0,000000	0,000000
	STEPD	54,00	346	155	5	799494	0,014245	0,031250	0,020809
	EMZD1	47,50	118	156	4	799722	0,032787	0,025000	0,028460
	EMZD2	30,00	948	154	6	798892	0,006289	0,037500	0,014879
	FPDD1	46,67	44	157	3	799796	0,063830	0,018750	0,034491
	FPDD2	46,67	44	157	3	799796	0,063830	0,018750	0,034491
	FSDD1	46,67	50	157	3	799790	0,056604	0,018750	0,032467
	FSDD2	46,67	50	157	3	799790	0,056604	0,018750	0,032467
FTDD	46,67	46	157	3	799794	0,061224	0,018750	0,033775	
50K	ADWIN	48,00	752	155	5	1999088	0,006605	0,031250	0,014196
	DDM	NaN	119	160	0	1999721	0,000000	0,000000	-0,000069
	EDDM	30,00	561	156	4	1999279	0,007080	0,025000	0,013156
	ECDD	NaN	0	160	0	1999840	0,000000	0,000000	0,000000
	STEPD	83,33	864	151	9	1998976	0,010309	0,056250	0,023900
	EMZD1	100,00	161	159	1	1999679	0,006173	0,006250	0,006131
	EMZD2	87,35	1708	126	34	1998132	0,019518	0,212500	0,064168
	FPDD1	67,50	67	156	4	1999773	0,056338	0,025000	0,037478
	FPDD2	67,50	67	156	4	1999773	0,056338	0,025000	0,037478
	FSDD1	77,50	72	156	4	1999768	0,052632	0,025000	0,036221
	FSDD2	77,50	72	156	4	1999768	0,052632	0,025000	0,036221
FTDD	67,50	69	156	4	1999771	0,054795	0,025000	0,036960	
100K	ADWIN	286,96	978	137	23	3998862	0,022977	0,143750	0,057380
	DDM	203,33	123	157	3	3999717	0,023810	0,018750	0,021094
	EDDM	106,67	525	157	3	3999315	0,005682	0,018750	0,010250
	ECDD	NaN	0	160	0	3999840	0,000000	0,000000	0,000000
	STEPD	213,33	1735	142	18	3998105	0,010268	0,112500	0,033863
	EMZD1	213,33	262	157	3	3999578	0,011321	0,018750	0,014519
	EMZD2	236,06	2213	127	33	3997627	0,014693	0,206250	0,054916
	FPDD1	218,75	76	152	8	3999764	0,095238	0,050000	0,068980
	FPDD2	218,75	76	152	8	3999764	0,095238	0,050000	0,068980
	FSDD1	228,75	87	152	8	3999753	0,084211	0,050000	0,064860
	FSDD2	228,75	87	152	8	3999753	0,084211	0,050000	0,064860
FTDD	228,75	78	152	8	3999762	0,093023	0,050000	0,068172	

Tabela 28 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados *Random RBF*.

Tamanho	Métodos	DD	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	50,00	385	157	3	799455	0,007732	0,018750	0,011733
	DDM	NaN	80	160	0	799760	0,000000	0,000000	-0,000141
	EDDM	46,67	324	157	3	799516	0,009174	0,018750	0,012834
	ECDD	NaN	0	160	0	799840	0,000000	0,000000	0,000000
	STEPD	47,50	378	156	4	799462	0,010471	0,025000	0,015876
	EMZD1	NaN	120	160	0	799720	0,000000	0,000000	-0,000173
	EMZD2	42,86	581	153	7	799259	0,011905	0,043750	0,022449
	FPDD1	46,67	47	157	3	799793	0,060000	0,018750	0,033434
	FPDD2	46,67	47	157	3	799793	0,060000	0,018750	0,033434
	FSDD1	46,67	53	157	3	799787	0,053571	0,018750	0,031579
	FSDD2	46,67	53	157	3	799787	0,053571	0,018750	0,031579
FTDD	46,67	48	157	3	799792	0,058824	0,018750	0,033102	
50K	ADWIN	170,00	527	157	3	1999313	0,005660	0,018750	0,010158
	DDM	NaN	96	160	0	1999744	0,000000	0,000000	-0,000062
	EDDM	80,00	345	158	2	1999495	0,005764	0,012500	0,008371
	ECDD	NaN	0	160	0	1999840	0,000000	0,000000	0,000000
	STEPD	79,17	890	148	12	1998950	0,013304	0,075000	0,031406
	EMZD1	140,00	138	158	2	1999702	0,014286	0,012500	0,013289
	EMZD2	104,00	999	140	20	1998841	0,019627	0,125000	0,049344
	FPDD1	62,50	71	156	4	1999769	0,053333	0,025000	0,036462
	FPDD2	62,50	71	156	4	1999769	0,053333	0,025000	0,036462
	FSDD1	30,00	80	158	2	1999760	0,024390	0,012500	0,017405
	FSDD2	30,00	80	158	2	1999760	0,024390	0,012500	0,017405
FTDD	62,50	72	156	4	1999768	0,052632	0,025000	0,036221	
100K	ADWIN	306,88	677	144	16	3999163	0,023088	0,100000	0,047972
	DDM	NaN	94	160	0	3999746	0,000000	0,000000	-0,000031
	EDDM	300,00	329	159	1	3999511	0,003030	0,006250	0,004295
	ECDD	NaN	0	160	0	3999840	0,000000	0,000000	0,000000
	STEPD	197,50	1776	136	24	3998064	0,013333	0,150000	0,044598
	EMZD1	316,67	249	157	3	3999591	0,011905	0,018750	0,014891
	EMZD2	242,50	1217	128	32	3998623	0,025620	0,200000	0,071484
	FPDD1	196,67	80	151	9	3999760	0,101124	0,056250	0,075393
	FPDD2	196,67	80	151	9	3999760	0,101124	0,056250	0,075393
	FSDD1	201,11	85	151	9	3999755	0,095745	0,056250	0,073359
	FSDD2	201,11	86	151	9	3999754	0,094737	0,056250	0,072971
FTDD	184,00	79	150	10	3999761	0,112360	0,062500	0,083773	

Tabela 29 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados *SEA*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	NaN	292	160	0	799548	0,000000	0,000000	-0,000270
	DDM	NaN	186	160	0	799654	0,000000	0,000000	-0,000216
	EDDM	44,55	631	149	11	799209	0,017134	0,068750	0,033938
	ECDD	35,54	1050	104	56	798790	0,050633	0,350000	0,132701
	STEPD	45,56	212	115	45	799628	0,175097	0,281250	0,221719
	EMZD1	53,57	464	146	14	799376	0,029289	0,087500	0,050298
	EMZD2	47,14	217	153	7	799623	0,031250	0,043750	0,036748
	FPDD1	NaN	6	160	0	799834	0,000000	0,000000	-0,000039
	FPDD2	NaN	6	160	0	799834	0,000000	0,000000	-0,000039
	FSDD1	NaN	6	160	0	799834	0,000000	0,000000	-0,000039
	FSDD2	70,00	22	159	1	799818	0,043478	0,006250	0,016411
FTDD	NaN	6	160	0	799834	0,000000	0,000000	-0,000039	
50K	ADWIN	149,26	578	133	27	1999262	0,044628	0,168750	0,086642
	DDM	120,00	192	157	3	1999648	0,015385	0,018750	0,016897
	EDDM	70,00	762	159	1	1999078	0,001311	0,006250	0,002688
	ECDD	72,08	2630	64	96	1997210	0,035216	0,600000	0,145136
	STEPD	81,57	429	71	89	1999411	0,171815	0,556250	0,309056
	EMZD1	133,85	582	108	52	1999258	0,082019	0,325000	0,163140
	EMZD2	138,16	272	122	38	1999568	0,122581	0,237500	0,170534
	FPDD1	NaN	18	160	0	1999822	0,000000	0,000000	-0,000027
	FPDD2	NaN	18	160	0	1999822	0,000000	0,000000	-0,000027
	FSDD1	NaN	18	160	0	1999822	0,000000	0,000000	-0,000027
	FSDD2	113,33	43	157	3	1999797	0,065217	0,018750	0,034928
FTDD	NaN	18	160	0	1999822	0,000000	0,000000	-0,000027	
100K	ADWIN	237,37	650	84	76	3999190	0,104683	0,475000	0,222929
	DDM	340,00	178	159	1	3999662	0,005587	0,006250	0,005867
	EDDM	183,33	812	157	3	3999028	0,003681	0,018750	0,008218
	ECDD	109,58	5297	42	118	3994543	0,021791	0,737500	0,126627
	STEPD	126,37	878	58	102	3998962	0,104082	0,637500	0,257527
	EMZD1	216,32	828	84	76	3999012	0,084071	0,475000	0,199765
	EMZD2	238,57	305	83	77	3999535	0,201571	0,481250	0,311417
	FPDD1	373,33	28	157	3	3999812	0,096774	0,018750	0,042581
	FPDD2	373,33	27	157	3	3999813	0,100000	0,018750	0,043285
	FSDD1	373,33	28	157	3	3999812	0,096774	0,018750	0,042581
	FSDD2	283,33	65	148	12	3999775	0,155844	0,075000	0,108088
FTDD	373,33	28	157	3	3999812	0,096774	0,018750	0,042581	

Tabela 30 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados SEA.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	0,00	330	158	2	799510	0,006024	0,012500	0,008392
	DDM	46,32	140	141	19	799700	0,119497	0,118750	0,118947
	EDDM	50,00	542	150	10	799298	0,018116	0,062500	0,033292
	ECDD	36,14	1023	103	57	798817	0,052778	0,356250	0,136707
	STEPD	45,95	271	123	37	799569	0,120130	0,231250	0,166445
	EMZD1	45,11	519	113	47	799321	0,083039	0,293750	0,155876
	EMZD2	48,82	229	126	34	799611	0,129278	0,212500	0,165533
	FPDD1	NaN	5	160	0	799835	0,000000	0,000000	-0,000035
	FPDD2	NaN	5	160	0	799835	0,000000	0,000000	-0,000035
	FSDD1	NaN	5	160	0	799835	0,000000	0,000000	-0,000035
	FSDD2	NaN	16	160	0	799824	0,000000	0,000000	-0,000063
FTDD	NaN	5	160	0	799835	0,000000	0,000000	-0,000035	
50K	ADWIN	155,00	508	144	16	1999332	0,030534	0,100000	0,055123
	DDM	100,00	214	153	7	1999626	0,031674	0,043750	0,037135
	EDDM	114,58	931	136	24	1998909	0,025131	0,150000	0,061219
	ECDD	71,16	2560	65	95	1997280	0,035782	0,593750	0,145534
	STEPD	82,42	662	94	66	1999178	0,090659	0,412500	0,193255
	EMZD1	112,00	1177	90	70	1998663	0,056135	0,437500	0,156545
	EMZD2	126,13	587	98	62	1999253	0,095532	0,387500	0,192280
	FPDD1	NaN	12	160	0	1999828	0,000000	0,000000	-0,000022
	FPDD2	NaN	13	160	0	1999827	0,000000	0,000000	-0,000023
	FSDD1	NaN	12	160	0	1999828	0,000000	0,000000	-0,000022
	FSDD2	136,67	45	157	3	1999795	0,062500	0,018750	0,034191
FTDD	NaN	12	160	0	1999828	0,000000	0,000000	-0,000022	
100K	ADWIN	253,24	661	92	68	3999179	0,093278	0,425000	0,199043
	DDM	220,00	252	155	5	3999588	0,019455	0,031250	0,024608
	EDDM	160,67	1151	130	30	3998689	0,025402	0,187500	0,068917
	ECDD	111,54	5171	43	117	3994669	0,022126	0,731250	0,127055
	STEPD	126,44	1370	73	87	3998470	0,059712	0,543750	0,180105
	EMZD1	167,33	2112	74	86	3997728	0,039126	0,537500	0,144913
	EMZD2	195,43	961	79	81	3998879	0,077735	0,506250	0,198305
	FPDD1	336,67	20	157	3	3999820	0,130435	0,018750	0,049440
	FPDD2	336,67	21	157	3	3999819	0,125000	0,018750	0,048398
	FSDD1	336,67	20	157	3	3999820	0,130435	0,018750	0,049440
	FSDD2	197,06	50	143	17	3999790	0,253731	0,106250	0,164171
FTDD	336,67	20	157	3	3999820	0,130435	0,018750	0,049440	

Tabela 31 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando NB no Conjunto de Dados *Sine*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	40,31	141	0	160	799699	0,531561	1,000000	0,729018
	DDM	50,18	118	46	114	799722	0,491379	0,712500	0,591603
	EDDM	38,66	1034	48	112	798806	0,097731	0,700000	0,261235
	ECDD	9,94	1297	2	158	798543	0,108591	0,987500	0,327193
	STEPD	13,69	216	0	160	799624	0,425532	1,000000	0,652240
	EMZD1	20,40	718	9	151	799122	0,173763	0,943750	0,404750
	EMZD2	25,51	224	13	147	799616	0,396226	0,918750	0,603247
	FPDD1	19,19	3	0	160	799837	0,981595	1,000000	0,990753
	FPDD2	20,19	3	0	160	799837	0,981595	1,000000	0,990753
	FSDD1	18,50	3	0	160	799837	0,981595	1,000000	0,990753
	FSDD2	17,50	5	0	160	799835	0,969697	1,000000	0,984729
FTDD	18,94	3	0	160	799837	0,981595	1,000000	0,990753	
50K	ADWIN	41,44	192	0	160	1999648	0,454545	1,000000	0,674167
	DDM	88,26	137	45	115	1999703	0,456349	0,718750	0,572672
	EDDM	79,00	1650	60	100	1998190	0,057143	0,625000	0,188808
	ECDD	9,87	3396	2	158	1996444	0,044457	0,987500	0,209344
	STEPD	14,62	589	2	158	1999251	0,211513	0,987500	0,456952
	EMZD1	29,30	1347	18	142	1998493	0,095366	0,887500	0,290801
	EMZD2	34,63	458	13	147	1999382	0,242975	0,918750	0,472411
	FPDD1	20,13	2	0	160	1999838	0,987654	1,000000	0,993807
	FPDD2	20,63	2	0	160	1999838	0,987654	1,000000	0,993807
	FSDD1	19,44	2	0	160	1999838	0,987654	1,000000	0,993807
	FSDD2	18,06	5	0	160	1999835	0,969697	1,000000	0,984731
FTDD	19,69	2	0	160	1999838	0,987654	1,000000	0,993807	
100K	ADWIN	40,38	162	0	160	3999678	0,496894	1,000000	0,704893
	DDM	137,39	157	41	119	3999683	0,431159	0,743750	0,566260
	EDDM	173,30	2054	66	94	3997786	0,043762	0,587500	0,160243
	ECDD	10,20	6955	9	151	3992885	0,021250	0,943750	0,141475
	STEPD	13,19	1148	0	160	3998692	0,122324	1,000000	0,349699
	EMZD1	50,60	2136	26	134	3997704	0,059031	0,837500	0,222264
	EMZD2	49,80	750	13	147	3999090	0,163880	0,918750	0,387983
	FPDD1	18,94	3	0	160	3999837	0,981595	1,000000	0,990754
	FPDD2	19,88	3	0	160	3999837	0,981595	1,000000	0,990754
	FSDD1	18,31	6	0	160	3999834	0,963855	1,000000	0,981761
	FSDD2	17,00	12	0	160	3999828	0,930233	1,000000	0,964484
FTDD	18,56	3	0	160	3999837	0,981595	1,000000	0,990754	

Tabela 32 – Análise das Detecções de Mudanças de Conceitos Abruptas, Utilizando HT no Conjunto de Dados *Sine*.

Tamanho	Métodos	DM	FP	FN	VP	VN	Precision	Recall	MCC
20K	ADWIN	40,19	263	2	158	799577	0,375297	0,987500	0,608671
	DDM	48,11	112	1	159	799728	0,586716	0,993750	0,763522
	EDDM	32,56	907	78	82	798933	0,082912	0,512500	0,205787
	ECDD	10,25	1314	3	157	798526	0,106730	0,981250	0,323342
	STEPD	12,83	240	1	159	799600	0,398496	0,993750	0,629194
	EMZD1	24,05	628	2	158	799212	0,201018	0,987500	0,445359
	EMZD2	31,46	159	3	157	799681	0,496835	0,981250	0,698153
	FPDD1	18,38	7	0	160	799833	0,958084	1,000000	0,978813
	FPDD2	19,94	6	0	160	799834	0,963855	1,000000	0,981758
	FSDD1	18,38	7	0	160	799833	0,958084	1,000000	0,978813
	FSDD2	17,00	19	0	160	799821	0,893855	1,000000	0,945428
FTDD	18,38	7	0	160	799833	0,958084	1,000000	0,978813	
50K	ADWIN	41,00	617	0	160	1999223	0,205920	1,000000	0,453714
	DDM	70,06	165	0	160	1999675	0,492308	1,000000	0,701617
	EDDM	79,88	1133	78	82	1998707	0,067490	0,512500	0,185823
	ECDD	10,27	3293	10	150	1996547	0,043567	0,937500	0,201909
	STEPD	11,84	489	2	158	1999351	0,244204	0,987500	0,491010
	EMZD1	35,38	903	2	158	1998937	0,148916	0,987500	0,383388
	EMZD2	43,90	175	1	159	1999665	0,476048	0,993750	0,687772
	FPDD1	17,56	11	0	160	1999829	0,935673	1,000000	0,967299
	FPDD2	18,63	8	0	160	1999832	0,952381	1,000000	0,975898
	FSDD1	17,56	11	0	160	1999829	0,935673	1,000000	0,967299
	FSDD2	15,69	18	0	160	1999822	0,898876	1,000000	0,948087
FTDD	17,56	11	0	160	1999829	0,935673	1,000000	0,967299	
100K	ADWIN	40,31	1281	0	160	3998559	0,111034	1,000000	0,333164
	DDM	91,31	133	0	160	3999707	0,546075	1,000000	0,738957
	EDDM	139,01	1258	79	81	3998582	0,060493	0,506250	0,174916
	ECDD	14,70	6658	11	149	3993182	0,021889	0,931250	0,142637
	STEPD	11,82	884	1	159	3998956	0,152445	0,993750	0,389176
	EMZD1	48,44	912	0	160	3998928	0,149254	1,000000	0,386290
	EMZD2	55,31	178	0	160	3999662	0,473373	1,000000	0,688006
	FPDD1	17,06	10	0	160	3999830	0,941176	1,000000	0,970141
	FPDD2	17,63	10	0	160	3999830	0,941176	1,000000	0,970141
	FSDD1	17,06	10	0	160	3999830	0,941176	1,000000	0,970141
	FSDD2	15,44	24	0	160	3999816	0,869565	1,000000	0,932502
FTDD	17,06	10	0	160	3999830	0,941176	1,000000	0,970141	