



Pós-Graduação em Ciência da Computação

ERAYLSON GALDINO DA SILVA

PREVISÃO DE SÉRIES TEMPORAIS USANDO SISTEMAS DE MÚLTIPLOS PREDITORES



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

RECIFE
2017

Eraylson Galdino da Silva

Previsão de Séries Temporais usando sistemas de múltiplos preditores

Este trabalho foi apresentado à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre Profissional em Ciência da Computação.

ORIENTADOR: Paulo Salgado G. de Mattos Neto
CO-ORIENTADOR: George D. da Cunha Cavalcanti

RECIFE
2017

Catálogo na fonte
Bibliotecário Jefferson Luiz Alves Nazareno, CRB4-1758

S586p Silva, Eraylson Galdino da.
Previsão de séries temporais usando sistemas de múltiplos preditores /
Eraylson Galdino da Silva. – 2017.
104 f.: il., fig., tab.

Orientador: Paulo Salgado Gomes Mattos Neto.
Dissertação (Mestrado Profissional) – Universidade Federal de
Pernambuco. CIn, Ciência da Computação, Recife, 2017.
Inclui referências.

1. Ciência da computação. 2. Aprendizado por máquina. 3. Seleção de
preditores. I. Mattos Neto, Paulo Salgado Gomes(orientador). II. Título.

004

CDD (23. ed.)

UFPE- MEI 2017-197

Eraylson Galdino da Silva

**Previsão de Séries Temporais Usando Sistemas de Múltiplos
Preditores**

Dissertação 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: 14/06/2017

BANCA EXAMINADORA

Prof. Dr. Paulo Salgado Gomes de Mattos Neto
Centro de Informática / UFPE

Prof. Dr. João Fausto Lorenzato de Oliveira
Faculdade de Ciências e Tecnologia de Garanhuns / UPE

Prof. Dr. Jarley Palmeira Nóbrega
Centro de Tecnologias Estratégicas do Nordeste / MCTIC

*Dedico esta Dissertação aos meus Pais, e toda a minha
Família.*

Agradecimentos

Agradeço primeiramente a Deus, por ter me dado saúde e perseverança para realizar esse objetivo.

Agradeço à minha Família, especialmente aos meus pais, Quitéria Celeste da Silva e Paulo Galdino da Silva, aos meus irmãos e irmãs, e aos meus sobrinhos, por terem me apoiado em todos os momentos, por terem me educado e me possibilitado momentos de felicidades.

Agradeço ao meu orientador Prof. Paulo Salgado Gomes de Mattos Neto e ao meu co-orientador Prof. George Darmiton da Cunha Cavalcanti, por toda a paciência, disponibilidade, compreensão e apoio que foram fundamentais para a execução e conclusão deste trabalho.

Agradeço aos Professores do Centro de Informática da UFPE, por todo o conhecimento compartilhando, que foram fundamentais para minha formação como Mestre. E aos professores do curso de Licenciatura em Computação na UPE, que me guiaram para o caminho da pesquisa.

E por fim, agradeço aos meus amigos pesquisadores do Centro de Informática da UFPE, que estiveram presente em momentos de discussões e descontração que contribuíram na minha jornada no Mestrado.

Se queres prever o futuro, estuda o passado.

—CONFÚCIO

Resumo

Sistema de Múltiplos Preditores (SMP) consiste no desenvolvimento de um conjunto de modelos para prever a mesma variável. Na literatura é possível encontrar trabalhos que mostram que a sua utilização em previsão de séries temporais pode melhorar a acurácia. Para tal, é construído um conjunto com diferentes preditores e a previsão final é obtida através da seleção ou combinação. A seleção de preditores consiste em encontrar um preditor dentro do conjunto que possa retornar a melhor previsão, podendo reduzir o custo computacional e aumentar a acurácia do sistema de previsão. Além da abordagem para retornar a previsão final é necessário que exista diversidade entre os preditores, a qual, consiste em desenvolver um conjunto de preditores que apresentem previsões diferentes para a mesma variável. Uma das técnicas para criar modelos diversos é particionar a série em diferentes amostras e treinar um modelo para cada amostra, possibilitando o desenvolvimento de preditores especialistas em padrões diferentes da série. Sabendo que a utilização de SMP pode aumentar a acurácia de sistemas de previsão de séries temporais, ainda existe a necessidade de pesquisas que apresentem em que circunstâncias a combinação é melhor que a seleção de preditores, além disso, quais as melhores formas de combinação e qual a influência de treinar os preditores através de partições diferentes da série. Para tal, esta dissertação realiza uma comparação entre diferentes abordagens de combinação: Média Simples, Mediana, Regressão Linear, Redes Neurais e Máquina de Vetor de Suporte; e uma abordagem de seleção de preditores que utiliza medida de distância entre os padrões para selecionar o melhor modelo, com o intuito de responder esses questionamentos. Para avaliar o desempenho das diferentes abordagens, foram utilizadas quatro séries: Mackey Glass, Laser, Microsoft e Goldman Sachs. Em cada série os desempenhos das arquiteturas foram avaliados através de um conjunto de métricas. Com base nos resultados obtidos, é possível inferir que em média os desempenhos apresentados pelas melhores abordagens de combinação: Redes Neurais, Regressão Linear e Média Simples, são melhores que a abordagem de seleção utilizada. Além disso, a relação entre a acurácia e a diversidade dos modelos pode variar de acordo com a abordagem de combinação utilizada.

Palavras-chave: Análise e Previsão de Séries Temporais. Combinação de Preditores. *Ensemble* de Preditores. Seleção de Preditores. Aprendizado de Máquina.

Abstract

Multiple Predictor Systems (MPS) consists in development a group of models for forecast same variable. According to some works in the literature, MPS can improve in time series forecast. For that, building a group with several models for same forecasting variable and the final forecast is to get through a combination or selection model. The selection of forecasts consists in finding the better model inside of a group that can improve efficiency and performance. The motivation for selection decreases the cost and to use a specific model for each pattern of the time series. Beyond the approach to get the final forecast of Ensemble, it is necessary that exist diversity among models forecasting. For that, one approach is to split the time series in different samples and train a model for each sample, thus making it possible to build off experts' models. In the knowledge that time series forecast can be better through MPS, it still exists the necessity for new research to find answers to some questions. For instance, in which circumstances are combination approaches better than selection? Which are the better approaches to combining? How important is it to train the models through different partitions of the series? For that, in present theses are developed with diverse approaches of combination: Mean Simple, Median, Linear Regression, Neural Networks and Support Vector Machine, and an approach to forecasting selection that use a distance measure between patterns for select the model. To verify the feasibility between different approaches, four series are used: Mackey-Glass, Laser, Microsoft, and Goldman Sachs. For each time series, the performances of the architectures were evaluated through several metrics. Based on the results obtained, it is possible to infer that on average, the performances presented by better combination approaches: Neural Network, Linear Regression, and Simple Mean are better than selection approach. Moreover, the relationship between accuracy and diversity for combination approach vary according to approach used.

Keywords: Analysis and Forecasting of Time Series. Combination of predictors. Ensemble of Predictors. Selection of Predictors. Machine Learning.

Lista de Figuras

2.1	Volume de vendas do comércio varejista de Pernambuco	23
2.2	Correlograma para os 20 primeiros pontos da Série de Volume de Vendas	24
2.3	Etapas para a construção de um modelo ARIMA	29
2.4	Rede Neural Artificial MLP com uma camada escondida	30
2.5	Representação de uma SVR com margens suaves	33
3.1	Etapas de treinamento da arquitetura de seleção	40
3.2	Exemplo de particionamento da Série	41
3.3	Exemplo dos lags selecionados através da função de auto correlação	42
3.4	Atividades de Teste da arquitetura de seleção	43
3.5	Fase de Treinamento dos Modelos de Previsão	45
3.6	Fase de Treinamento dos Modelos de Combinação	46
3.7	Fase de Teste dos Modelos de Combinação	48
4.1	Série Laser	51
4.2	Série Mackey-Glass	52
4.3	Série Microsoft	53
4.4	Série Goldman Sachs	54
5.1	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Goldman Sachs	62
5.2	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Goldman Sachs	63
5.3	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Goldman Sachs	64
5.4	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com <i>Perceptron</i> na série Goldman Sachs	65
5.5	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Goldman Sachs	66

5.6	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Mediana na série Goldman Sachs	67
5.7	Previsões das Arquiteturas para o conjunto de teste da série Goldman Sachs	69
5.8	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Microsoft	70
5.9	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Microsoft	71
5.10	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Microsoft	72
5.11	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com <i>Perceptron</i> na série Microsoft	73
5.12	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Microsoft	74
5.13	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Mediana na série Microsoft.	75
5.14	Previsões de Teste para a série Microsoft	77
5.15	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Laser	78
5.16	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Laser	79
5.17	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Laser	80
5.18	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com <i>Perceptron</i> na série Laser	81
5.19	Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Laser	82

5.20 Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Mediana na série Laser	83
5.21 Previsões das Arquiteturas: Combinação x Seleção para a série Laser	85
5.22 Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Mackey Glass	86
5.23 Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Mackey Glass	87
5.24 Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Mackey Glass	88
5.25 Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com <i>Perceptron</i> na série Mackey Glass	89
5.26 Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Mackey Glass	90
5.27 Os <i>Box-Plots</i> representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação por Mediana na série Mackey Glass	91
5.28 Previsões para a Série Mackey-Glass	92

Lista de Tabelas

4.1	Estatísticas da série Laser	50
4.2	Estatísticas da série Mackey-Glass	51
4.3	Estatísticas da série Microsoft	52
4.4	Estatísticas da série Goldman Sachs	54
4.5	Parâmetros para as partições das Séries	58
4.6	Parâmetros para a Rede MLP	58
4.7	Parâmetros para a SVR	59
5.1	Melhores taxas para as abordagens de combinação para Série Goldman Sachs	64
5.2	Resultados das abordagens de combinação para a série Goldman Sachs	65
5.3	Resultados dos Testes para as comparações par a par do <i>fitness</i> na série Goldman Sachs	66
5.4	Comparação entre as abordagens de combinação e a seleção para a série Goldman Sachs	68
5.5	Ranking dos melhores desempenhos médio para a série Goldman Sachs	68
5.6	Resultados dos testes estatísticos entre as abordagens de seleção e combinação para a série Goldman Sachs	68
5.7	Comparação entre as diferentes abordagens para a série Goldman Sachs	68
5.8	Melhores taxas para as abordagens de combinação Série Microsoft	72
5.9	Comparação entre as diferentes abordagens de combinação para a série Microsoft	74
5.10	Resultados dos Testes para o <i>fitness</i> na série Microsoft	74
5.11	Comparação entre as abordagens de combinação e a seleção para a série Microsoft	76
5.12	Ranking dos melhores desempenhos médio para a série Microsoft	76
5.13	Resultados do Teste para a série Microsoft	76
5.14	Comparação entre as diferentes abordagens para a série Microsoft	77
5.15	Melhores taxas para as abordagens de combinação Laser	80
5.16	Comparação entre as diferentes abordagens de combinação para a série Laser	82
5.17	Resultados dos Testes para o <i>Fitness</i> na série Laser	82
5.18	Comparação entre as abordagens de combinação e a seleção para a série Laser	84
5.19	Ranking dos melhores desempenhos médio para a série Laser	84
5.20	Resultados do Teste para a série Laser	84
5.21	Comparação entre as diferentes abordagens para a série Laser	84
5.22	Melhores taxas para as abordagens de combinação Série Mackey-Glass	87
5.23	Comparação entre as diferentes abordagens de combinação para a série Mackey Glass	88
5.24	Resultados dos Testes para o <i>Fitness</i> em relação a série Mackey Glass	88
5.25	Comparação entre as abordagens de combinação e seleção para a série Mackey Glass	90

5.26	Ranking dos melhores desempenhos médio para a série Mackey Glass	91
5.27	Resultados do Teste para a série Mackey-Glass	91
5.28	Comparação entre as diferentes abordagens para a série Mackey-Glass	92

Lista de Acrônimos

AIC	<i>Akaike Information Criterion</i>	35
ARIMA	<i>Auto-Regressive Integrated Moving Average</i>	18
ARV	<i>Average Relative Variance</i>	54
DTW	<i>Dynamic Time Warping</i>	42
ERLS	<i>Equality Restricted Least Squares</i>	35
FAC	Função de Autocorrelação	28
GARCH	Modelo Generalizado Autorregressivo para a Heteroscedasticidade Condicional	18
IRLS	<i>Iteratively Reweighted Least-Squares</i>	59
LASSO	<i>Least Absolute Shrinkage and Selector Operator</i>	35
MAPE	Erro Médio Absoluto em Percentual	53
MLP	<i>Multilayer Perceptron</i>	18
MSE	<i>Mean Squared Error</i>	54
NRLS	<i>Nonnegativity Restricted Least Squares</i>	35
POCID	<i>Prediction of Change in Direction</i>	53
RBF	<i>Radial Basis Function</i>	33
RLM	Regressão Linear Múltipla	47
RMSE	<i>Root Mean Squared Error</i>	54
RNA	Redes Neurais Artificiais	18
SVR	<i>Support Vector Regression</i>	18

Sumário

1	INTRODUÇÃO	18
1.1	Motivação	18
1.2	Objetivos	20
1.3	Estrutura da Dissertação	20
2	CONCEITOS E TÉCNICAS	22
2.1	Séries Temporais	22
2.1.1	Características das Séries Temporais	23
2.2	Previsão de Séries Temporais	25
2.2.1	Tratamento dos Dados	25
2.2.2	Modelagem de Série Temporal	26
2.2.3	Previsão	27
2.3	Modelos para Previsão de Séries Temporais	27
2.3.1	Modelos Estatísticos	27
2.3.2	Modelos de Aprendizado de Máquina	29
2.3.2.1	<i>Redes Neurais Artificiais</i>	29
2.3.2.2	<i>Máquina de Vetor de Suporte para Regressão</i>	32
2.4	Sistemas de Múltiplos Preditores	34
2.4.1	Combinação de Preditores	34
2.4.2	Seleção de Preditores	37
2.5	Considerações Finais	38
3	MÉTODO PROPOSTO	39
3.1	Arquitetura de Partição e Seleção de Preditores	39
3.1.1	Treinamento	39
3.1.1.1	<i>Particionamento</i>	39
3.1.1.2	<i>Seleção Lags</i>	41
3.1.1.3	<i>Seleção dos Dados de Validação</i>	42
3.1.1.4	<i>Treinamento dos preditores</i>	42
3.1.2	Teste	42
3.1.2.1	<i>Seleção do Modelo</i>	43
3.1.2.2	<i>Realização da Previsão</i>	43
3.1.3	Aprimoramento da Arquitetura	43

3.2	Arquitetura Proposta	44
3.2.1	Fase de Treinamento	44
3.2.1.1	<i>Treinamento dos Modelos de Previsão</i>	45
3.2.1.2	<i>Treinamento dos Modelos de Combinação</i>	46
3.2.2	Fase de Teste	48
3.2.2.1	<i>Previsão do Ensemble</i>	48
3.2.2.2	<i>Combinação</i>	48
4	METODOLOGIA DOS EXPERIMENTOS	50
4.1	Base de Dados	50
4.1.1	Série Laser	50
4.1.2	Série Mackey-Glass	51
4.1.3	Série Microsoft	52
4.1.4	Série Goldman Sachs	53
4.2	Métricas de Desempenho	53
4.3	Testes Estatísticos	56
4.3.1	Análise de Normalidade	56
4.3.2	Análise de Distribuições	56
4.4	Metodologia para a Análise Comparativa	57
4.4.1	Configuração dos Dados	57
4.4.2	Configuração para os Modelos	58
4.4.3	Processo de Comparação	59
4.5	Considerações Finais	59
5	ANÁLISE E RESULTADOS	61
5.1	Série Goldman Sachs	61
5.1.1	Interseção entre as Partições	61
5.1.2	Análise das Abordagens de Combinação	65
5.1.3	Escolha da Arquitetura	67
5.2	Série Microsoft	69
5.2.1	Interseção entre as Partições	70
5.2.2	Análise das Abordagens de Combinação	73
5.2.3	Escolha da Arquitetura	75
5.3	Série Laser	78
5.3.1	Interseção entre as Partições	78
5.3.2	Análise das Abordagens de Combinação	81
5.3.3	Escolha da Arquitetura	83
5.4	Série Mackey Glass	85
5.4.1	Interseção entre as Partições	85
5.4.2	Análise das Abordagens de Combinação	87

5.4.3	Escolha da Arquitetura	89
5.5	Considerações Finais	93
6	CONCLUSÃO E TRABALHOS FUTUROS	94
6.1	Resumo	94
6.2	Questões de Pesquisa	95
6.2.1	Quando a combinação é melhor que a seleção?	95
6.2.2	Qual a relação entre a diversidade nos dados e as abordagens de combinação?	96
6.2.3	Quais as melhores abordagens de combinação?	96
6.3	Contribuições	96
6.4	Limitações	97
6.5	Trabalhos Futuros	97
	REFERÊNCIAS	98

1

INTRODUÇÃO

Este capítulo apresenta uma introdução ao contexto da pesquisa realizada e está dividido em 3 seções. Na primeira seção é introduzida a motivação do trabalho. Na segunda seção os objetivos são apresentados. E por fim, a estrutura da dissertação é apresentada.

1.1 Motivação

Série temporal é um conjunto de observações ordenadas no tempo sobre uma determinada variável de interesse e a atividade de previsão consiste em estimar os valores futuros dessa variável (OLIVEIRA; TORGO, 2014). Essa atividade vem se tornando comum em diversas áreas, tais como: meteorologia (ZHANG et al., 2016), demanda de produtos (LIU et al., 2015), mercado financeiro (ZHANG et al., 2016), consumo de tráfego (TANG; DU; JI, 2016) e em concentrações de poluentes na atmosfera (MATTOS NETO et al., 2014).

Tanto na academia como no mercado, os modelos Estatísticos e os de Aprendizado de Máquina são os mais utilizados para previsão de séries temporais. Dentre os modelos estatísticos, os mais conhecidos são: *Auto-Regressive Integrated Moving Average* (ARIMA) (BOX et al., 2015) e o Modelo Generalizado Autorregressivo para a Heteroscedasticidade Condicional (GARCH) (BOLLERSLEV, 1986). Dentre os modelos de aprendizado de máquina, os mais utilizados são: as Redes Neurais Artificiais (RNA) do tipo *Multilayer Perceptron* (MLP) (ZHANG; PATUWO; HU, 1998), o *Support Vector Regression* (SVR) (SANTAMARÍA-BONFIL; REYES-BALLESTEROS; GERSHENSON, 2016), Redes Neurais Recorrentes (GUO et al., 2016), *Extreme Learning Machine* (GUO; LIEW, 2016) e Redes Fuzzy (DAS; BISOI; DASH, 2015).

O modelo estatístico mais utilizado, o ARIMA, tem como particularidade a simplicidade na compreensão e implementação. Porém, é um modelo linear e de forma geral este fato pode vir a ser um limitador, já que séries do mundo real geralmente apresentam padrões não lineares (ZHANG, 2003). Já as RNAs, segundo (KHASHEI; BIJARI, 2010) são mais atrativas para problemas de previsões, por diversos fatores: (i) são auto adaptativas e orientadas a dados, facilitando sua modelagem para problemas diferentes; (ii) têm capacidade de generalização, podendo inferir para padrões não apresentados na etapa de treinamento; (iii) são aproximadores

universais de funções, de forma que a RNA contendo a configuração correta pode aproximar qualquer função (HORNİK; STINCHCOMBE; WHITE, 1989).

Segundo (WICHARD; OGORZALEK, 2004) uma prática de aprendizado de máquina que vem sendo utilizada para melhorar a acurácia em sistemas de previsão é a utilização de um Sistema de Múltiplos Preditores (SMP), a qual consiste em criar um conjunto de modelos preditores e retornar uma previsão final com base em seleção ou combinação dos modelos. Na seleção de preditores, apenas um modelo será escolhido para realizar a previsão final. Enquanto que na combinação, a previsão de todos os modelos será utilizada para compor a previsão final.

No contexto de construção de um sistema de múltiplos preditores, o trabalho de (NOVA FILHO; PESSOA, 2015) apresenta uma nova arquitetura para previsão, a qual divide a série em diferentes partições e treina um modelo para cada partição e por fim, seleciona um único modelo para realizar a previsão final. Essa abordagem tem o objetivo de desenvolver preditores especializados para cada comportamento encontrado na série temporal. Os resultados obtidos mostram que essa arquitetura tem um desempenho mais acurado que os modelos MLP e SVR utilizados individualmente.

A seleção de preditores é muitas vezes uma abordagem que pode apresentar variações no seu desempenho e dependendo da diversidade entre os modelos, uma mudança de comportamento nos dados pode acarretar na queda de desempenho do sistema de previsão (ZOU; YANG, 2004). Uma alternativa a essa abordagem é a combinação de preditores, a qual segundo (HIBON; EVGENIOU, 2005) reduz o risco em comparação com a seleção de apenas um preditor, pois os resultados obtidos através da combinação apresentam uma variância menor, se comparado com os resultados da seleção.

Dessa forma, vem sendo cada vez mais discutida a utilização da combinação de preditores, onde a previsão final é obtida através da composição de diferentes previsões. Dentre as técnicas de combinações, as mais utilizadas são as abordagens lineares: média simples, mediana e combinação linear ponderada (MENDES-MOREIRA et al., 2012). A média simples e a mediana podem ser influenciadas por *outliers*, acarretando no mal desempenho do sistema de previsão. Além disso, essas abordagens não levam em consideração o comportamento individual de cada modelo como mostra (JOSE; WINKLER, 2008). Enquanto que para a combinação linear ponderada, se os seus pesos forem mal ajustados, podem incorporar erros à previsão final (TIMMERMANN, 2006).

Outra forma de combinar as previsões dos modelos, é através de modelos não lineares (ADHIKARI; AGRAWAL, 2012) como MLP e SVR. No trabalho de (WIDODO; BUDI, 2011) vários preditores do tipo MLP são treinados utilizando o mesmo conjunto de treinamento, e no final suas previsões são combinadas através de uma Rede Neural do tipo MLP. Mesmo apresentando resultados melhores que a utilização de apenas um modelo, essa arquitetura de combinação poderá ter desempenho superior, caso os modelos sejam treinados com dados diferentes, semelhante ao trabalho de (NOVA FILHO; PESSOA, 2015).

Contudo, com base no contexto atual de sistemas de previsão de séries temporais, é

possível destacar que a utilização de um conjunto de preditores pode melhorar a acurácia do sistema. Dessa forma, o presente trabalho busca aprimorar a arquitetura desenvolvida por (NOVA FILHO; PESSOA, 2015) através da combinação de preditores e com isso responder as seguintes questões importantes para a escolha da abordagem a ser utilizada no desenvolvimento de um SMP: em quais circunstâncias a combinação é melhor que a seleção de preditores? quais as melhores formas de combinação, seja por abordagens lineares ou não lineares? e qual a importância de treinar os preditores através de amostras diferentes?.

1.2 Objetivos

O objetivo principal deste trabalho de Mestrado é investigar a viabilidade da combinação de preditores em relação a seleção, respondendo as seguintes questões:

1. Quando a combinação é melhor que a seleção?
2. Quais as melhores abordagens de combinação?
3. Qual a relação entre a diversidade nos dados e as abordagens de combinação?

A fim de atingir o objetivo principal da pesquisa, os seguintes objetivos específicos foram definidos:

1. Comparar os desempenhos obtidos entre as diferentes abordagens de combinação e as diferentes arquiteturas;
2. Investigar em que circunstâncias a seleção e a combinação apresentam melhores desempenhos;
3. Investigar quais as melhores abordagens de combinação;
4. Investigar a relação entre a diversidade de dados e o desempenho das abordagens de combinação;

1.3 Estrutura da Dissertação

A estrutura desta dissertação está dividida em 6 capítulos:

1. **Introdução:** apresenta de forma introdutória sobre o contexto de previsões de séries temporais, bem como a motivação e os objetivos da realização deste trabalho.
2. **Conceitos e técnicas:** apresenta os conceitos e técnicas fundamentais para o embasamento teórico desta dissertação;
3. **Método Proposto:** apresenta as etapas que compõem a arquitetura proposta;

4. **Metodologia dos Experimentos:** apresenta as séries temporais utilizadas, destacando suas características e comportamentos, os parâmetros utilizados de cada modelo, as métricas de avaliação, e o processo de execução dos experimentos;
5. **Resultados Experimentais:** os experimentos realizados são descritos, e seus resultados são apresentados e discutidos;
6. **Conclusões:** apresenta um resumo deste trabalho junto com as conclusões obtidas com base nos experimentos, além das limitações e propostas para trabalhos futuros;

2

CONCEITOS E TÉCNICAS

Neste capítulo são apresentados os conceitos e as técnicas utilizadas na pesquisa. A seção 2.1 apresenta o conceito e as características de séries temporais. A seção 2.2 descreve as etapas para a previsão de séries temporais. Na seção 2.3 os modelos de Aprendizado de Máquina são apresentados, com destaque para as RNAs do tipo MLP e SVR. Na seção 2.3 o conceito de Sistemas de Múltiplos Preditores e suas principais abordagens são descritas. E por fim, na seção 2.4 são apresentadas as considerações finais do capítulo.

2.1 Séries Temporais

Séries temporais podem ser representadas de forma contínua, onde as observações são realizadas continuamente no tempo através de intervalos, ou de forma discreta, onde as observações são realizadas em tempos específicos (BROCKWELL; DAVIS, 2013). No presente trabalho são utilizadas séries discretas e podem ser representadas da seguinte forma:

$$Z_t = \{z_t, t = 1, \dots, N\}, \quad (2.1)$$

tal que, t é o tempo atual da série, também chamado de ponto atual da série ou índice temporal de Z , e N é o número de observações. Alguns exemplos de séries temporais são:

- Valores diário da cotação de moedas;
- Índices de Bolsa de Valores;
- Histórico da temperatura local ou global;
- Histórico da quantidade de passageiros em aeroportos ou rodoviárias;
- Taxas de desemprego;
- Valores médios de poluentes na atmosfera;
- Produção de eletricidade;

2.1.1 Características das Séries Temporais

As principais características da maioria das séries temporais são a tendência e a sazonalidade (COWPERTWAIT; METCALFE, 2009). Tendência é uma mudança sistemática nos valores de forma crescente ou decrescente na série em relação ao tempo. Sazonalidade é um comportamento periódico e recorrente (ZHANG; QI, 2005).

Segundo (MAHALAKSHMI; SRIDEVI; RAJARAM, 2016), principalmente para séries da área de negócios e economia, é importante identificar o comportamento de tendência e sazonalidade. Para tal, existem métodos simples, que removem ou ajustam a sazonalidade e a tendência das séries. Outra característica importante é a correlação entre as observações no decorrer do tempo (COWPERTWAIT; METCALFE, 2009).

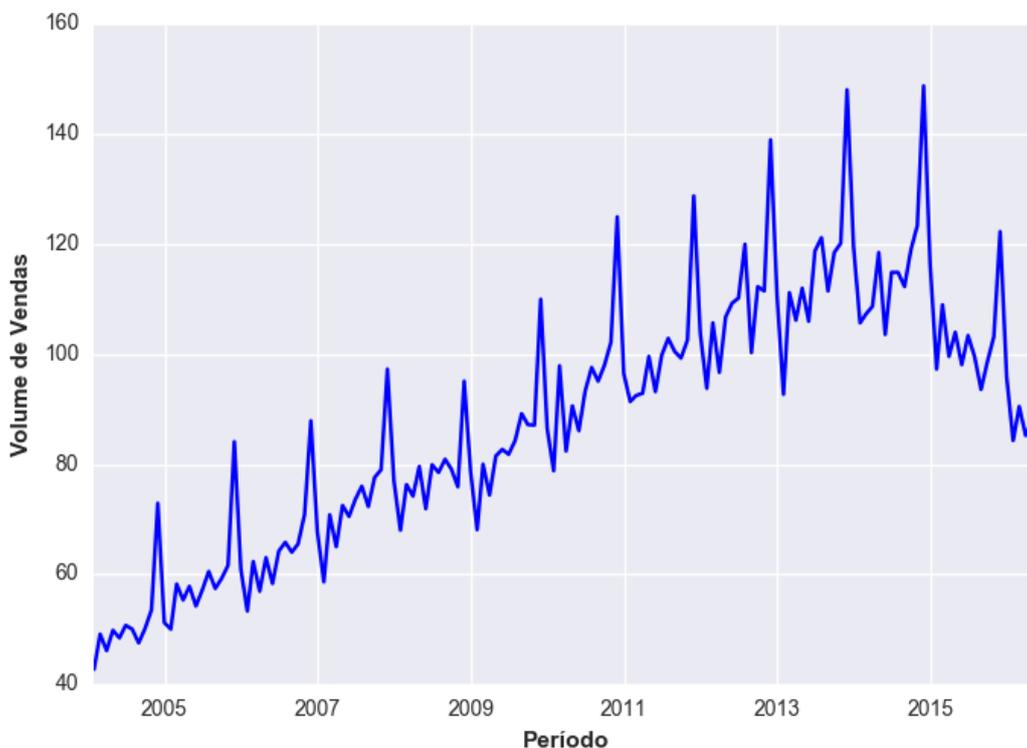


Figura 2.1: Volume de vendas do comércio varejista de Pernambuco. Obtida pelo IBGE¹

A Figura 2.1 mostra o volume mensal de vendas do mercado varejista de Pernambuco, de 2004 a 2016. Através dessa figura é possível observar que existe um comportamento de tendência de crescimento em boa parte da série, onde no decorrer dos anos o volume de vendas vai aumentando, e um comportamento de sazonalidade que formam picos em frequências semelhantes.

A Figura 2.2 mostra um correlograma, frequentemente utilizado para identificar características da série, dentre elas, quais retardos temporais (*Lags*) têm relação com o ponto futuro da

¹<http://seriesestatisticas.ibge.gov.br>

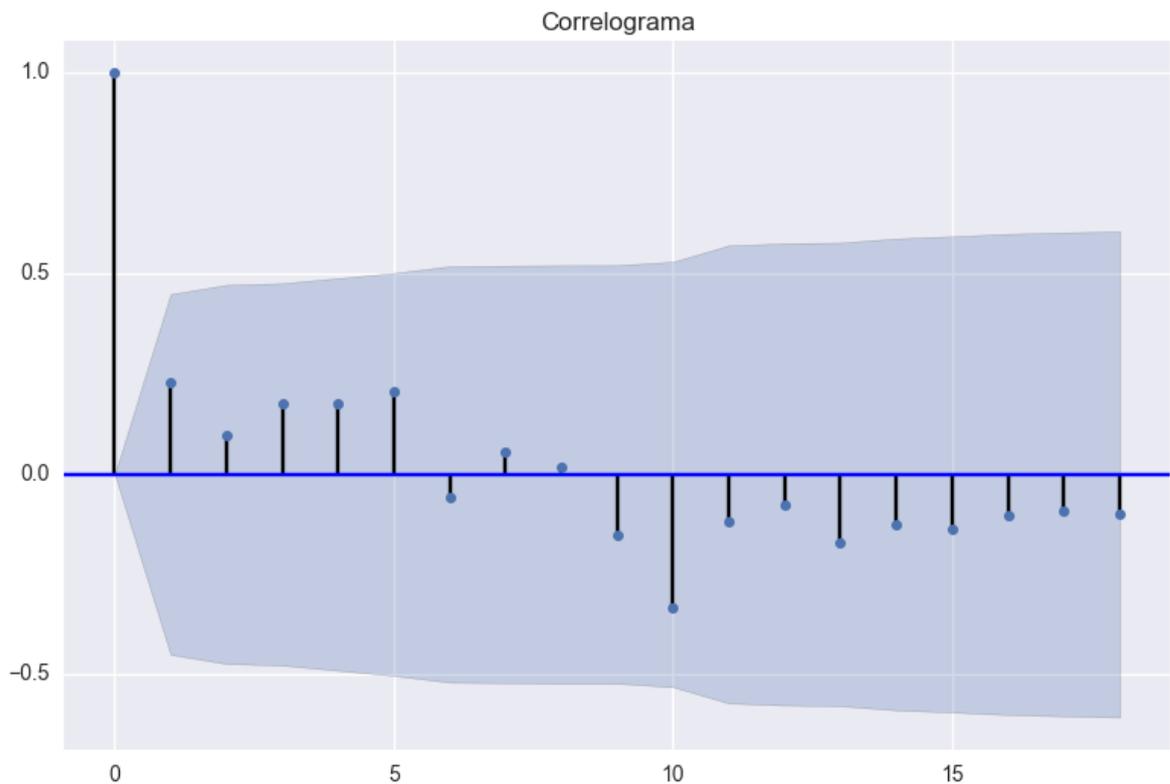


Figura 2.2: Correlograma para os 20 primeiros pontos da Série de Volume de Vendas

série. No eixo Y tem a taxa de correlação, o valor 0 representa nenhuma correlação, o valor 1 representa correlação máxima positiva e o valor -1 representa correlação máxima negativa. E no eixo X os respectivos *Lags*. A região sombreada representa os intervalos de confiança, os *Lags* que estiverem fora dessa região são considerados significantes. Através dessa figura é possível observar que os pontos anteriores $\{z_{t-1}, \dots, z_{t-20}\}$ não apresentam uma taxa de correlação significativa ao ponto z_t . Em muitas séries, o comportamento de correlação é difícil de encontrar e exige técnicas mais precisas que o correlograma (OLIVEIRA; LUDERMIR, 2014).

Séries Temporais podem ser caracterizadas como estacionárias, quando a série tem uma média e desvio padrão constantes no tempo, refletindo assim um comportamento estável (ZHANG, 2003), ou não estacionária, no caso em que diferentes médias podem ser encontradas ao segmentar a série temporal (KARTHIKEYAN; KUMAR, 2013). Em caso de séries que são não estacionárias, é necessário que seja utilizado algum filtro que as transformem em estacionárias. O filtro mais comum é a diferenciação e consiste em diferenciar os pontos da série de forma iterativa até a série se tornar estacionária. Dessa forma, a primeira diferenciação da série $Z = \{z_1, z_2, \dots, z_t\}$ pode ser representada como:

$$\Delta Z_t = Z_t - Z_{t-1} \quad (2.2)$$

Ao utilizar a série diferenciada para criar um modelo de previsão, a previsão obtida não representa o valor original da série, dessa forma é necessário que seja realizado o procedimento inverso,

também conhecido por integração da série e consiste na soma do valor previsto com o valor anterior da série:

$$\hat{z}_t = y_t + z_{t-1}, \quad (2.3)$$

tal que y_t é o valor previsto e z_{t-1} é o ponto original da série.

Segundo (MAHALAKSHMI; SRIDEVI; RAJARAM, 2016) diversos problemas de pesquisas podem ser aplicados as séries temporais, tais como:

- **Classificação de Séries Temporais:** Consiste em encontrar um modelo que classifique os padrões inclusos na série.
- **Previsão de Séries Temporais:** Tem como objetivo prever valores futuros da série com base nos valores históricos;
- **Associação Temporal:** Busca encontrar associações e relacionamentos entre os atributos da série;
- **Agrupamento de Séries Temporais:** Consiste em agrupar padrões da série que apresentam similaridade;

2.2 Previsão de Séries Temporais

A presente pesquisa teve o foco para o problema de previsão de séries temporais. Para tal, o processo pode ser representado como uma função de seus valores passados:

$$\hat{Z}_{t+1} = f(z_t, z_{t-1}, \dots, z_{t-r+1}), \quad (2.4)$$

em que \hat{Z}_{t+1} representa a previsão, o z_t é o ponto atual da série, r é o tamanho da dimensão de entrada.

Dessa forma o problema de previsão consiste em encontrar um modelo capaz de representar o comportamento gerador da série e através dele realizar as previsões. Para tal, são utilizadas técnicas de análises e métodos para previsões. As diferentes composições dessas técnicas e métodos formam as arquiteturas ou sistemas de previsão e apresentam três etapas fundamentais. A etapa de tratamento dos dados ou pré-processamento, a etapa de modelagem e a etapa de previsão.

2.2.1 Tratamento dos Dados

A etapa de tratamento exige um conhecimento a respeito dos dados e de técnicas que possam realizar o pré-processamento dos dados com o objetivo de extrair informações que contribuam na escolha dos modelos a serem utilizados na próxima etapa.

Nesta etapa, algumas atividades devem ser realizadas: primeiro, é importante conhecer as características apresentadas na série. Para tal, é necessário a utilização de técnicas para saber se a série é estacionária, se apresenta tendência e sazonalidade. Caso a série seja não estacionária, é nessa etapa onde é utilizado um filtro para transformar a série em estacionária, o filtro mais comum é a diferenciação (Equação 2.2).

A segunda atividade é definir a janela temporal a ser utilizada. Uma janela temporal é composta por *lags* da série que apresentam algum tipo de relação com o ponto futuro. Para encontrar os *lags* relevantes é possível utilizar técnicas da estatística, por exemplo função de autocorrelação, que pode ser representada através de um correlograma, de forma que, os pontos com correlação significativa serão utilizados para compor a janela temporal. Outra abordagem é através de algoritmos de otimização, em (RIBEIRO et al., 2011) e (OLIVEIRA; LUDERMIR, 2014) são utilizados algoritmos de otimização como uma alternativa viável para encontrar os *lags* relevantes, segundo os autores através desses algoritmos é possível reduzir a influência dos ruídos e dados irrelevantes contidos na série temporal, podendo melhorar a acurácia do modelo de previsão.

Outra atividade comum nesta etapa é a transformação intervalar dos valores da série, também chamado de normalização ou padronização dos dados, a qual consiste em mudar os valores da série de forma padronizada em intervalos menores, por exemplo, entre [0, 1] utilizando a seguinte equação:

$$Z_t = \frac{Z_t - Z_{min}}{Z_{max} - Z_{min}}, \quad (2.5)$$

em que Z_{max} e Z_{min} é o maior e o menor valor da série, respectivamente, e Z_t é o ponto atual que será transformado. A normalização da série é importante para abordagens de previsão com funções que contenham o seu intervalo entre [0, 1] ou [-1, 1] (ZHANG; PATUWO; HU, 1998).

2.2.2 Modelagem de Série Temporal

A etapa de modelagem tem como objetivo criar um modelo que represente o comportamento temporal dos dados utilizando alguma técnica de previsão. É nesta etapa em que os modelos são estimados, no caso de modelos estatísticos, ou treinados, no caso de abordagens de aprendizado de máquina como as RNAs.

A escolha do melhor modelo pode ser realizada através de medidas sobre o erro da previsão:

$$\varepsilon_t = z_t - \hat{z}_t, \quad (2.6)$$

em que z_t é o valor desejado e \hat{z}_t é o valor previsto. Ao realizar a avaliação do modelo é importante verificar se existe o problema de *overfitting* (HAWKINS, 2004), o qual consiste em ter um modelo que apresenta um bom desempenho em prever os padrões em que foi modelado, e um mal desempenho em prever os novos padrões. Esse problema é muito frequente em modelos

de Aprendizado de Máquina e as técnicas mais comuns para evitar esse tipo de problema são validação cruzada e *holdout* (KOHAVI et al., 1995).

Os modelos de previsão são compostos por parâmetros, e nessa etapa de modelagem é realizado o ajuste desses parâmetros de forma que consigam representar o comportamento gerador da série temporal. Tanto para os modelos estatísticos quanto para os modelos de aprendizado de máquina, os parâmetros são frequentemente ajustados com base no erro, ou seja, na diferença entre a previsão e o valor desejado. Porém, para os modelos estatísticos a escolha dos valores para os parâmetros e comumente realizada de forma analítica, enquanto nos modelos de aprendizado de máquina, os parâmetros são ajustados através de algoritmos de aprendizado ou através de algoritmos de otimização (YOLCU; EGRIOGLU; ALADAG, 2013).

2.2.3 Previsão

Esta é a última etapa, onde o modelo já tem seus parâmetros ajustados e prontos para realizar a previsão. Porém, em muitos casos, devido às incertezas na estimação ou treinamento do modelo é necessário a utilização de mais de um modelo, ou de técnicas para melhorar a sua previsão final (NEUMAN, 2003). Dessa forma, uma atividade que pode ser realizada nessa etapa é a combinação, a qual pode acontecer através de uma composição utilizando operação matemática para juntar as previsões de modelos diferentes, como mostrado no trabalho de (OLIVEIRA; LUDERMIR, 2016), ou uma combinação através de outros modelos (WIDODO; BUDI, 2011). Outra atividade que pode ser realizada é a seleção do modelo que será responsável pela previsão final (NOVA FILHO; PESSOA, 2015).

O presente trabalho tem suas contribuições nesta etapa, apresentando as melhores formas de realizar a combinação das previsões obtidas.

2.3 Modelos para Previsão de Séries Temporais

A atividade de previsão de séries temporais exige a utilização de um ou mais modelos que consigam capturar os padrões presentes na série. Para tal, existem diversos modelos na literatura, onde cada um tem sua particularidade. De forma genérica é possível dividir esses modelos em dois tipos, os estatísticos e os de aprendizado de máquina ou inteligência computacional (NOWICKA-ZAGRAJEK; WERON, 2002).

2.3.1 Modelos Estatísticos

Os modelos estatísticos são assim chamados por utilizarem técnicas estatísticas para a sua composição. Os mais conhecidos são: Auto Regressivo (AR) (DICKEY; FULLER, 1979), o modelo de Média Móveis (MA) (BOX; PIERCE, 1970), o modelo ARIMA (BOX et al., 2015) e o GARCH (BOLLERSLEV, 1986). Dentre esses modelos, o ARIMA é o mais popular.

O modelo $ARIMA(p,d,q)$ consiste na composição dos modelos $AR(p)$ que pode ser expresso através da equação 2.7 e $MA(q)$ que pode ser expresso pela equação 2.8, junto com uma técnica de diferenciação de ordem (d) da série, onde a mais conhecida é expressa através da equação 2.2.

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \dots + A_p Z_t + \varepsilon_t. \quad (2.7)$$

$$Z_t = \varepsilon_t + B_1 \varepsilon_{t-1} + B_2 \varepsilon_{t-2} + \dots + B_q \varepsilon_t. \quad (2.8)$$

A equação 2.7 representa o modelo AR de ordem (p), onde Z_t é o ponto t -ésimo da série, o A representa os coeficientes autoregressivos de ordem (p) e ε_t representa um ruído branco com média zero e variância σ^2 . Dessa forma, o $AR(p)$ modela uma série a qual é gerada por uma soma ponderada dos últimos valores (p) junto com um valor retirado de uma série de ruído branco. A formação do modelo AR consiste em três etapas. Na primeira etapa é definido o valor da ordem do processo de previsão, para isso é muito utilizado o autocorrelograma mostrado na Figura 2.2. Na segunda etapa os parâmetros de A são estimados, para tal, é comum utilizar o método dos mínimos quadrados. E por fim, o valor do ruído é definido utilizando a equação 2.6. Após a definição dos parâmetros, a previsão é a saída da equação do modelo AR.

A equação 2.8 representa o modelo MA de ordem (q), onde ε_t é o t -ésimo termo da série de ruído, B_p são os parâmetros que caracterizam a série. A formação do modelo MA consiste em duas etapas: a primeira é semelhante ao modelo AR, onde é definido o valor de ordem (q) do modelo, para tal pode ser utilizado o autocorrelograma. Na segunda etapa, os valores dos parâmetros B são estimados através do método da máxima verossimilhança. Após encontrar os valores dos parâmetros a previsão é a saída da equação do modelo MA.

A junção dessas duas equações resulta no modelo ARMA, o qual é utilizado em séries estacionárias. Já o modelo ARIMA é simplesmente o modelo ARMA junto com um filtro representado por a letra I e que tem o papel de tornar a série estacionária. Dessa forma o modelo ARIMA pode ser representado por a equação 2.9:

$$Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \dots + A_p Z_t + \varepsilon_t - B_1 \varepsilon_{t-1} + B_2 \varepsilon_{t-2} + \dots + B_q \varepsilon_t \quad (2.9)$$

A fim de facilitar o processo de formação do modelo ARIMA, Box e Jenkins (BOX et al., 2015) elaboraram um procedimento iterativo para encontrar o modelo adequado. Esse procedimento é dividido em três etapas principais: identificação, estimação e diagnóstico, e pode ser ilustrado através da Figura 2.3.

A etapa de identificação tem o objetivo de encontrar as correlações entre as observações dentro da série temporal. Para tal, são utilizadas as técnicas de Função de Autocorrelação (FAC) e a Função de Correlação Parcial. Com base nos valores obtidos, é possível determinar os valores de ordem (p) e (q) para os modelos AR e MA, respectivamente.

A etapa de estimação tem o objetivo de estimar os parâmetros do modelo escolhido na etapa anterior. Para tal, as técnicas de estimação através de mínimos quadrados ou de máxima

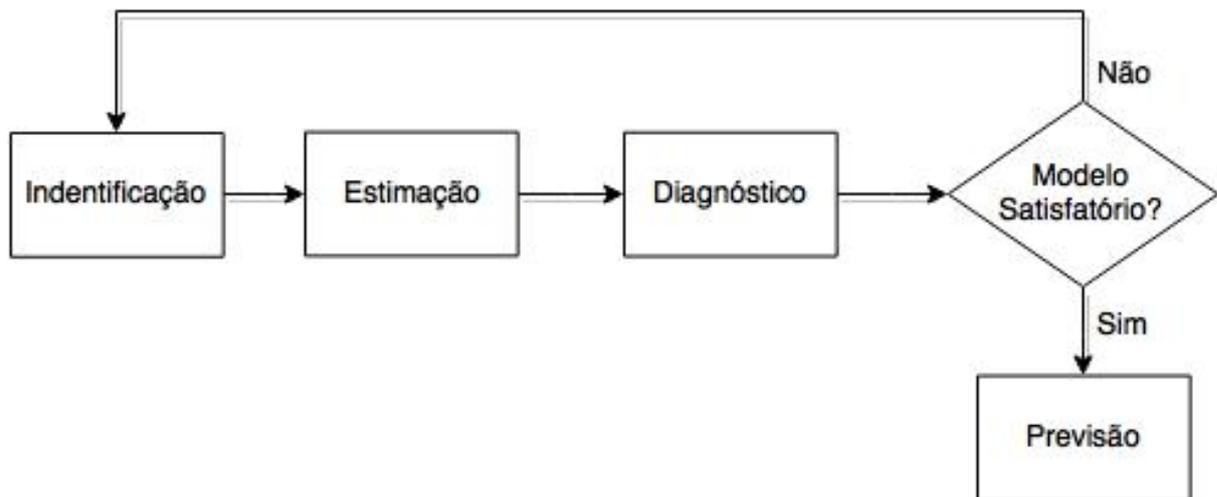


Figura 2.3: Etapas para a construção de um modelo ARIMA adaptado de (BOX et al., 2015)

verossimilhança são bastante utilizadas.

A etapa de diagnóstico tem o objetivo de analisar se o modelo resultante das etapas anteriores é estatisticamente adequado aos dados, ou seja, se representa muito bem o efeito gerador da série temporal. Para isso, é realizada uma análise no ruído gerado pelo o modelo, e caso seja um ruído branco, o modelo é o adequado para realizar a previsão.

O modelo ARIMA, mesmo sendo bastante utilizado, apresenta desempenho de previsão inferior a Rede Neural, principalmente em séries não lineares (KOHZADI et al., 1996).

2.3.2 Modelos de Aprendizado de Máquina

Os modelos de Aprendizado de Máquina são construídos a partir de dados do conjunto de amostra. Nos métodos de aprendizagem supervisionada os modelos são elaborados com base na relação entre a entrada e a saída. Estes modelos estão sendo cada vez mais utilizados para problemas de previsão de séries temporais, seja com Redes Neurais Recorrentes (GUO et al., 2016), *Extreme Learning Machine* (GUO; LIEW, 2016), Redes Fuzzy (DAS; BISOI; DASH, 2015), SVR (SAMARAWICKRAMA; HEMAPALA; JAYASEKARA, 2016) ou MLP (TAKEYASU; HIGUCHI; TSUCHIDA, 2016).

2.3.2.1 Redes Neurais Artificiais

Rede Neural Artificial é uma técnica de Aprendizado de Máquina inspirada no sistema biológico neural, como o cérebro, para processar informações (HAYKIN; NETWORK, 2004). Segundo (DONALDSON; KAMSTRA, 1996), uma RNA é um conjunto de funções de transferências que relacionam as variáveis de entrada com as variáveis de saída. Dessa forma, uma RNA é composta por um conjunto de unidades de processamentos simples e paralelos, conhecidos

por nodos ou neurônios, e tem o papel de armazenar o conhecimento através de pesos e funções matemáticas.

Esses neurônios são distribuídos de forma conjunta em uma ou mais camadas e estão conectados através de pesos, os quais servem como ponderadores sobre a informação a ser processada pelos neurônios da rede. Dessa forma, cada nodo da RNA é condicionado a receber um sinal, o qual é ponderado através dos pesos das conexões de entrada, e em seguida, esse sinal é processado por uma função matemática, conhecida como função de ativação ou função de transferência, resultando em um novo sinal de saída que é propagado pela a rede para a próxima camada.

A relação entre a saída (Z_t) e as entradas ($Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$) de uma RNA tem a seguinte expressão matemática:

$$Z_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left(\beta_{0j} + \sum_{i=1}^p \beta_{ij} z_{t-1} \right) + \varepsilon_t, \quad (2.10)$$

tal que, α_j ($j = 1, 2, \dots, q$) e β_{ij} ($i = 1, 2, \dots, p; j = 1, 2, \dots, q$) são os parâmetros do modelo, conhecidos como os pesos de conexões. O p é o número de neurônios de entrada, q é o número de neurônios da camada escondida, e g é a função de transferência ou de ativação de cada neurônio, onde a função logística é a mais utilizada e pode ser expressa através da seguinte equação:

$$g(x) = \frac{1}{1 + \exp(-x)}, \quad (2.11)$$

em que, x é o valor ponderado do peso da conexão através do valor de entrada. Essa função restringe a amplitude da saída do neurônio, tipicamente no intervalo de $[0, 1]$ ou $[-1, 1]$. Porém esses intervalos podem ser diferentes de acordo com a função utilizada.

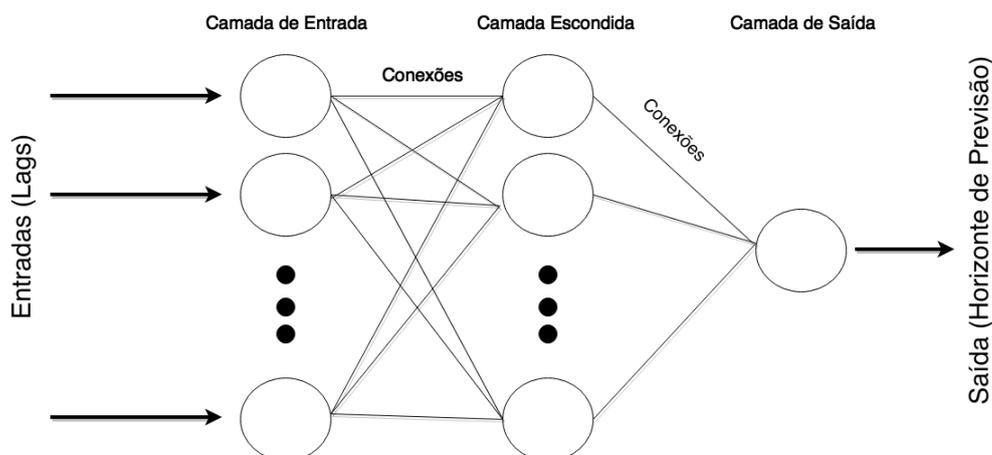


Figura 2.4: Rede Neural Artificial MLP com uma camada escondida. Adaptada de (MATTOS NETO, 2008)

A Figura 2.4 mostra de forma simplificada a arquitetura de uma RNA do tipo Perceptron Multi-Camadas. Quando todos os neurônios de uma camada estiverem ligados a todos os

neurônios das camadas adjacentes, a rede é chamada de RNA MLP completamente conectada.

A primeira camada é composta por neurônios de entrada. A quantidade de neurônios nessa camada denota a quantidade de *Lags* (em problemas de previsão de séries temporais uni-variadas). Além da camada de entrada, uma MLP pode possuir uma ou mais camadas intermediárias, onde geralmente estão a maior quantidade de neurônios da rede. Os neurônios das camadas intermediárias são responsáveis por processar os dados da camada anterior através de funções de transferência e enviar os resultados para a camada de saída. Na camada de saída, onde a quantidade de neurônios denota a quantidade de saídas necessárias do problema, é realizado um novo processamento que retorna a resposta final da rede. Dependendo da quantidade de neurônios nas camadas escondidas, dois problemas podem ocorrer: *underfitting* ou o *overfitting* (COULIBALY; ANCTIL; BOBEE, 2000). O *underfitting* acontece quando são utilizados poucos neurônios e com isso não conseguem capturar os padrões de forma adequada. O *overfitting* acontece quando são utilizados muitos neurônios e com isso a rede se ajusta demasiadamente aos dados de treinamento, perdendo a capacidade de generalização.

A modelagem de Redes Neurais acontece através da etapa de treinamento, que será responsável por ajustar os pesos da rede. Geralmente uma RNA do tipo MLP é treinada através de um processo supervisionado, ou seja, na etapa de treinamento a rede tem acesso a um conhecimento prévio (a saída desejada), e os ajustes são realizados com base na diferença da saída apresentada pela a rede e a saída desejada.

O algoritmo de treinamento para RNA do tipo MLP mais utilizado é o *Backpropagation* (RUMELHART et al., 1988) e consiste em propagar a correção do erro da última camada até à primeira de forma iterativa (HAYKIN; NETWORK, 2004). Esse erro pode ser calculado através de métricas como o MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y} - Y)^2 \quad (2.12)$$

Como dito anteriormente, vários fatores influenciam para que as RNAs sejam comumente utilizadas. Segundo (KHASHEI; BIJARI, 2010), as RNAs são auto adaptativas e orientadas aos dados, ou seja, são modeladas com base nos dados, e ao utilizar da aprendizagem supervisionada fazem o mapeamento entre a entrada e a saída. Dessa forma as RNA podem ser modeladas para problemas diferentes, apenas passando novos dados. As RNAs têm capacidade de generalização e com isso podem inferir os padrões ainda não vistos e até mesmo com ruídos, característica importante para problemas de séries temporais. Por fim, são aproximadores universais de funções, tal que, uma RNA pode com a quantidade correta de neurônios modelar a função geradora da série temporal e com isso apresentar bom desempenho nas previsões.

Dentre as limitações da utilização de uma RNA, é possível observar que são consideradas uma "caixa preta", ou seja, é difícil explicar como o modelo retornou determinada resposta, assim como é complicado explicar as ligações entre os neurônios e a relação com os padrões obtidos. Outra limitação se dá pelo fato de apresentar parâmetros sensíveis: a quantidade de camadas

escondidas, a quantidade de neurônios em cada camada escondida e o algoritmo de treinamento, são parâmetros que estão completamente relacionados com o desempenho do modelo e que podem variar de acordo com o problema. Para tal é necessário conhecimento a respeito do contexto, ou a utilização de algoritmos de otimização (WANG; ZENG; CHEN, 2015).

2.3.2.2 Máquina de Vetor de Suporte para Regressão

Máquinas de vetor de suporte para regressão (SVR) é um método proposto por (VAPNIK; VAPNIK, 1998) baseado na Máquina de Vetor de Suporte para Classificação (SVM) (VAPNIK, 1963) que consiste na construção de hiperplanos compostos por padrões de treinamento e utilizados para classificar os novos padrões. O SVR é um modelo não linear para resolver problemas de regressão e utilizado para previsão de séries temporais.

O processo de aplicação de SVR para previsões de séries temporais consiste em encontrar uma função $f(x)$ com liberdade de erro menor ou igual a ε . A etapa de treinamento é semelhante a resolução de problemas de otimização de funções convexas ou programação quadrática, a qual fornece sempre um único valor ótimo e não existem mínimos locais (ALWEE; HJ SHAMSUDDIN; SALLEHUDDIN, 2013). O modelo SVR pode ser representado pela seguinte expressão:

$$f(x) = (\mathbf{W} \cdot \phi(x)) + b, \quad (2.13)$$

onde \mathbf{W} é um vetor de pesos, "b" é o limiar e ambos são a solução do problema de otimização, $\phi(x)$ é uma função kernel. Dessa forma o algoritmo de treinamento tem o papel de encontrar o vetor de pesos \mathbf{W} e um limiar "b" em um conjunto de dados com l amostras, que minimize a função:

$$\frac{1}{2} \|\mathbf{W}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \quad (2.14)$$

em que obedeça as seguintes restrições:

$$y_i - \langle \mathbf{w}, x_i \rangle - b \leq \varepsilon + \xi_i \quad (2.15)$$

$$\langle \mathbf{w}, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \quad (2.16)$$

$$\xi_i, \xi_i^* \geq 0, \quad (2.17)$$

em que ξ_i^* e ξ_i são as variáveis de folga responsáveis por medir o custo dos erros de previsão e foram criadas caso no processo de treinamento não encontre uma função que limite os erros em ε . y_i é o valor desejado, $C > 0$ é o fator de regularização, responsável por regularizar a complexidade da função e a quantidade de desvio maiores que ε que serão tolerados (SMOLA; SCHÖLKOPF, 2004). Dessa forma, o valor de C está relacionado com a complexidade da função

(quanto maior o valor de C , maior é a complexidade) e com a suavidade da função (quanto menor o valor de C , mais suave será a função). O ε representa a região de custo, que é contabilizada apenas se o valor predito cair nessa região.

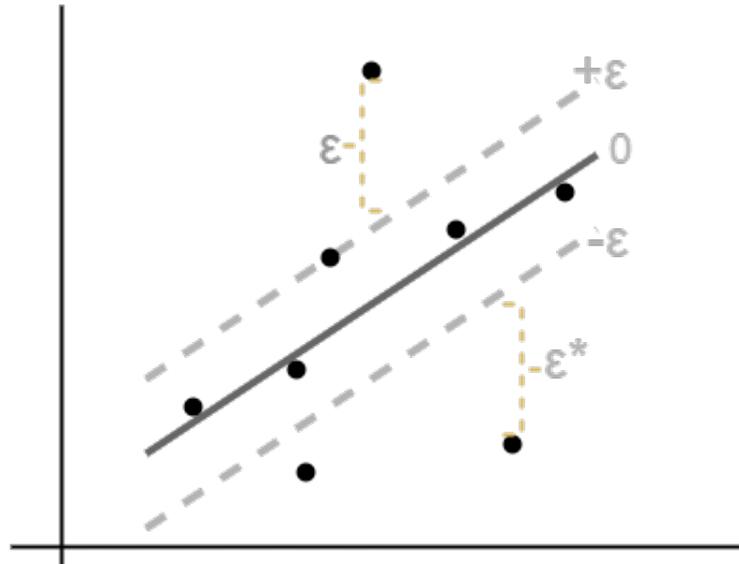


Figura 2.5: Representação de uma SVR com margens suaves. Adaptado de (SMOLA; SCHÖLKOPF, 2004)

Através da Figura 2.5 é possível visualizar a representação de uma SVR com margens suaves. Como mostra na figura, existe um hiperplano linear que separa os dados, esse hiperplano é construído com base no teorema de Cover, o qual relata que um padrão não-linear separável pode ser linearmente separável em uma dimensão maior, caso a transformação para o novo espaço seja não-linear e que a dimensionalidade seja suficientemente grande (HAYKIN; NETWORK, 2004). Dessa forma, a transformação do espaço de entrada para outro espaço com uma dimensionalidade maior, é realizado através de funções *Kernel* (BERGMAN, 1970). Dentre as diversas funções *Kernel*, disponíveis na literatura, a *Radial Basis Function* (RBF) é a mais utilizada em problemas de previsão de séries temporais (WU; CHENG, 2010), principalmente por ter um treinamento mais rápido e ser bastante flexível (XIN et al., 2012). A função RBF tem a seguinte expressão:

$$\exp\left(-\frac{1}{2\sigma^2}|x-x_i|^2\right), \quad (2.18)$$

em que σ é a largura dos núcleos da função de base radial. A escolha da função de *Kernel* é importante para o desempenho do modelo SVR e influencia nos parâmetros do modelo.

Entre as limitações de modelos SVR, as principais são: alta sensibilidade aos parâmetros que são escolhidos pelos usuários (C , ε , e o σ caso utilize o RBF), o alto nível de complexidade do algoritmo. Dependendo da função *Kernel* escolhida e da quantidade de dados, o treinamento pode ser mais lento que de RNAs do tipo MLP e exigir mais memória da máquina (BURGES, 1998). Além desses, caso o modelo SVR seja mal ajustado pode acarretar em problemas de *underfitting* ou *overfitting* semelhantes as RNAs.

2.4 Sistemas de Múltiplos Preditores

Sistemas de Múltiplos Preditores (SMP) podem ser definidos como um conjunto de modelos diferentes $(f_1(z), f_2(z), \dots, f_n(z))$ com o objetivo de prever a mesma variável z_t , onde n é o tamanho do conjunto. A previsão final de um SMP é realizada com base na combinação de todas as previsões obtidas ou através da seleção de uma previsão.

Segundo (ROLI; GIACINTO; VERNAZZA, 2001) o processo de construção de um sistema de múltiplos preditores pode ser dividido em duas etapas:

1. **Geração do *Ensemble*:** Essa etapa consiste em criar o conjunto de modelos, também conhecido como o *Pool* de preditores. No entanto, é necessário que exista diversidade entre os modelos. Existem diferentes formas de gerar um *Ensemble* com modelos diversos. A mais comum é através dos dados de treinamento, onde cada modelo é treinado utilizando uma amostra de dados diferentes. Para tal, existem diferentes técnicas: *Bagging* (MORETTI et al., 2015), o *AdaBoost* (LIU et al., 2015), e particionando a série em diferentes amostras (NOVA FILHO; PESSOA, 2015). Outra forma de tornar os modelos diferentes é através de suas configurações, ou seja, mesmo um *Ensemble* sendo composto de RNAs treinadas com o mesmo conjunto de dados, elas podem apresentar previsões diferentes, caso tenham a quantidade de neurônios, o número de camadas escondidas, ou o algoritmo de treinamento, diferentes. Outra forma bastante comum é criar um *Ensemble* com modelos de diferentes características, como exemplo, o trabalho de (ADHIKARI, 2015) onde é utilizado um *Ensemble* composto por uma SVR, um ARIMA, uma MLP e uma RNA Recorrente.
2. **Integração:** Essa etapa consiste em integrar as previsões dos modelos em apenas uma previsão final. Para tal, existem dois grupos de técnicas de integração: a combinação e a seleção. Na combinação todas as previsões são utilizadas para compor a previsão final do *Ensemble*. Já na seleção, apenas um modelo é escolhido para ser o responsável pela previsão.

2.4.1 Combinação de Preditores

A combinação de preditores de diferentes modelos teve início no trabalho de (BATES; GRANGER, 1969). Desde então vem sendo objeto de estudo e apresentando bons resultados (MENDES-MOREIRA et al., 2012). Como dito anteriormente, a combinação de preditores consiste em utilizar alguma técnica para combinar todas as previsões, resultando em apenas uma previsão.

Dentre as diversas técnicas de combinação, as mais utilizadas são:

- **Combinação Linear Ponderada:** consiste em combinar as previsões de forma linear através de pesos pré-estabelecidos. De forma genérica, é possível representar uma

combinação linear através da seguinte expressão:

$$y_f = \sum_{i=1}^n w(i) \cdot \hat{y}_i, \quad (2.19)$$

em que, y_f é a resposta final, y_i é a previsão do modelo i , e $w(i)$ é o peso atribuído à previsão y_i . A escolha do valor de $w(i)$ pode ser realizada de diversas formas, sendo mais comum através do erro de treinamento:

$$w_i = \frac{e_i}{\sum_{i=1}^n e_i}, \quad (2.20)$$

tal que, cada previsão é ponderada com base no erro e_i do modelo em relação aos outros modelos. O ponto negativo dessa técnica está na dificuldade de encontrar os pesos que representem corretamente a relação entre os modelos.

- **Combinação Linear por Regressão:** consiste em modelar a relação entre as previsões obtidas e o valor desejado. Para tal, são estimados pesos e ajustados de forma que minimize a soma dos erros ao quadrado. De forma genérica uma função de regressão pode ter a seguinte representação:

$$y_f = w_0 + \sum_{i=1}^n w_i \cdot \hat{y}_i, \quad (2.21)$$

tal que, w_0 é uma constante e w_i é o peso associado a previsão y_i .

Diversos tipos de funções de regressão linear podem ser aplicadas em combinação de preditores. O trabalho de (GRANGER; RAMANATHAN, 1984) é um dos pioneiros e utiliza a regressão linear múltipla para realizar a combinação entre as previsões dos modelos. No trabalho de (AKSU; GUNTER, 1992) é apresentado diferentes técnicas de regressão, a *Equality Restricted Least Squares* (ERLS) em que os pesos são restritos a soma da unidade e a *Nonnegativity Restricted Least Squares* (NRLS) que obriga os pesos a não serem negativos. No trabalho de (NÓBREGA; OLIVEIRA, 2013) diferentes abordagens de regressão são utilizadas: *Bayesian Model Averaging* que utiliza *Akaike Information Criterion* (AIC) para encontrar os pesos ótimos, o *Least Absolute Shrinkage and Selector Operator* (LASSO) que utiliza limites para a soma dos valores absolutos dos pesos, e *Kalman Filter Regression Model* que utiliza do Filtro de Kalman (HARVEY, 1990) para estimar os pesos entre as previsões de cada modelo.

- **Combinação Linear por Perceptron:** utiliza-se do modelo mais simples de uma rede neural, o *perceptron*, para modelar a relação linear entre as previsões dos modelos. Para tal, é realizado o treinamento do *perceptron* de forma iterativa, passando como entrada as previsões dos modelos e como saída desejada o valor real

a ser previsto. Um modelo do tipo *perceptron* tem a seguinte representação:

$$y_f = f\left(b + \sum_{i=1}^n w_i \cdot \hat{y}_i\right), \quad (2.22)$$

em que, w_i são os pesos que representam a relação entre as previsões y_i e são ajustados através de um processo de minimização do erro utilizando gradiente e descendente, a combinação é obtida através da função de saída $f(\cdot)$.

- **Combinação por Média Simples:** A média simples é uma das técnicas mais utilizadas, e mostra uma tendência central em relação ao conjunto de previsões. Essa média pode ser utilizada com ou sem um conjunto de pesos. A combinação por Média tem a seguinte expressão:

$$y_f = \frac{1}{n} \sum_{i=1}^n w(i) \cdot y_i. \quad (2.23)$$

Caso não sejam utilizados pesos para as previsões o valor de $w(i)$ será igual a 1. Os principais pontos negativos dessa técnica, é o fato de ser sensível a valores aberrantes e distribuições distorcidas, ou seja, é uma técnica indicada para um conjunto de previsões que tenham uma distribuição normal (JOSE; WINKLER, 2008).

- **Combinação por Mediana:** A mediana é considerada uma medida mais robusta que a média (KOURENTZES; BARROW; CRONE, 2014) e tem seu funcionamento da seguinte forma: Caso o tamanho do *Ensemble* seja um número ímpar, os valores das previsões são ordenados e o valor médio é retornado para ser a previsão final, caso contrário, será retornado como previsão final a média dos dois valores centrais do *Ensemble*.
- **Combinação não Linear:** A combinação não linear utiliza de funções não lineares para realizar a combinação. De forma genérica, uma técnica não linear pode ser expressa da seguinte forma:

$$y_f = g(x), \quad (2.24)$$

em que $g(x)$ representa uma função não linear que recebe como entrada o conjunto de previsões $x = \{y_1, y_2, \dots, y_n\}$. Através dessa representação é possível analisar que o problema de combinação de previsões é semelhante a um problema de regressão, onde o conjunto de previsões são as entradas e a previsão combinada é a saída. Dessa forma, vários modelos de aprendizado de máquina podem ser utilizados para realizar a combinação, como mostra o trabalho de (SHI; DA XU; LIU, 1999), que utiliza uma RNA do tipo MLP, o trabalho de (HE; XU, 2005) utiliza uma RNA do tipo *Self-Organized-Maps* (SOM), e o trabalho de (FIORDALISO, 1998) que utiliza Redes *Fuzzy*.

Segundo (LEMKE, 2010), a combinação de preditores através de abordagens não

lineares necessita de novos estudos, e desde então, poucos trabalhos com abordagens de combinação não lineares podem ser encontrados na literatura.

A utilização de técnicas lineares para combinação tem como motivação a simplicidade de implementar e compreender, porém, geralmente ignoram a relação entre os modelos (ADHIKARI; AGRAWAL, 2012), além de que, a relação entre as previsões dos modelos e a previsão desejada nem sempre é uma relação linear (SHI; DA XU; LIU, 1999). Enquanto que os modelos não lineares, apesar de serem mais complexos, podem ser capazes de capturar tanto os padrões lineares como não lineares. Dessa forma, se faz necessário que sejam realizados estudos comparativos entre as abordagens de combinação linear e não linear a fim de verificar em quais situações cada abordagem apresenta o melhor desempenho.

2.4.2 Seleção de Preditores

A seleção de preditores tem como objetivo encontrar um modelo do *Ensemble* que apresente resultados melhores que o conjunto completo.

Segundo (YU et al., 2005), existem três formas genéricas para seleção de preditores. A primeira é através de análise gráfica e estatística (como o auto correlograma e autocorrelações parciais). A segunda é através de testes de hipóteses. A terceira forma é através de um método de seleção formal e bem definido como o AIC (AKAIKE, 1998) que utiliza uma medida de discrepância entre dois modelos, ou o BIC (SCHWARZ et al., 1978) que considera aproximação entre os modelos através da probabilidade Bayesiana.

No entanto, uma forma mais simples de separar as técnicas de seleção é através de abordagens estáticas e dinâmicas.

As abordagens estáticas consistem na seleção dos modelos durante o processo de treinamento, através do desempenho dos modelos no conjunto de treinamento ou de validação. Essa abordagem pode ser considerada uma técnica de corte, pois está reduzindo o conjunto de preditores a um único modelo.

No trabalho de (ADHIKARI; VERMA; KHANDELWAL, 2015) é utilizado uma arquitetura de seleção estática, onde é criado um *ranking* através do desempenho dos modelos no conjunto de treinamento e o modelo que apresentar melhor desempenho será utilizado. Segundo o autor, essa arquitetura apresenta melhores resultados que a combinação utilizando a média.

As abordagens dinâmicas consistem em realizar a seleção durante a utilização do *Ensemble*, ou seja, para cada novo conjunto de entrada um modelo diferente poderá ser selecionado para realizar a previsão. No trabalho de (ROSA; BARBOSA; RIBEIRO, 2016) é realizado uma seleção do preditor baseado em contexto para previsão de localização móvel. O trabalho de (NOVA FILHO; PESSOA, 2015), também apresenta uma arquitetura de seleção dinâmica, onde o preditor é escolhido com base na similaridade do seu conjunto de treinamento com os novos padrões de teste.

Segundo (TSOUMAKAS; PARTALAS; VLAHAVAS, 2008) a utilização da seleção de preditores é importante por duas razões: eficiência e performance, pois a utilização de um conjunto de modelos para serem combinados resultam em um maior custo computacional se comparado a selecionar apenas um modelo. Além disso, uma grande quantidade de modelos pode acarretar na existência de modelos com diversas variações de desempenhos e ao combiná-los, a performance do *Ensemble* poderá ser reduzida.

2.5 Considerações Finais

Neste segundo capítulo foram definidos alguns conceitos e características a respeito de séries temporais, por exemplo: tendência, sazonalidade, séries estacionárias e retardos temporais relevantes (*Lags*). Também foi apresentado o problema de previsão de séries temporais e as três principais etapas para a construção de uma arquitetura de previsão: tratamento dos dados, modelagem de série temporal e previsão. Em cada etapa foram apresentadas técnicas importantes para o desenvolvimento da arquitetura.

Também foram abordados diferentes tipos de modelos para a previsão de séries temporais. O modelo estatístico ARIMA, e os modelos de aprendizado de máquina: RNA e SVR, os quais se destacam em relação aos estatísticos por apresentarem na literatura bons resultados, principalmente em séries com padrões não lineares.

E por fim, o conceito de sistemas de múltiplos preditores foi apresentado junto com as diferentes abordagens de integração, seja por combinação ou seleção. Ambas abordagens apresentam suas vantagens, a combinação por apresentar resultados com menos variância e levar em consideração diferentes perspectivas do problema de previsão, através de diferentes especialistas. E a seleção que pode apresentar um menor custo computacional e previsões mais precisas que a combinação.

3

MÉTODO PROPOSTO

Neste capítulo é apresentado o método proposto e está dividido nas seguintes seções: na seção 3.1 é descrito as etapas da arquitetura de seleção utilizada, na Seção 3.2 é apresentada a arquitetura de combinação proposta. Por fim as considerações finais do capítulo.

3.1 Arquitetura de Partição e Seleção de Preditores

A arquitetura desenvolvida por (NOVA FILHO; PESSOA, 2015) foi utilizada como objeto de estudo na presente dissertação, por apresentar uma abordagem de treinamento dos modelos através de diferentes partições, criando modelos especialistas em determinados comportamentos da série e por apresentar desempenhos melhores que abordagens que utilizam de apenas um modelo. Com isso, essa arquitetura será utilizada como base de comparação para as abordagens de combinação.

A arquitetura pode ser dividida em duas fases: Treinamento e Teste.

3.1.1 Treinamento

A fase de treinamento dos modelos foi dividida em quatro etapas: particionamento, seleção de *lags*, seleção dos dados de validação e treinamento dos preditores. As diversas etapas da fase de treinamento da arquitetura podem ser vistas na Figura 3.1.

3.1.1.1 Particionamento

Nessa etapa a série temporal é dividida em diversas partições $J = \{j_1, j_2, \dots, j_q\}$, podendo existir interseção entre as janelas adjacentes, sendo q o número de partições. O objetivo dessa etapa é gerar amostras menores da série, de forma que cada partição consiga representar um intervalo específico da série, e espera-se que em cada intervalo esteja contido comportamentos diferentes da série. Dessa forma, ao treinar um modelo para cada partição, o resultado será um conjunto de especialistas em diferentes comportamentos da série.

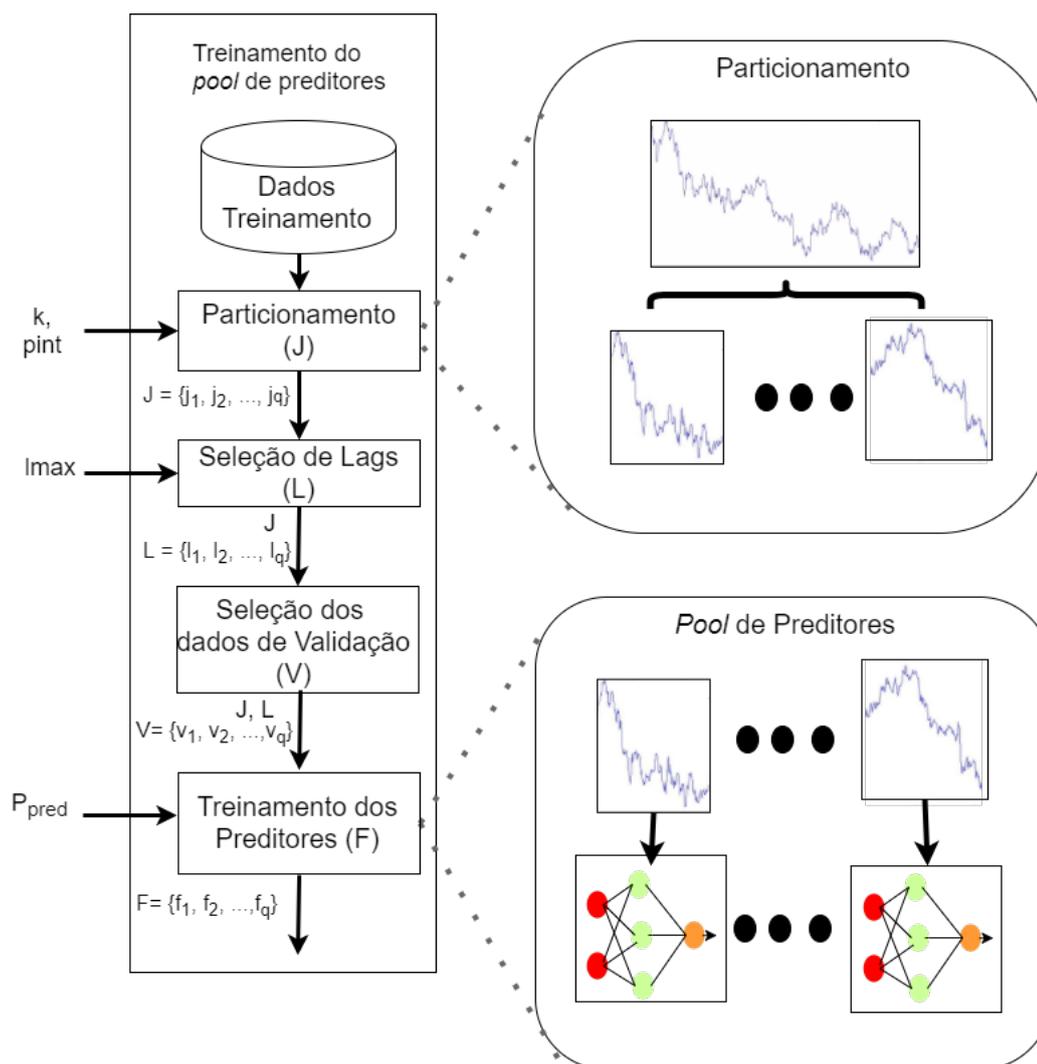


Figura 3.1: Etapas de treinamento da arquitetura de seleção. Figura adaptada de (NOVA FILHO; PESSOA, 2015)

O processo de particionamento da série depende de dois parâmetros principais: o tamanho da janela (k) e o percentual de interseção entre as janelas adjacentes ($pint$). Através desses parâmetros e do tamanho da série (n) é possível encontrar a quantidade de partições (q) utilizando a seguinte equação:

$$q = \left\lceil \frac{n - (k \cdot pint)}{k \cdot (1 - pint)} \right\rceil \quad (3.1)$$

Definir o tamanho das partições corretamente é de grande importância para o desempenho da arquitetura. Caso seja definido janelas com tamanhos pequenos, por exemplo uma janela com 10% do tamanho da série, pode ocorrer de o modelo não ser bem treinado, por não possuir dados suficientes. Caso seja definido janelas com tamanhos grandes, por exemplo uma janela com mais de 70% da série de treinamento, poucos preditores serão treinados, e com isso, existe uma grande chance dos modelos não se tornarem especialistas nos diferentes padrões contidos na série.

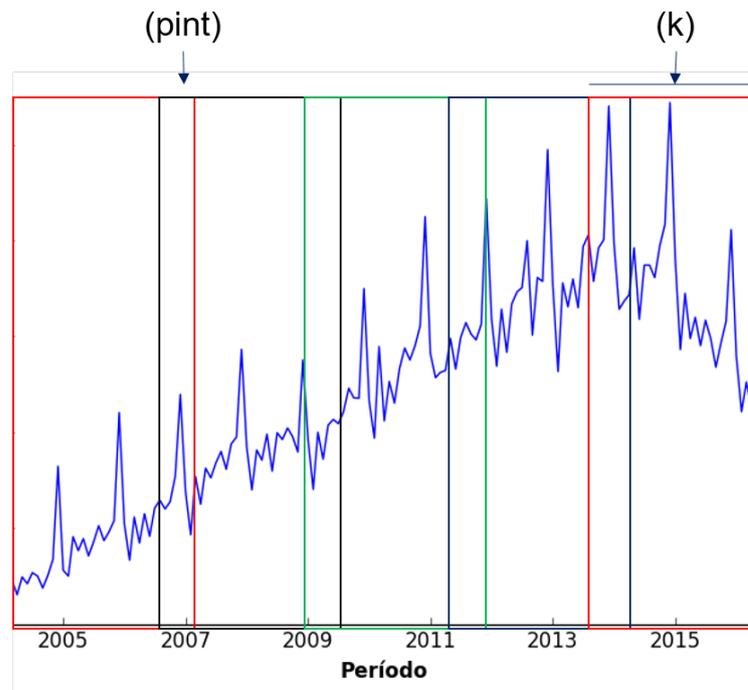


Figura 3.2: Exemplo de particionamento da Série

3.1.1.2 Seleção de Lags

Essa etapa tem como objetivo selecionar os *lags* relevantes de cada partição. O processo de seleção de *lags* consiste em analisar através da função de autocorrelação. Para tal, é definido um número máximo de *lags* ($lmax$) a serem analisados. Dessa forma, os *lags* entre 1 e ($lmax$) que apresentarem autocorrelação significativa serão utilizados para compor as entradas dos modelos de previsão.

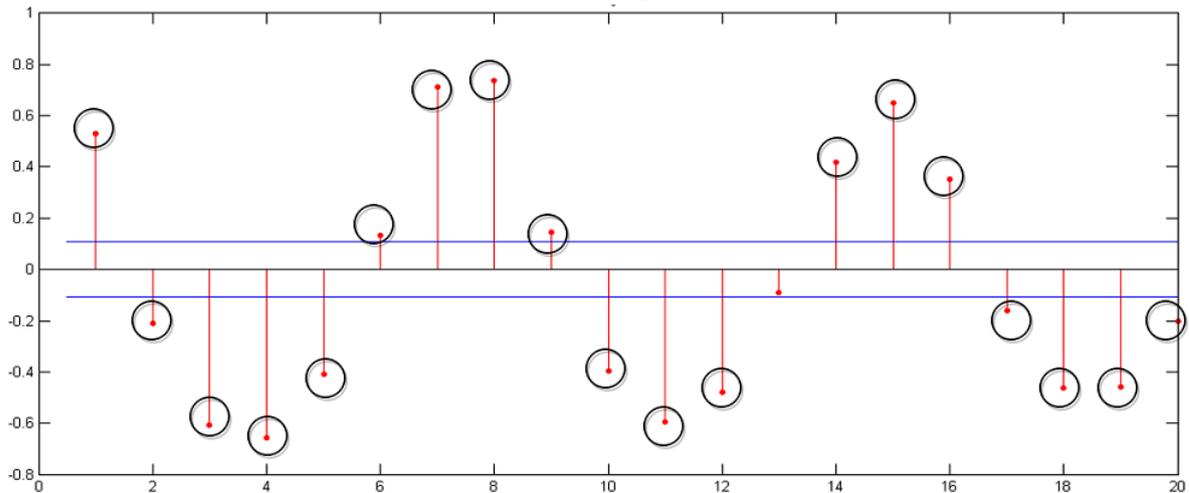


Figura 3.3: Exemplo dos lags selecionados através da função de auto correlação

3.1.1.3 Seleção dos Dados de Validação

Essa etapa tem como objetivo selecionar os dados de validação para cada partição. Como o objetivo da arquitetura é criar modelos especialistas, logo não é necessário que os modelos tenham alta capacidade de generalização para outro tipo de comportamento. Portanto nessa etapa de validação é realizada uma busca na série por padrões com comportamento semelhante a partição de treinamento de cada modelo.

Dessa forma, o conjunto de validação é composto pelos dados da partição mais semelhante. O grau de semelhança é calculado através do algoritmo de distância entre duas séries temporais, o *Dynamic Time Warping* (DTW) (SAKOE; CHIBA, 1978). A busca é realizada calculando a distância da partição avaliada em relação as outras partições que não a intersectam. Caso não existam as partições que não a intersectam, serão selecionados os dados das partições diferentes da partição avaliada, utilizando os dados que não estejam no intervalo de interseção.

3.1.1.4 Treinamento dos preditores

Nessa etapa é realizado o treinamento de cada modelo de previsão. Para tal, será utilizado uma partição para o treinamento e outra partição para a validação. Dessa forma, o número de modelos treinados será respectivo ao número de partições criadas (q). Ao término dessa etapa, pretende-se que existam modelos especialistas no comportamento específico da partição, a qual foi utilizada para o seu treinamento.

3.1.2 Teste

A fase de teste da arquitetura consiste em simular a utilização da arquitetura para novos dados e se divide em duas atividades: seleção do modelo e realização da previsão, como mostra na Figura 3.4

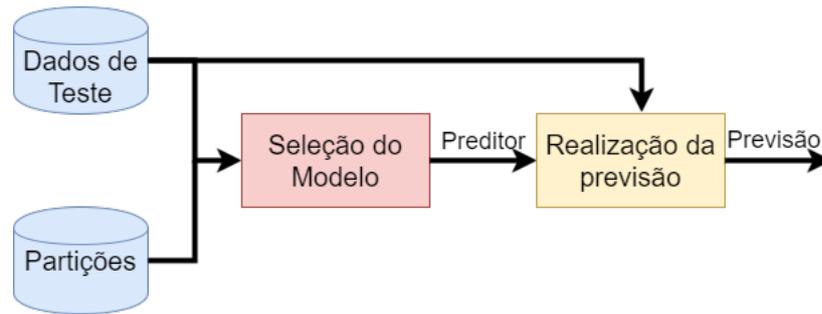


Figura 3.4: Atividades de Teste da arquitetura de seleção

3.1.2.1 Seleção do Modelo

Essa atividade tem como entrada a janela que antecede o valor a ser previsto, junto com as partições utilizadas para treinar e validar os modelos. O processo consiste em através da medida de distância DTW, encontrar a partição que seja mais semelhante a janela de teste. O resultado dessa etapa é um conjunto ordenado de distâncias correspondente as partições. Em seguida, o modelo de previsão correspondente a partição com menor distância será selecionada para fazer a previsão. Como a seleção do modelo é realizada de forma dinâmica, assim que mudar a janela que antecede o valor a ser previsto é provável que o modelo escolhido também seja outro. Dessa forma, deseja-se que para cada comportamento diferente no conjunto de teste, o modelo que foi treinado com o comportamento mais semelhante seja utilizado para realizar a previsão.

3.1.2.2 Realização da Previsão

Essa atividade tem como entrada o modelo escolhido para realizar a previsão e a série que antecede o valor a ser previsto. Nesse momento os *lags* relevantes serão utilizados como entrada para o modelo de previsão, resultando na previsão final.

3.1.3 Aprimoramento da Arquitetura

No trabalho de (NOVA FILHO; PESSOA, 2015) a arquitetura apresenta resultados melhores que a utilização dos modelos SVR e MLP de forma isolada. No entanto, ainda existem alguns pontos na arquitetura que podem ser aprimorados para encontrar resultados melhores.

Primeiramente, o desempenho da arquitetura é sensível aos parâmetros que define a quantidade de partições. Essa quantidade pode variar de acordo com a série utilizada e o modelo de previsão. Caso sejam utilizados modelos que necessitem de uma grande quantidade de dados para serem treinados, serão necessárias partições maiores, e com isso, séries temporais que contém poucos pontos poderão resultar em modelos mal treinados. Por outro lado, caso seja uma série com muitos pontos temporais e modelos que necessitem de uma partição de treinamento pequena, terá como resultado uma grande quantidade de modelos para serem selecionados e com

isso o custo computacional do modelo será elevado. Uma alternativa é escolher os parâmetros através de algum algoritmo de otimização.

Outro ponto é em relação a técnica utilizada para obter a previsão final, a seleção. Como visto em capítulos anteriores, utilizar a seleção de preditores mesmo para modelos especialistas pode deixar de fora a previsão de um outro modelo que pode contribuir para melhorar o desempenho final da arquitetura. Além disso, a seleção geralmente apresenta uma alta variância em seus resultados (ZOU; YANG, 2004), podendo assim aumentar o risco na tomada de decisão a respeito da previsão obtida (HIBON; EVGENIOU, 2005). Uma alternativa é juntar a previsão de todos os especialistas, ou seja, utilizar uma abordagem de combinação de preditores, que possibilite integrar as diferentes previsões em uma previsão final mais acurada e com menor variabilidade. Para tal, a arquitetura proposta nesse trabalho realiza esse aprimoramento, utilizando diferentes abordagens de combinação e analisando os seus comportamentos.

3.2 Arquitetura Proposta

A arquitetura proposta consiste no desenvolvimento de um sistema com múltiplos preditores utilizando abordagens lineares e não lineares para realizar a combinação da previsão e pode dividida em duas fases: treinamento e teste.

Antes da etapa de treinamento, a série temporal a ser utilizada passa por uma etapa de pré-processamento. Nesta etapa é realizado a normalização dos dados utilizando a Equação 2.5, que normaliza os dados no intervalo $[0, 1]$. Essa normalização é importante para os dados ficarem dentro do intervalo da função de ativação utilizada. Após a normalização é realizado uma verificação de estacionariedade, ou seja, é analisado se a série segue uma média e desvio padrão constante, para tal é utilizado um teste estatístico que verifica se diferentes intervalos da série contém médias iguais. Caso a série seja não estacionária, a Equação 2.2 será utilizada de forma iterativa até a série se tornar estacionária.

Após realizar o pré-processamento na série, é realizado a divisão da série em dois conjuntos: treinamento e teste. Uma parte do conjunto de treinamento é utilizado para realizar a validação dos parâmetros escolhidos para os modelos de previsão.

3.2.1 Fase de Treinamento

A fase de treinamento tem o objetivo de tornar os modelos prontos para serem utilizados. Para tal será necessário realizar o treinamento dos modelos de previsão, que segue as mesmas etapas da arquitetura de seleção e o treinamento dos modelos de combinação.

3.2.1.1 *Treinamento dos Modelos de Previsão*

A fase de treinamento tem como entrada o conjunto de dados de treinamento que passou pelo pré-processamento. Nessa fase acontece o treinamento dos modelos que irão compor o *Ensemble* de Preditores. A Figura 3.5 mostra o processo para o treinamento dos modelos de previsão, composto por duas etapas: Partição e Treinamento.

Na etapa de partição o conjunto de treinamento é dividido em q partições, a quantidade de partições será resultado da equação 3.1, e para tal, é necessário definir os parâmetros: tamanho das partições k e taxa de interseção ($pint$) entre as partições. O término dessa etapa resultará em q partições, que serão utilizadas para o treinamento dos modelos.

Ainda nessa etapa é realizado a seleção dos *lags* a serem utilizados em cada partição. Para tal, é utilizado a função de autocorrelação, e observado quais *lags* possuem autocorrelação significativa. É definido um número máximo de *lags* ($lmax$) a serem analisados. Dessa forma, os *lags* entre 1 e ($lmax$) que apresentarem autocorrelação significativa serão utilizados para compor as entradas dos modelos de previsão.

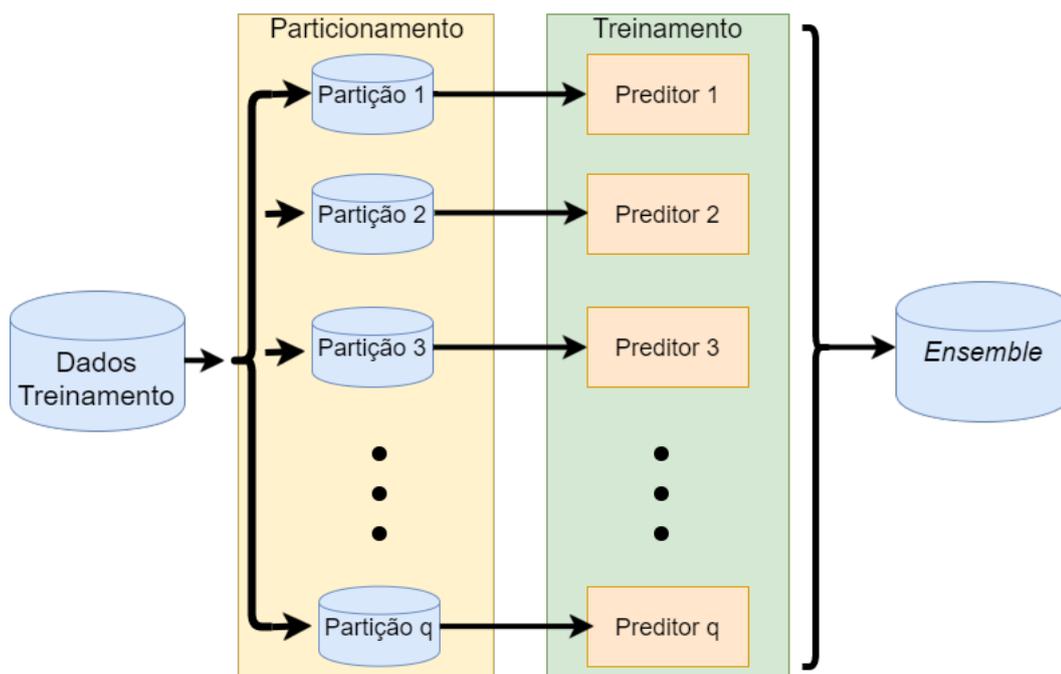


Figura 3.5: Fase de Treinamento dos Modelos de Previsão

Na etapa de treinamento serão treinados q modelos, ou seja, um modelo para cada partição criada. Na presente dissertação o modelo utilizado foi RNA do tipo MLP, por três razões: (i) por ser o modelo de Aprendizado de Máquina mais comum em previsão de séries temporais (WANG; ZENG; CHEN, 2015). (ii) pelo o fator aleatório contido no treinamento (SCHMIDT; KRAAIJVELD; DUIN, 1992), ou seja, cada RNA tem seus pesos iniciais gerados aleatoriamente e isso é uma forma de tornar os modelos diferentes, que é um ponto importante para compor um *Ensemble* de preditores. (iii) e o último motivo se dar ao fato de que no trabalho de (NOVA FILHO; PESSOA, 2015) foi selecionado RNAs e com isso para efeito comparativo é importante que

sejam os mesmo modelos. Ao término dessa etapa, o *Ensemble* estará formado com todos os modelos que foram treinados.

3.2.1.2 Treinamento dos Modelos de Combinação

Após a etapa de treinamento dos modelos de previsão, a qual resultou em um *Ensemble* com q preditores, é realizado o treinamento dos modelos de combinação. Essa etapa tem como entrada o conjunto de dados de treinamento utilizado na etapa anterior e o conjunto de preditores. O objetivo é treinar os modelos de combinação com base na previsão realizada pelos os modelos contidos no *Ensemble*.

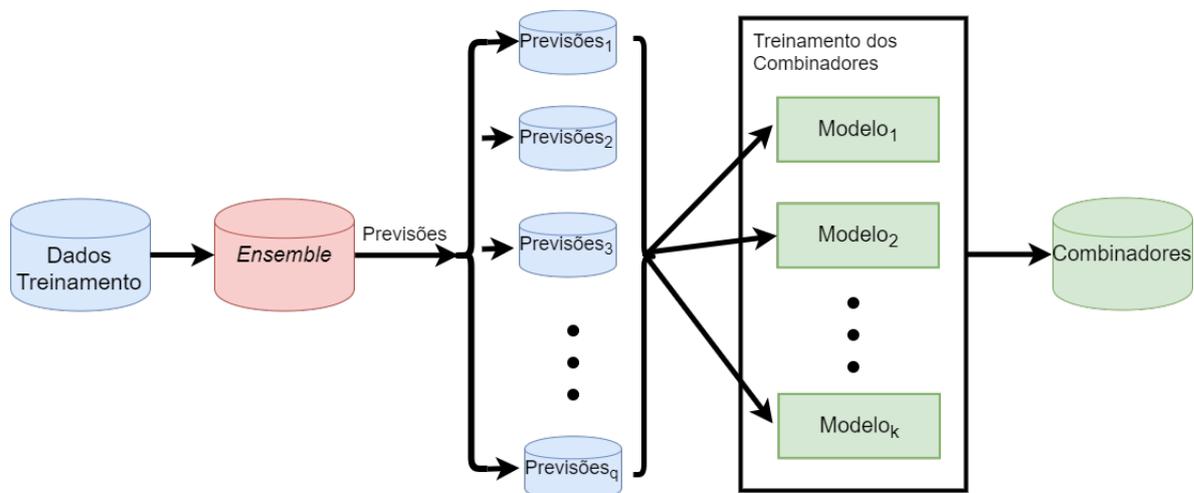


Figura 3.6: Fase de Treinamento dos Modelos de Combinação

A Figura 3.6 ilustra o processo de treinamento, através dela é possível observar que o conjunto completo de treinamento é passado novamente para todos os modelos treinados com o objetivo de obter as suas previsões. Dessa forma, cada modelo resultará em uma nova série de previsões. A composição das q séries de previsões dos modelos irão resultar no conjunto de treinamento dos modelos de combinação. Esse conjunto pode ser expresso pela seguinte matriz:

$$\mathbf{X} = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1j} \\ Y_{21} & Y_{22} & Y_{23} & \dots & Y_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{i1} & Y_{i2} & Y_{i3} & \dots & Y_{ij} \end{bmatrix}, \quad (3.2)$$

em que, Y_{ij} é a previsão do modelo j para o padrão i . Dessa forma a matriz X é composta por todas as previsões dos modelos para os respectivos padrões, com isso, a quantidade de colunas será igual a quantidade de partições q e conseqüentemente igual a quantidade de modelos dentro do *Ensemble*, já a quantidade de linhas será a mesma da quantidade de padrões na série de treinamento dos modelos de previsão.

Após a construção do conjunto de previsões X , o mesmo será dividido em dois subconjuntos, um para treinar os modelos de combinação e outro para a validação. Após a divisão, os

modelos serão treinados para realizar a combinação. Para tal, cada modelo terá como entrada o conjunto de previsões dos modelos do *Ensemble*: $x = \{Y_1, Y_2, \dots, Y_j\}$ e como saída o valor final da previsão da série real em um passo à frente \hat{Z}_{t+1} . Cada configuração do modelo é validada através dos dados de validação após o treinamento, escolhendo as configurações que apresente os melhores desempenhos nos dados de validação.

Dessa forma, o problema de previsão utilizando os modelos de combinação da arquitetura pode ser expresso por: $\hat{Z}_f = g(x)$. Onde deseja-se que a diferença entre \hat{Z}_f e Z_{t+1} seja mínima.

Na presente dissertação, a função $g(x)$ representa os modelos utilizados para a combinação: MLP, SVR, *Perceptron* e Regressão Linear Múltipla (RLM). Os modelos distintos foram escolhidos por apresentarem diferentes formas de resolução do problema de combinação. A MLP tem como motivação para ser um modelo de combinação, pois já foi utilizada em outros trabalhos e apresentou bons resultados. A SVR foi utilizada por conta que em problemas de regressão é considerado um dos modelos com melhores desempenhos (KAVOUSI-FARD; SAMET; MARZBANI, 2014). O *Perceptron*, por ser a forma mais simples de uma RNA para problemas lineares, podendo ser uma abordagem linear que apresente bons resultados. E por fim, a RLM, que é uma forma de combinação já conhecida na literatura (GRANGER; RAMANATHAN, 1984).

Além desses modelos, outras duas formas de combinação foram utilizadas na análise de comportamento. A Média simples e a Mediana, ambas são encontradas em diversos trabalhos na literatura (KOURENTZES; BARROW; CRONE, 2014).

De forma simplificada, o processo de treinamento dos modelos de combinação pode ser representado através do Algoritmo 1.

Algoritmo 1: Treinamento dos modelos de combinação

Entrada: Série de Treinamento

Saída: Modelos de combinação

início

X = obterPrevisãoDosModelos(entradaTreinamento);
dadosTreinamento, dadosValidação = dividirDados(X);

repita

modeloTreinado = treinarModelo(dadosTreinamento);
previsoesValidação = obterPrevisões(modeloTreinado, dadosValidação);
desempenhoDoModelo = calcularDesempenho(previsoesValidação);

se *desempenhoDoModelo* > *desempenhoAtual* **então**

desempenhoAtual = desempenhoDoModelo;
melhorModelo = modeloTreinado;

fim

até número máximo de modelos de combinação;

retorna o conjunto de modelos treinados;

fim

3.2.2 Fase de Teste

Após os modelos de previsão e combinação terem sido treinados, inicia-se a fase de Teste, que tem como objetivo testar a generalização do modelo para novos dados. Como ilustra a Figura 3.7, o processo de teste acontece em duas etapas: previsão dos modelos do *Ensemble* e combinação.

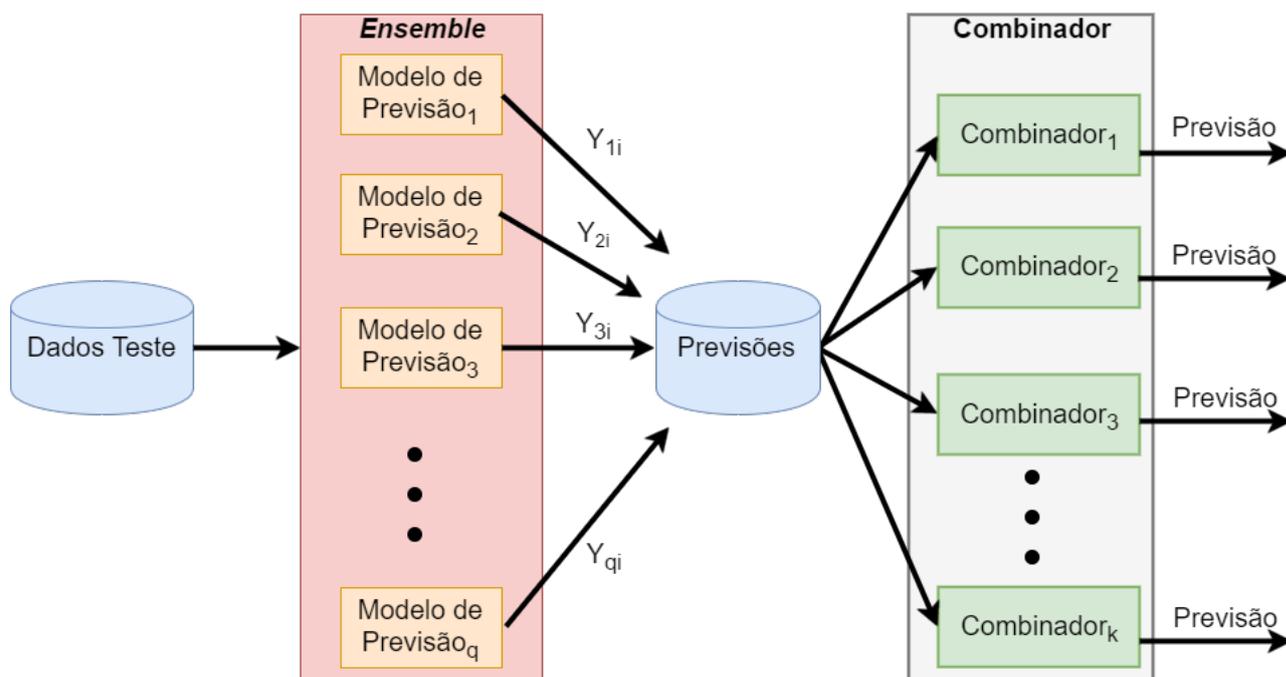


Figura 3.7: Fase de Teste dos Modelos de Combinação

3.2.2.1 Previsão do Ensemble

Essa etapa tem como entrada a janela da série de teste com os *lags* selecionados que antecedem o valor a ser previsto e o *Ensemble* de Preditores. Tem como objetivo realizar as previsões de cada modelo contido no *Ensemble*.

Para tal, considerando $Z_i = \{z_1, z_2, \dots, z_l\}$ como o padrão de entrada (i), onde z_l são os *Lags*, o Z_i será passado para cada modelo do *Ensemble* que retornará sua previsão y_{ji} , resultando assim em um conjunto de previsões $\{y_{1i}, y_{2i}, \dots, y_{ji}\}$, onde j é a quantidade de modelos de previsão no *Ensemble*.

3.2.2.2 Combinação

Essa etapa tem como entrada o conjunto de previsões: $\{y_{1i}, y_{2i}, \dots, y_{ji}\}$ e o conjunto de modelos de combinação. Tem como objetivo realizar a combinação das previsões obtidas. Para tal, cada modelo de combinação irá receber o conjunto de previsões como entrada e retornar a previsão combinada. Ao término dessa etapa, serão obtidas as previsões finais resultantes da

combinação dos modelos e serão utilizadas para a análise de desempenho através de diferentes métricas.

Em seguida, a arquitetura de combinação proposta foi apresentada, a qual tem como objetivo obter um desempenho mais acurado que a abordagem de seleção e analisar qual o comportamento de diferentes abordagens de combinação. As abordagens utilizadas podem ser caracterizadas como linear: média simples, mediana, regressão linear e *perceptron*; ou não linear: RNA e SVR. Porém, a arquitetura de combinação não se restringe a essas abordagens, podendo ser substituídas por outras abordagens de combinação.

Através dessas arquiteturas é possível realizar uma análise comparativa entre as abordagens de combinação e a abordagem de seleção e verificar em quais séries temporais cada abordagem obtém os melhores desempenhos.

4

METODOLOGIA DOS EXPERIMENTOS

Neste capítulo são apresentadas as bases de dados, medidas de desempenho, testes estatísticos e o processo da análise comparativa, utilizados na pesquisa.

4.1 Base de Dados

Para avaliar o desempenho das abordagens implementadas, foram utilizadas quatro séries temporais distintas: Laser, Mackey-Glass, Microsoft e Goldman Sachs.

4.1.1 Série Laser

Laser é uma série estacionária e caótica gerada por experimentos utilizando um laser infravermelho do laboratório de física no Instituto Santa Fé¹. Séries caóticas são caracterizadas por conterem alta sensibilidade à pequenas variações, de forma que, uma pequena variação no passado pode acarretar em grandes mudanças no futuro (LORENZ, 1963). A série Laser já foi citada em diversos trabalhos: (WU; LEE, 2015; JIA et al., 2016; CHEN; LEE, 2015; MIRANIAN; ABDOLLAHZADE, 2013; SERGIO; LIMA; LUDERMIR, 2016). A Figura 4.1 mostra a série utilizada e a Tabela (4.1) apresenta as estatísticas da série.

Estatística	Valores
Média	59,894
Desvio Padrão	46,875
Mínimo	2
Máximo	255
Quantidade de pontos	1000

Tabela 4.1: Estatísticas da série Laser

¹<https://www-psych.stanford.edu/andreas/Time-Series/SantaFe.html>

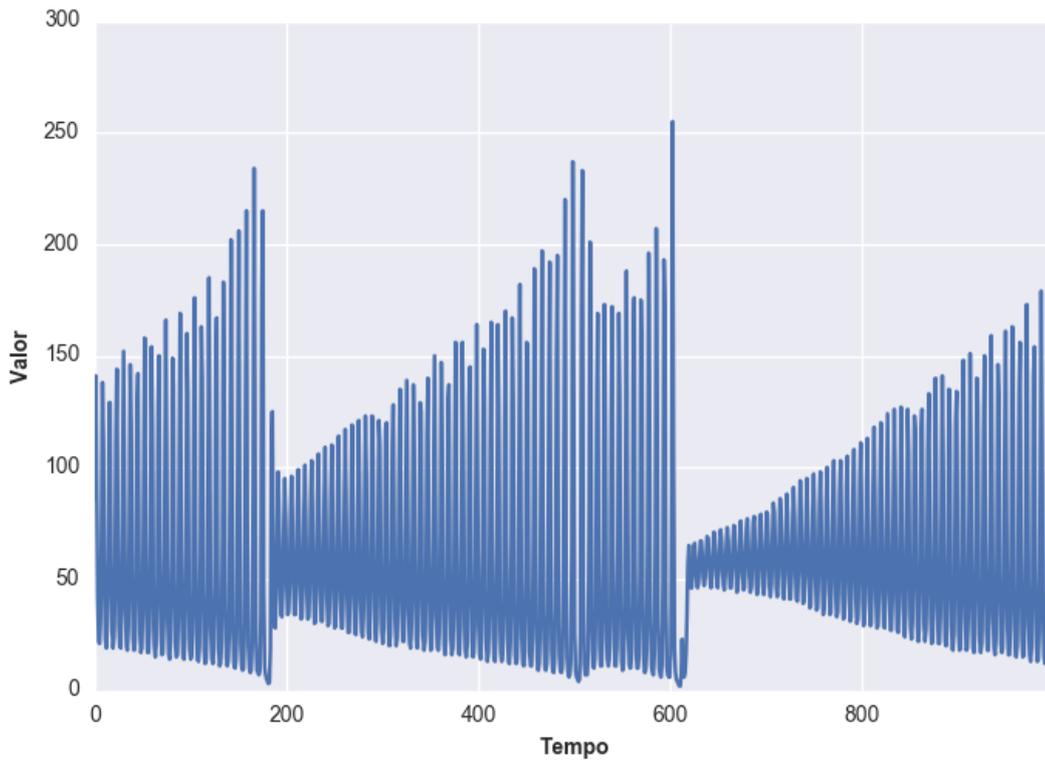


Figura 4.1: Série Laser

4.1.2 Série Mackey-Glass

Mackey-Glass é uma série estacionária e caótica, gerada pela Equação 4.1:

$$\frac{dx}{dt} = \frac{\beta x_{\tau}}{1 + x_{\tau}^n} - \gamma x, \quad (4.1)$$

em que β, γ, τ, n são número reais.

Segundo (ZHAO et al., 2014) a série Mackey-Glass tem alta complexidade de previsão, tornando uma série muito utilizada para avaliar o desempenho do modelo de previsão. Na literatura a série pode ser encontrada em diversos trabalhos: (MIRANIAN; ABDOLLAHZADE, 2013; LI; PI; JIANG, 2013; SERGIO; LIMA; LUDERMIR, 2016; GAXIOLA et al., 2014).

Estatística	Valores
Média	0,863
Desvio Padrão	0,318
Mínimo	0,2020
Máximo	1,369
Quantidade de pontos	700

Tabela 4.2: Estatísticas da série Mackey-Glass

A Figura 4.2 mostra a série completa utilizada, e a Tabela (4.2) apresenta as estatísticas

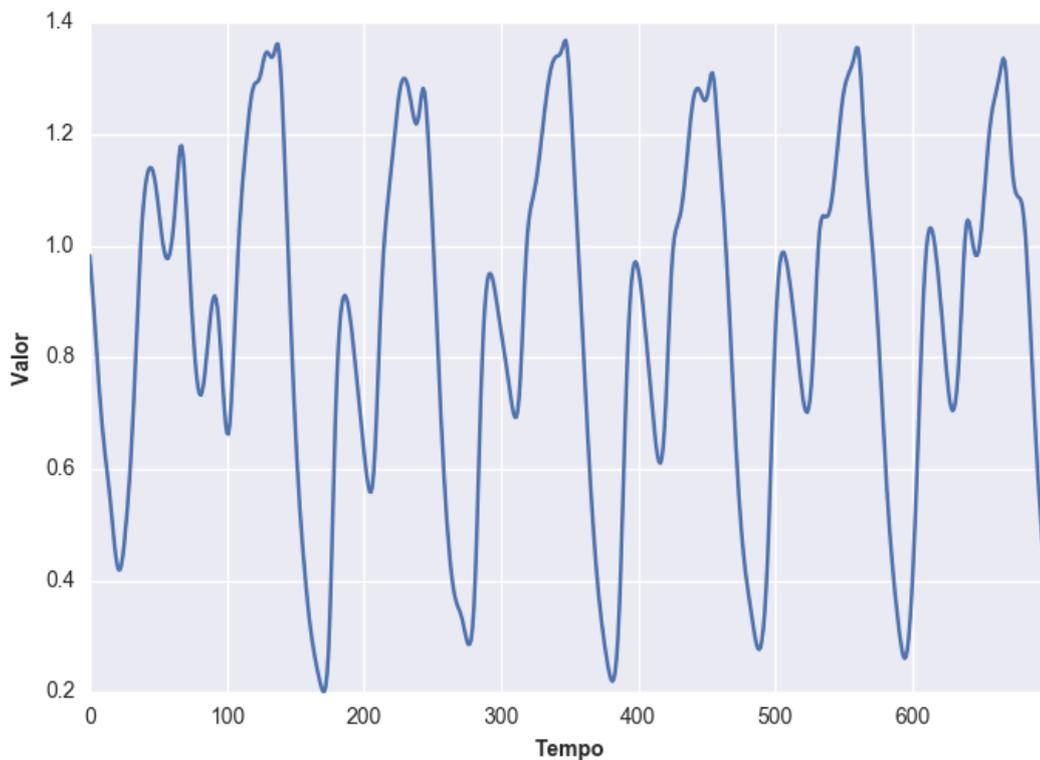


Figura 4.2: Série Mackey-Glass

da série.

4.1.3 Série Microsoft

A série Microsoft é uma base de dados contendo os valores de fechamento diário da ação referente à empresa Microsoft na bolsa NASDAQ¹. É uma série não estacionária, contém alta volatilidade ao longo do período e forte comportamento de aleatoriedade, características comum em séries financeiras (COWPERTWAIT; METCALFE, 2009). A utilização da série pode ser encontrada em vários trabalhos da literatura: (KAZEM et al., 2013; TICKNOR, 2013; NGUYEN; SHIRAI; VELCIN, 2015; ALAM; LJUNGEHED, 2015).

Estatística	Valores
Média	25,98
Desvio Padrão	2,473
Mínimo	21,08
Máximo	31,45
Quantidade de pontos	754

Tabela 4.3: Estatísticas da série Microsoft

¹<http://finance.yahoo.com/quote/MSFT>

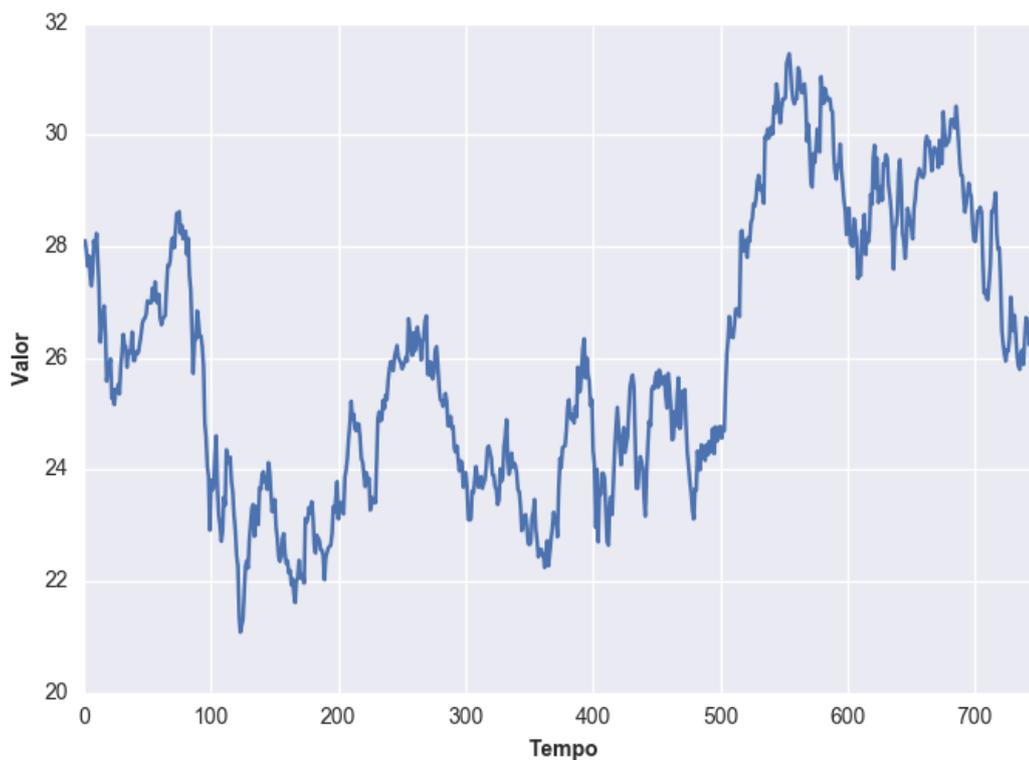


Figura 4.3: Série Microsoft

A Figura 4.3 mostra a série Microsoft no período de 04/01/2010 até 31/12/2012. E a Tabela 4.3 mostra as estatísticas da série.

4.1.4 Série Goldman Sachs

É uma série não estacionária financeira contendo o fechamento diário da ação do grupo financeiro Goldman Sachs na bolsa de *New York* (NYSE)¹. Assim como a série Microsoft, a série Goldman Sachs tem alto fator aleatório, tornando a série difícil de ser prevista corretamente. A sua utilização pode ser encontrada em alguns trabalhos da literatura: (TICKNOR, 2013; ALAM; LJUNGEHED, 2015; FRANKEL, 2014). A Figura 4.4 mostra a série Goldman Sachs no período de 04/01/2010 até 31/12/2012. A Tabela 4.4 mostra as estatísticas da série.

4.2 Métricas de Desempenho

Para comparar o desempenho dos modelos foram utilizadas diversas métricas, de forma que cada modelo possa ser avaliado através de diferentes perspectivas. Para tal, as seguintes métricas de avaliação foram utilizadas: Erro Médio Absoluto em Percentual (MAPE), *Prediction*

¹<http://finance.yahoo.com/quote/GS-PC>

Estatística	Valores
Média	127,98
Desvio Padrão	23,979
Mínimo	85,29
Máximo	176,78
Quantidade de pontos	754

Tabela 4.4: Estatísticas da série Goldman Sachs

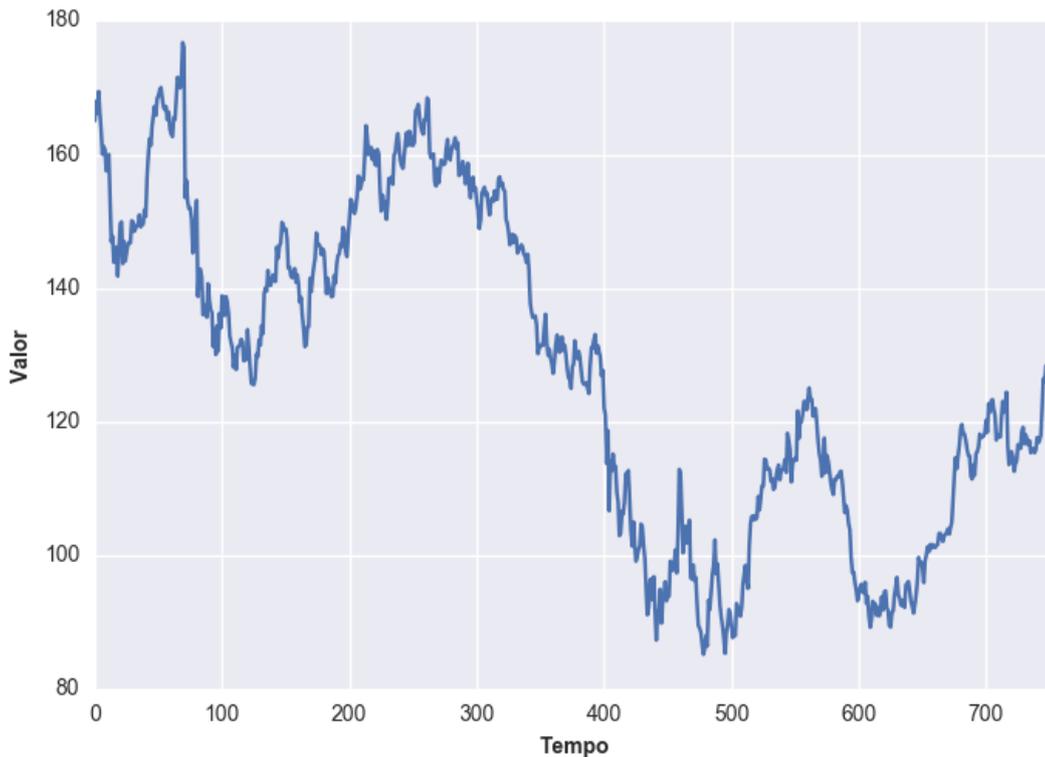


Figura 4.4: Série Goldman Sachs

of Change in Direction (POCID), Average Relative Variance (ARV), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Theil e Fitness.

O MAPE calcula o percentual de erro médio da previsão independente da escala dos valores e indica a margem de erro da previsão em percentual. Ao utilizar o MAPE para avaliar o modelo, quanto mais próximo de zero o valor do MAPE melhor é o modelo. O MAPE é calculado através da Equação 4.3:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{desejado(i) - previsto(i)}{desejado(i)} \right|, \quad (4.2)$$

tal que, N é o número total de pontos da série.

O POCID calcula a frequência de acertos do modelo quanto à tendência do valor futuro. Por exemplo, se o valor previsto segue a mesma tendência do valor real, logo o POCID será

incrementado com o valor de 1, caso contrário, não terá incremento. O valor do POCID pode variar entre $[0, 100]$, de forma que, quanto mais próximo de 100 o valor, melhor é o modelo. O POCID é calculado através das Equações 4.3 e 4.4:

$$POCID = \frac{100}{N} \sum_{i=1}^N X_i, \quad (4.3)$$

em que X_i é calculado através da equação:

$$X_i = \begin{cases} 1, & \text{se } (desejado(t) - desejado(t-1))(previsto(i) - previsto(t-1)) > 0 \\ 0, & \text{caso contrário} \end{cases} \quad (4.4)$$

O ARV compara o desempenho do modelo em relação à previsão da média da série temporal. Caso $ARV = 1$, o modelo avaliado tem um desempenho igual a utilizar a média da série como previsão, caso $ARV > 1$, o modelo é inferior, caso contrário, onde $ARV < 1$, o modelo é melhor que utilizar a média como previsão. Para calcular o ARV é utilizado a Equação 4.5:

$$ARV = \frac{\sum_{i=1}^N (previsto(i) - desejado(i))^2}{\sum_{i=1}^N (desejado(i) - \overline{desejado})^2} \quad (4.5)$$

O *Theil* avalia se o modelo é melhor que o modelo *random walk* (REIF, 2009). Um modelo *Random Walk* é considerado o modelo mais simples para realizar a previsão, pois utiliza apenas o valor anterior do ponto a ser previsto, acrescido de um valor aleatório. Ao utilizar essa métrica para avaliar o modelo, caso *Theil* = 1 indica que o modelo tem o mesmo desempenho do *random walk*, caso contrário indica que o modelo é melhor. O *Theil* é calculado através da Equação 4.6:

$$Theil = \frac{\sum_{i=1}^N (desejado(i) - previsto(i))^2}{\sum_{i=1}^N (desejado(i) - desejado(i-1))^2} \quad (4.6)$$

O MSE calcula o erro quadrático médio das previsões, dessa forma, quanto mais próximo de zero o valor do MSE, melhor são as previsões do modelo avaliado. O cálculo é realizado através da Equação 4.7:

$$MSE = \frac{1}{N} \sum_{i=1}^N (desejado(i) - previsto(i))^2 \quad (4.7)$$

Fitness é uma função apresentada em (MATTOS NETO et al., 2015), composta de diferentes métricas, utilizada com o objetivo de avaliar o modelo levando em consideração diferentes perspectivas representadas por apenas um valor. Para tal, quanto maior o valor *fitness* obtido melhor é o modelo. O cálculo é realizado através da Equação 4.9:

$$Fitness = \frac{POCID}{1 + MSE + MAPE + Theil + ARV} \quad (4.8)$$

Essa função foi utilizada principalmente para definir quais as melhores configurações

para os modelos. Tal que, a configuração escolhida, teve como resultado o maior valor *fitness* para o conjunto de validação.

4.3 Testes Estatísticos

Para realizar a comparação entre os desempenhos dos modelos, se faz necessário a utilização de testes estatísticos, que avaliam se as amostras dos resultados apresentam diferenças estatisticamente significativas. O processo para a utilização dos testes estatísticos consiste em duas etapas: análise de normalidade e análise de distribuições.

4.3.1 Análise de Normalidade

Para definir qual teste de hipótese entre as amostras será utilizado é necessário que antes se faça um teste de normalidade nos dados, principalmente quando a amostra utilizada é pequena. Para tal, foi utilizado o teste *Shapiro-Wilk* que consiste nos seguintes passos:

1. Elaboração das Hipóteses:

$$H_0 : A amostra provém de uma população Normal$$

$$H_1 : A amostra não provém de uma população Normal$$

2. Definir o nível de significância do teste (α), neste trabalho o $\alpha = 0.05$, dessa forma o nível de confiança é de 95% ;
3. Realiza o cálculo da estatística, obtendo o valor $W_{calculado}$;
4. Realiza a tomada de decisão, em rejeitar a H_0 caso o $W_{calculado} < W_\alpha$, onde W_α é o valor crítico da distribuição de *Shapiro-Wilk*;

4.3.2 Análise de Distribuições

A análise de distribuições tem como objetivo conhecer se as amostras apresentam semelhanças. Para tal, caso a amostra venha de distribuição normal é utilizado o teste *t-Student* que consiste nos seguintes passos:

1. Elaboração das Hipóteses:

$$H_0 : \mu_1 = \mu_2 : As amostras tem as mesmas variância.$$

$$H_1 : \mu_1 \neq \mu_2 : As amostras não tem as mesmas variância.$$

2. Definir o nível de significância do teste (α), neste trabalho o $\alpha = 0.05$, dessa forma o nível de confiança é de 95% ;

3. Realiza o cálculo da estatística, obtendo o valor $t_{calculado}$;
4. Realiza a tomada de decisão, em rejeitar a H_0 caso o $t_{calculado} \neq t_{\alpha}$, onde t_{α} é o valor crítico da distribuição *t-Student*.

Caso a amostra não venha de uma distribuição normal, o Teste Wilcoxon será utilizado através dos seguintes passos:

1. Elaboração das Hipóteses:

$H_0 : \Delta = 0$: A mediana da diferença entre as amostras é nula.

$H_1 : \Delta \neq 0$: A mediana da diferença entre as amostras não é nula.

2. Definir o nível de significância do teste (α), neste trabalho o $\alpha = 0.05$, dessa forma o nível de confiança é de 95% ;
3. Realiza o cálculo da estatística, obtendo o valor $w_{calculado}$;
4. Realiza a tomada de decisão, em rejeitar a H_0 caso o $w_{calculado} < w_{\alpha 1}$, ou $w_{calculado} > w_{\alpha 2}$ onde $w_{\alpha 1}$ e $w_{\alpha 2}$ são os valores críticos da distribuição Wilcoxon.

4.4 Metodologia para a Análise Comparativa

Essa seção consiste em especificar o processo e os parâmetros utilizados para comparar as abordagens implementadas, e está dividida em 3 etapas: configuração dos dados, configuração dos modelos e processo de comparação.

4.4.1 Configuração dos Dados

Cada série temporal utilizada passa por três etapas: A primeira etapa consiste em realizar a normalização dos dados, utilizando a Equação 2.5. Na segunda etapa é realizada uma análise para verificar se a série é estacionária, caso não seja estacionária, será utilizada a Equação 2.2 para aplicar a diferenciação na série de forma à torna-la estacionária . A terceira etapa consiste em dividir a série em dois conjuntos, o primeiro é o conjunto de treinamento utilizado para treinar e validar os modelos, e o segundo é o conjunto de teste, utilizado para avaliar o desempenho dos modelos.

Para a divisão do conjunto de dados, foram utilizados 75% da série para treinamento e validação, e os 25% restantes para teste. Os dados de validação são selecionados dinamicamente como mostra na Seção 3.1.1.3.

4.4.2 Configuração para os Modelos

Após os conjuntos de dados estarem definidos e tratados, começa a etapa de treinamento dos modelos de previsão. As configurações utilizadas nessa etapa foram baseadas nos parâmetros que apresentaram os melhores resultados no trabalho de (NOVA FILHO; PESSOA, 2015).

Parâmetro	Goldman Sachs	Microsoft	Mackey-Glass	Laser
Interseção entre as Partições (pint)	50%, 70%, 90%			
Lags Máximo (lmax)	20	30	30	20
Tamanho das Partições	150	350	450	350

Tabela 4.5: Parâmetros para as partições das Séries

Na Tabela (4.5) são apresentados os parâmetros que foram utilizados para definir o número de partições em cada série com base na Equação 3.1. A taxa de interseção entre as partições (*pint*) foi variada com o objetivo de testar o comportamento das abordagens em relação a diversidade nas partições. Dessa forma, foi analisado para cada série qual a melhor taxa de interseção e utilizada junto com os outros parâmetros já estabelecidos. E o valor do tamanho das partições e dos lags máximos foram utilizados os valores que apresentaram os melhores resultados no trabalho de (NOVA FILHO; PESSOA, 2015).

O modelo de previsão utilizado foi a RNA do tipo MLP treinada através do algoritmo de aprendizado *BFGS-Quasi Newton* (FLETCHER, 2013). Para a sua utilização foi necessário realizar um processo para selecionar os parâmetros da MLP para cada partição. Os parâmetros das arquiteturas foram variados de acordo com a Tabela (4.6).

Parâmetro	Valores
Nós Escondidos	1, 5, 10, 15, 20, 25
Taxas de Aprendizagem	0.001, 0.003, 0.005, 0.01

Tabela 4.6: Parâmetros para a Rede MLP

O processo de seleção de parâmetros consiste em três etapas. Na primeira etapa, uma combinação de parâmetros é passada para o modelo. Na segunda etapa, os modelos são treinados com as configurações recebidas e armazenados em um banco de dados para uso posterior. A terceira etapa é realizada após todas as configurações serem treinadas. Nessa etapa, o modelo é avaliado através do desempenho obtido no conjunto de validação. O modelo que apresentar o maior valor da função *Fitness* será o escolhido para representar a partição. Esse processo será repetido para cada partição, de forma que ao fim do treinamento, seja formado o *Ensemble* de preditores.

Após o *Ensemble* de preditores está formado é o momento de treinar as abordagens de combinação. Dessa forma, as previsões obtidas no conjunto de treinamento e validação são apresentadas para as abordagens de combinação. As abordagens serão treinadas através do

conjunto de previsão para a série de treinamento e avaliadas com base no valor da função *Fitness* para o conjunto de validação utilizado nos preditores.

As RNAs do tipo MLP utilizadas para a combinação tiveram o mesmo conjunto de parâmetros apresentados na Tabela (4.6). A SVR utilizada para realizar a combinação é constituída de uma função *kernel* RBF e os seus parâmetros foram variados de acordo com a Tabela (4.7). O *Perceptron* utilizou a função Hiperbólica Tangente Sigmoide (HARRINGTON, 1993) como função de transferência. E por fim, a RLM utilizou o *Iteratively Reweighted Least-Squares* (IRLS) como técnica de estimação para os pesos.

Parâmetro	Valores
Complexidade(C)	1, 1.5, ..., 10
Variável de folga (ϵ)	0.001, 0.051, ..., 1
Gamma (RBF)	1, 1.5, ..., 100

Tabela 4.7: Parâmetros para a SVR

4.4.3 Processo de Comparação

Para responder as questões de pesquisa elaboradas no início da dissertação, é necessário realizar um processo de comparação entre todas as abordagens utilizadas. Esse processo pode ser dividido em duas etapas. A primeira consiste em encontrar o melhor valor para o parâmetro (*pint*), para tal, é realizado uma análise exploratória dos desempenhos através de *box-plots*, média, desvio padrão. E a segunda etapa, consiste em verificar entre as abordagens de combinação e a arquitetura de seleção proposta por (NOVA FILHO; PESSOA, 2015) qual apresenta o melhor desempenho, para tal serão utilizadas as métricas apresentadas anteriormente, assim como, os testes estatísticos para inferir se as abordagens apresentam diferenças significativas. As amostras em análise correspondem aos desempenhos obtidos nas cinco execuções realizadas para cada abordagem.

4.5 Considerações Finais

A metodologia desta dissertação descreve as bases de dados, as configurações, as métricas de avaliação e o processo de análise comparativa, utilizados para atingir os objetivos de pesquisa. Para tal, são utilizadas diferentes séries temporais: Laser, Mackey-Glass, Microsoft e Goldman Sachs e através das métricas de desempenho apresentadas, as abordagens utilizadas são avaliadas levando em considerações diferentes perspectivas.

Para complementar as métricas, são utilizados testes estatísticos de análise de hipóteses, levando em considerações às amostras obtidas com base nas cinco execuções realizadas para cada abordagem.

As configurações dos dados e modelos, são selecionadas com base no desempenho do preditor, obtido de acordo com cada combinação de parâmetros. Com isso, é realizado um processo iterativo onde cada modelo é treinado com uma configuração diferente e armazenado em uma base de dados, a fim de selecionar o melhor preditor para cada partição dos dados.

O processo de análise comparativa consiste em duas etapas, onde a primeira tem o objetivo de verificar qual a taxa de interseção que corresponde aos melhores desempenhos nas abordagens de combinação, e a segunda, consiste em comparar entre as abordagens de combinação e a seleção quais apresentam melhores resultados, de forma que, a melhor abordagem é a que apresentar melhores desempenhos na maioria das métricas.

5

ANÁLISE E RESULTADOS

Este capítulo apresenta os resultados das previsões obtidas através das diferentes abordagens para as séries temporais utilizadas. As análises são realizadas a partir das diferentes métricas de desempenho apresentadas no capítulo anterior utilizando os resultados obtidos em cinco execuções de cada abordagem.

Os experimentos foram realizados utilizando um notebook com processador Intel Core i5-4200U com 8 GB de memória RAM. As abordagens foram implementadas utilizando o Matlab R2014a¹.

O capítulo está dividido em cinco seções. As quatro primeiras correspondem aos resultados obtidos nas diferentes bases de dados. E na última seção estão as considerações em relação aos resultados obtidos.

5.1 Série Goldman Sachs

A série é composta por 754 pontos, destes, os 567 primeiros foram utilizados para treinamento e validação, e os últimos 187 pontos foram utilizados para teste.

5.1.1 Interseção entre as Partições

Esta etapa teve como objetivo encontrar qual a taxa de interseção apresenta o melhor desempenho entre as abordagens de combinação. Para tal, foram coletados os resultados das previsões utilizando as taxas de 50%, 70% e 90%, que resultam em *Ensembles* compostos por 7, 11 e 29 modelos de previsão, respectivamente. Cada modelo foi treinado por uma partição contendo 150 pontos, e cada partição foi composta por janelas de no máximo 20 *lags*. O intuito de variar as taxas é verificar o comportamento dessas abordagens de acordo com a diversidade entre as partições adjacentes.

Para a abordagem de combinação utilizando MLP, através da Figura 5.1 é possível observar em forma de *box-plot*, que para a métrica MAPE os melhores desempenhos foram para

¹<http://www.mathworks.com/products/matlab>

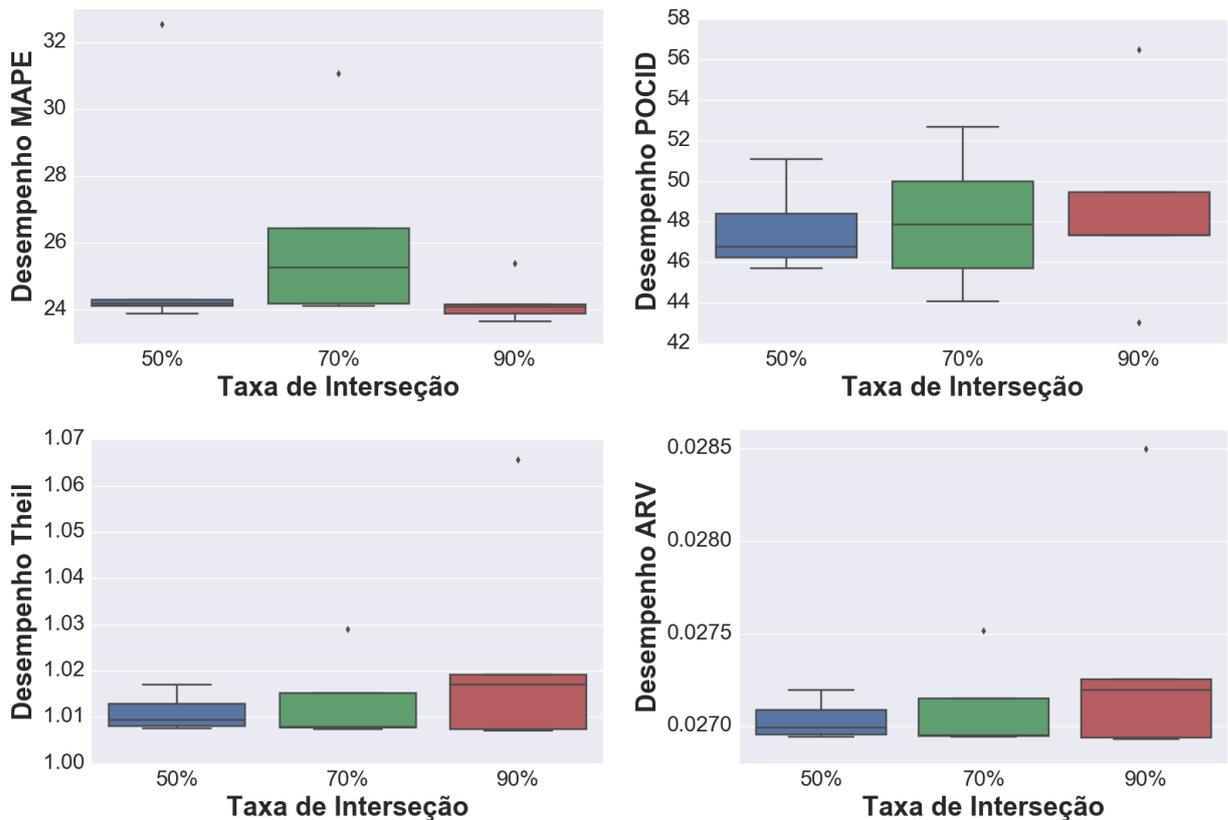


Figura 5.1: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Goldman Sachs

as taxas 50% e 90%. Para o POCID a taxa de 90% apresentou em média melhores desempenhos. Para as métricas *Theil* e *ARV* utilizar a taxa de 50% apresentam melhores desempenhos. Através desses resultados, é possível observar que na maioria das métricas as partições com 50% de interseção entre os dados, apresentam melhores resultados para a abordagem de combinação através da MLP, com isso, para as próximas comparações, será utilizado o desempenho da MLP com taxa de 50%.

Para a abordagem de combinação utilizando SVR, através dos resultados mostrados na Figura 5.2 é possível observar que em relação as métricas *MAPE*, *Theil* e *ARV*, os desempenhos estão melhores distribuídos para a taxa de 90% de interseção. Apenas para a métrica *POCID*, é que utilizar 70% de interseção apresentam melhores resultados. Como a taxa de 90% apresentou os melhores desempenhos na maioria das métricas, serão utilizados os seus resultados para as próximas comparações em relação à abordagem de combinação com SVR.

Para a abordagem de combinação utilizando Regressão Linear, é possível observar nos *box-plots* da Figura 5.3 que para as métricas *MAPE*, *Theil* e *ARV* os melhores resultados foram encontrados utilizando a taxa de interseção de 50%. Enquanto que para a métrica *POCID*, foi a taxa de 70%. No entanto, como a taxa de 50% apresentou melhores desempenhos na maioria das métricas, essa será utilizada para as próximas comparações utilizando a abordagem de

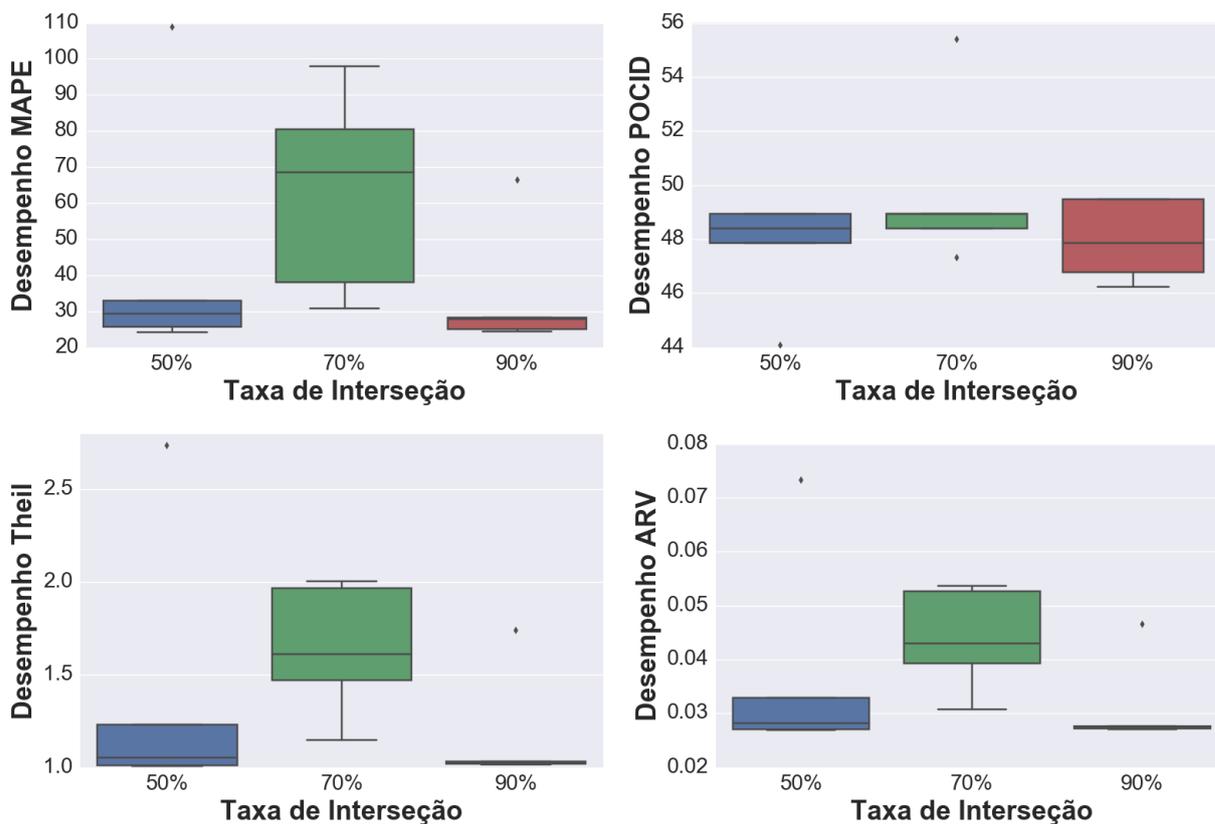


Figura 5.2: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Goldman Sachs

combinação através de Regressão Linear.

Para a abordagem de combinação utilizando *Perceptron*, é possível observar na Figura 5.4 que para as métricas POCID, *Theil* e ARV os melhores resultados são alcançados utilizando a interseção de 70%, enquanto que para a métrica MAPE foi a taxa de 50% que apresentou os melhores resultados. Com isso, para futuras comparações utilizando a abordagem de combinação através do *Perceptron*, serão utilizados os resultados com interseção de 70% pois apresentou melhores desempenhos na maioria das métricas.

Para a abordagem de combinação utilizando a Média Simples, é possível observar nos *box-plots* da Figura 5.5 que para as métricas *Theil* e ARV, os melhores desempenhos estão distribuídos utilizando a interseção de 90%. Enquanto que para as métricas MAPE e POCID é possível observar que em relação ao melhor desempenho da amostra, as taxas 50% para o MAPE e 70% para o POCID apresentam melhores resultados, porém tem uma maior variância, enquanto isso os resultados para a taxa 90% apresentam uma distribuição mais concentrada. Dessa forma, para futuras comparações utilizando a abordagem de combinação através de Média Simples, serão utilizados os resultados com interseção de 90%.

Para a abordagem de combinação utilizando a Mediana, é possível observar através da Figura 5.6 que para as métricas *Theil* e ARV os melhores desempenhos foram utilizando a taxa

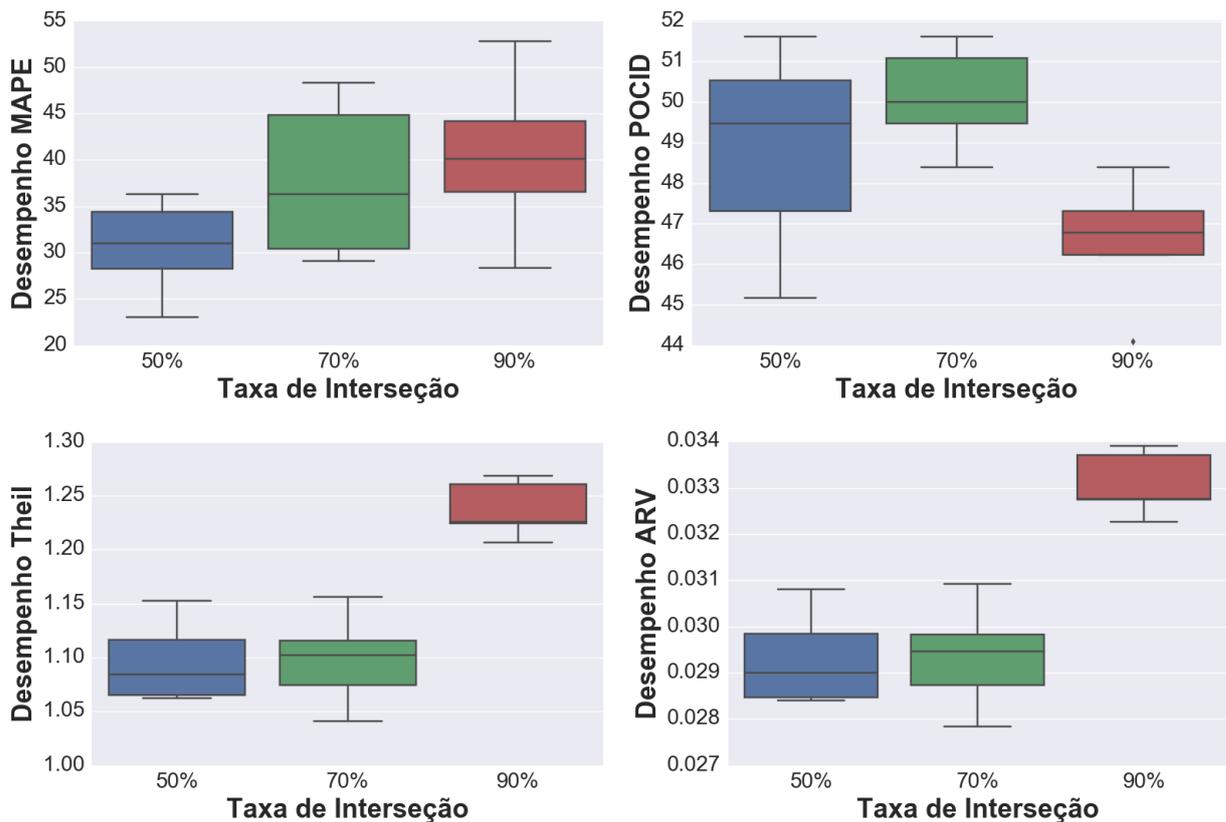


Figura 5.3: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Goldman Sachs

de 90%. Para a métrica MAPE o melhor resultado foi utilizando a taxa de 50%, porém conta com alta variância nos desempenhos, diferente dos resultados com taxa de 90% que apresentam uma variância menor. Enquanto que para a métrica POCID o melhor resultado foi obtido através da taxa de 70%, porém os resultados com a taxa de 90% estão em média melhores distribuídos. Contudo, a taxa de 90% será utilizada em comparações futuras utilizando a abordagem de combinação através da Mediana, por apresentar melhores resultados na maioria das métricas.

Abordagem	Interseção (%)
MLP	50
SVR	90
RLM	50
Perceptron	70
Média	90
Mediana	90

Tabela 5.1: Melhores taxas para as abordagens de combinação para Série Goldman Sachs

A Tabela (5.1) mostra as taxas escolhidas para cada abordagem com base no desempenho médio obtido. Através dessas tabelas é possível verificar que a MLP e a Regressão Linear apresentaram desempenhos melhores ao combinar modelos que tinham 50% dos seus

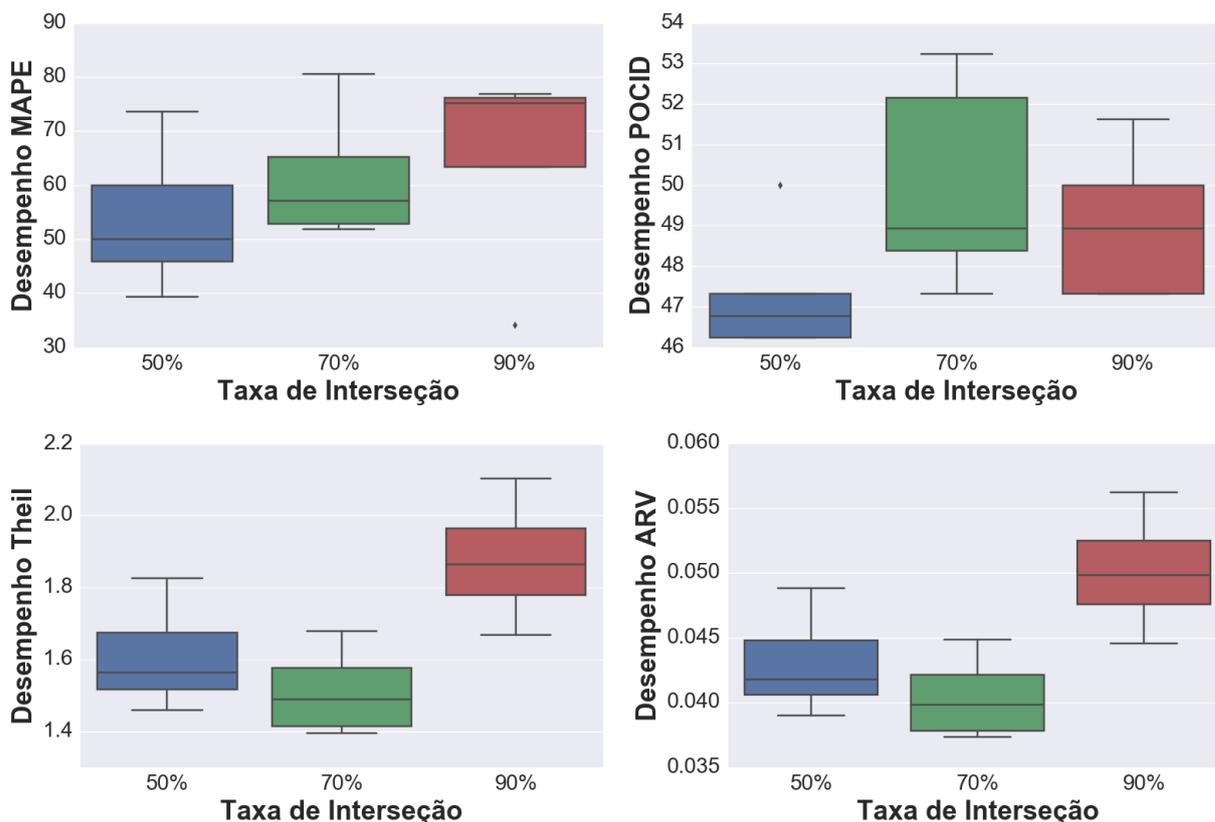


Figura 5.4: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com *Perceptron* na série Goldman Sachs

dados de treinamento compartilhados com outros modelos. Já as abordagens SVR, Média e Mediana, apresentaram melhores resultados para combinar modelos com uma taxa maior de dados compartilhados, e conseqüentemente uma taxa menor de diversidade entre as partições de treinamento.

5.1.2 Análise das Abordagens de Combinação

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
MLP	25.80(3.77)	47.63(2.17)	1.010(0.003)	0.0270(0.0001)
SVR	34.35(18.0)	47.95(1.49)	1.165(0.320)	0.0311(0.008)
RLM	30.58(5.26)	48.81(2.58)	1.095(0.038)	0.0293(0.0010)
Perceptron	61.50(11.88)	50.00(2.55)	1.510(0.117)	0.0403(0.0031)
Média	28.75(1.36)	48.81(1.58)	1.036(0.013)	0.0277(0.0003)
Mediana	25.84(1.49)	46.02(2.33)	1.035(0.005)	0.0276(0.0001)

Tabela 5.2: Resultados das abordagens de combinação para a série Goldman Sachs

Após conhecer quais as melhores taxas de interseção, é realizada a comparação entre as abordagens de combinação. Através da Tabela (5.2) é possível analisar o desempenho médio das

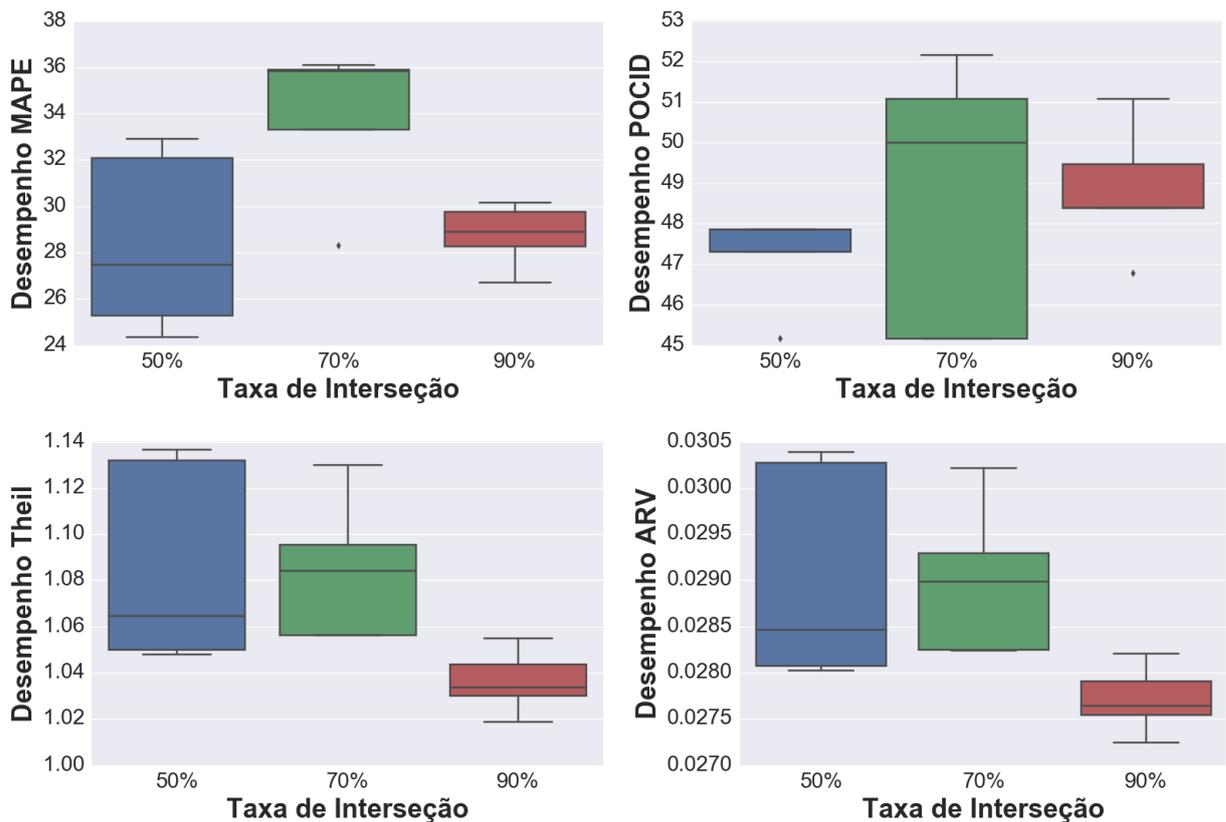


Figura 5.5: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Goldman Sachs

	SVR	RLM	Perceptron	Média	Mediana
MLP	Não Rejeita H_0	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
SVR	—	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
RLM	—	—	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
Perceptron	—	—	—	Rejeita H_0	Rejeita H_0
Média	—	—	—	—	Não Rejeita H_0

Tabela 5.3: Resultados dos Testes para as comparações par a par do *fitness* na série Goldman Sachs

execuções de acordo a métrica. Para as métricas MAPE, *Theil* e ARV as melhores abordagens foram as combinações através da MLP e através da Mediana. Já para o POCID as melhores abordagens foram utilizando o *Perceptron* e a Média.

Para verificar se realmente existe diferença entre os desempenhos das abordagens, foram realizados os testes estatísticos para as amostras (conjunto com o desempenho das 5 execuções) em relação ao *fitness*. Através do teste *Shapiro-Wilk* foi verificado que as amostras não seguem uma distribuição normal, com isso foi utilizado o teste *Wilcoxon* para verificar se existe diferença entre os desempenhos. A hipótese nula foi que os desempenhos de ambos modelos são iguais. Através da Tabela (5.3) é possível observar que a única abordagem que apresentou um *fitness* estatisticamente diferente das demais foi o *Perceptron*.

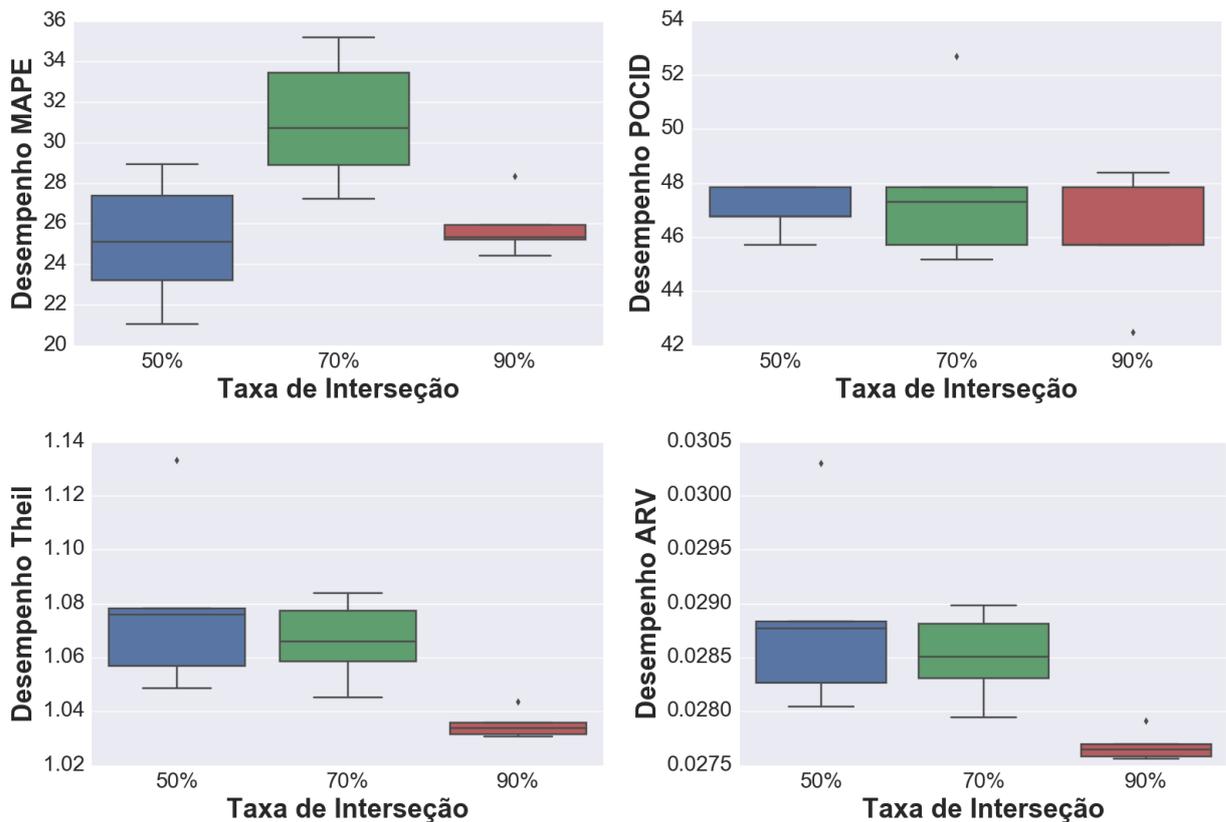


Figura 5.6: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Mediana na série Goldman Sachs

5.1.3 Escolha da Arquitetura

Depois de realizar a análise entre as abordagens de combinação e ter encontrado os melhores valores para a taxa de interseção entre as partições, é realizado uma comparação entre as abordagens de combinação e a abordagem de seleção. Para tal, foi obtido o desempenho da arquitetura de seleção utilizando os parâmetros que apresentaram os melhores resultados no trabalho de (NOVA FILHO; PESSOA, 2015), e comparado com os melhores desempenhos da arquitetura de combinação utilizando as diferentes abordagens de combinação.

Realizando os testes estatísticos para verificar se existe diferença significativa entre os desempenhos das abordagens de combinação com a abordagem de seleção, Tabela (5.6), verificou-se que em relação as métricas MAPE e *Fitness*, a hipótese nula não foi rejeitada, ou seja, que não existe diferença estatisticamente significativa entre os desempenho da seleção com a combinação utilizando MLP e Mediana.

Analisando a Tabela (5.4) que mostra os desempenhos médios, é possível observar que em relação ao MAPE, a seleção foi melhor que a combinação por Regressão e por *Perceptron*, em relação ao POCID a seleção foi melhor que a maioria das abordagens de combinação, perdendo apenas para a combinação por *Perceptron*, em relação ao *Theil* e o ARV a seleção foi melhor apenas para a combinação utilizando o *Perceptron*. A Tabela (5.5) mostra o *Ranking* das melhores

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
Seleção	28.92(4.18)	48.92(1.31)	1.169(0.100)	0.0312(0.0026)
MLP	25.80(3.77)	47.63(2.17)	1.010(0.003)	0.0270(0.0001)
SVR	34.35(18.0)	47.95(1.49)	1.165(0.320)	0.0311(0.008)
RLM	30.58(5.26)	48.81(2.58)	1.095(0.038)	0.0293(0.0010)
Perceptron	61.50(11.88)	50.00(2.55)	1.510(0.117)	0.0403(0.0031)
Média	28.75(1.36)	48.81(1.58)	1.036(0.013)	0.0277(0.0003)
Mediana	25.84(1.49)	46.02(2.33)	1.035(0.005)	0.0276(0.0001)

Tabela 5.4: Comparação entre as abordagens de combinação e a seleção para a série Goldman Sachs

Rank	MAPE	POCID	Theil	ARV
1°	Mediana	Perceptron	Comb. MLP	Comb. MLP
2°	Comb. MLP	Seleção	Comb. Mediana	Comb. Mediana
3°	Comb. Média	Comb. Média	Comb. Média	Comb. Média
4°	Seleção	Comb. Regressão	Comb. Regressão	Comb. Regressão
5°	Comb. Regressão	Comb. SVR	Comb. SVR	Comb. SVR
6°	Comb. SVR	Comb. MLP	Seleção	Seleção
7°	Comb. Perceptron	Comb. Mediana	Comb. Perceptron	Comb. Perceptron

Tabela 5.5: Ranking dos melhores desempenhos médio para a série Goldman Sachs

	Seleção x MLP	Seleção x Mediana
MAPE	Não Rejeita H_0	Não Rejeita H_0
POCID	Rejeita H_0	Rejeita H_0
Theil	Rejeita H_0	Rejeita H_0
ARV	Rejeita H_0	Rejeita H_0
Fitness	Não Rejeita H_0	Não Rejeita H_0

Tabela 5.6: Resultados dos testes estatísticos entre as abordagens de seleção e combinação para a série Goldman Sachs

abordagens, levando em consideração o desempenho médio das abordagens.

	MAPE	POCID	Theil	ARV
Seleção	21.78	50.53	1.0848	0.0290
Comb. MLP	23.89	51.07	1.0075	0.0269
Comb. SVR	24.28	49.46	1.0132	0.0270
Comb. RLM	23.01	51.61	1.0621	0.0284
Comb. Perceptron	51.87	53.22	1.3955	0.037
Comb. Média	26.69	51.07	1.0299	0.0272
Comb. Mediana	24.41	48.38	1.0307	0.027

Tabela 5.7: Comparação entre as diferentes abordagens para a série Goldman Sachs

A Tabela (5.7) mostra os melhores resultados encontrados nas abordagens implementadas neste trabalho. Através dos resultados contidos na tabela é possível observar que a abordagem

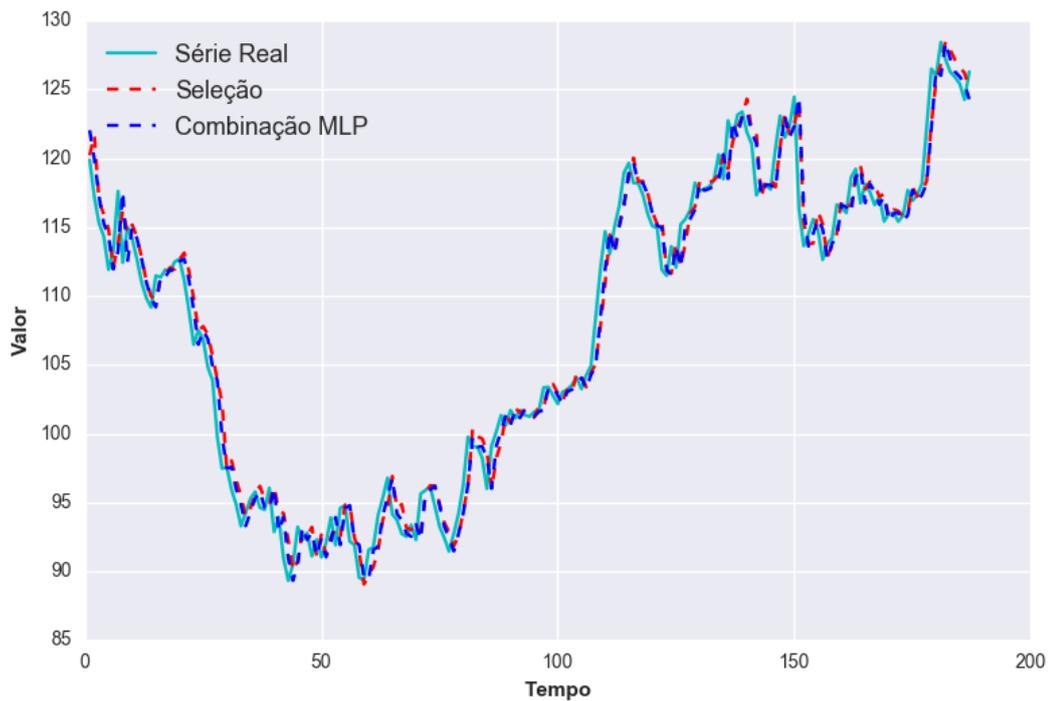


Figura 5.7: Previsões das Arquiteturas para o conjunto de teste da série Goldman Sachs

de seleção apresentou o melhor desempenho apenas para o MAPE.

Contudo, é possível inferir que para a série Goldman Sachs, em relação ao *fitness* que leva em consideração as diferentes métricas e atribui um peso maior para os resultados obtidos do MAPE, as melhores abordagens de combinação não apresentam diferença estatística em seus resultados. Porém, ao analisar os valores médios dos resultados obtidos nas métricas MAPE, Theil e ARV a maioria dos modelos de combinação apresentaram resultados melhores. Contudo, em valores absoluto as melhores abordagens de combinação apresenta um melhor resultado que a abordagem de seleção. Na Figura 5.7 é possível observar as previsões de teste para a abordagem de combinação utilizando MLP e a abordagem de seleção.

5.2 Série Microsoft

A série é composta por 754 pontos, onde os 567 iniciais foram utilizados para treinamento e validação, e os últimos 187 foram utilizados para teste.

O processo de experimentos para a série Microsoft foi dividido em três etapas. Primeira etapa, a análise da taxa de interseção entre os modelos de combinação. Segunda etapa, a análise de desempenho entre as abordagens de combinação. E por fim, a análise entre as abordagens de combinação em comparação com a abordagem de seleção.

5.2.1 Interseção entre as Partições

Nessa etapa foi realizado a comparação entre as diferentes taxas de interseção 50%, 70% e 90% que resultam em *Ensembles* compostos por 3, 4 e 7 modelos de previsão, respectivamente, onde cada preditor foi treinado por uma partição contendo 350 pontos e cada partição foi composta por janelas de no máximo 30 *lags*. O intuito de variar as taxas, como dito anteriormente, é verificar o comportamento dessas abordagens de acordo com a diversidade entre as partições adjacentes.

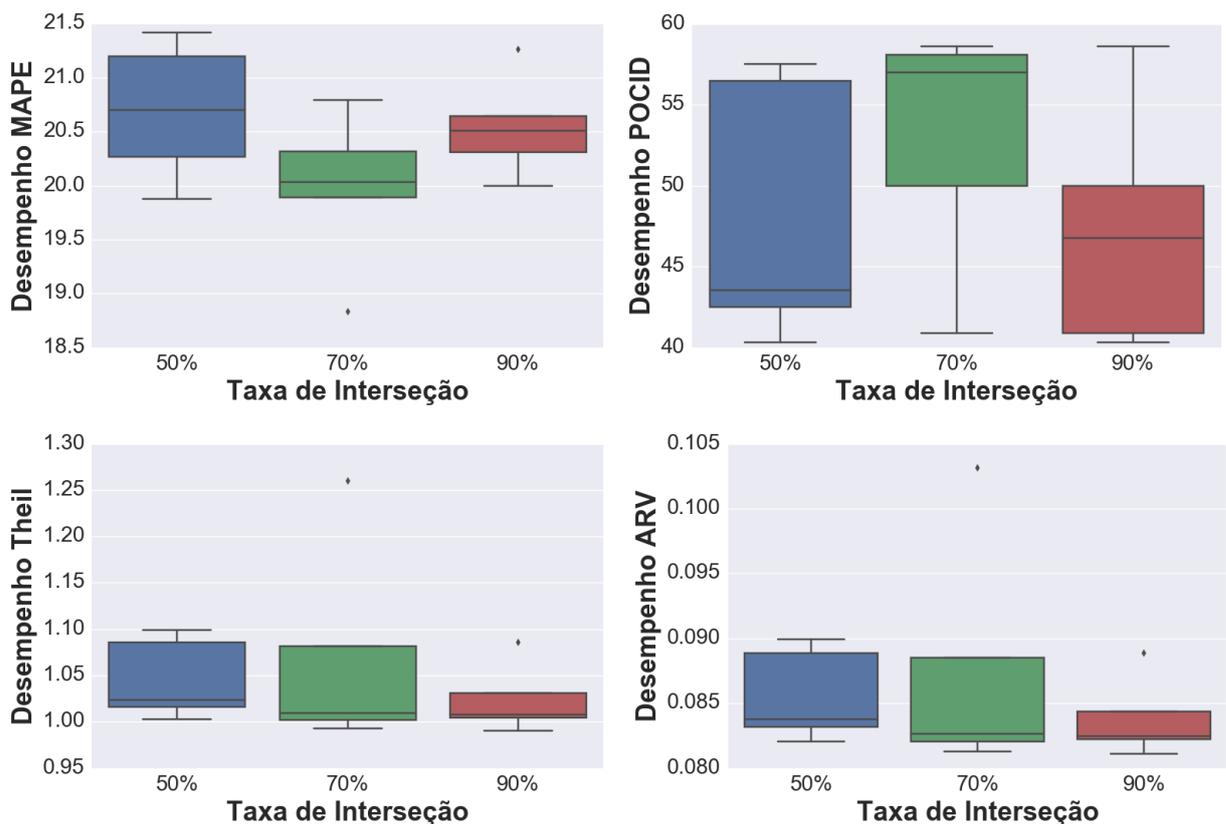


Figura 5.8: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Microsoft

Em relação a abordagem de combinação utilizando MLP, na Figura 5.8 é possível observar em forma de *box-plots* que para as métricas MAPE e POCID os melhores resultados foram utilizando a taxa de 70%. Enquanto que para as métricas *Theil* e ARV, os melhores resultados foram obtidos utilizando a taxa de 90%. Com isso, para os próximos experimentos utilizando a MLP como abordagem de combinação, serão utilizados os desempenhos para taxa de 70%, por apresentar menor custo computacional, pois irá realizar a combinação de apenas quatro modelos de previsão ao invés de oito.

Através dos *box-plots* apresentados na Figura 5.9 é possível analisar os desempenhos da abordagem de combinação utilizando SVR em relação as métricas. Em todas as métricas utilizadas: MAPE, POCID, *Theil* e ARV os melhores desempenhos foram utilizando a taxa de

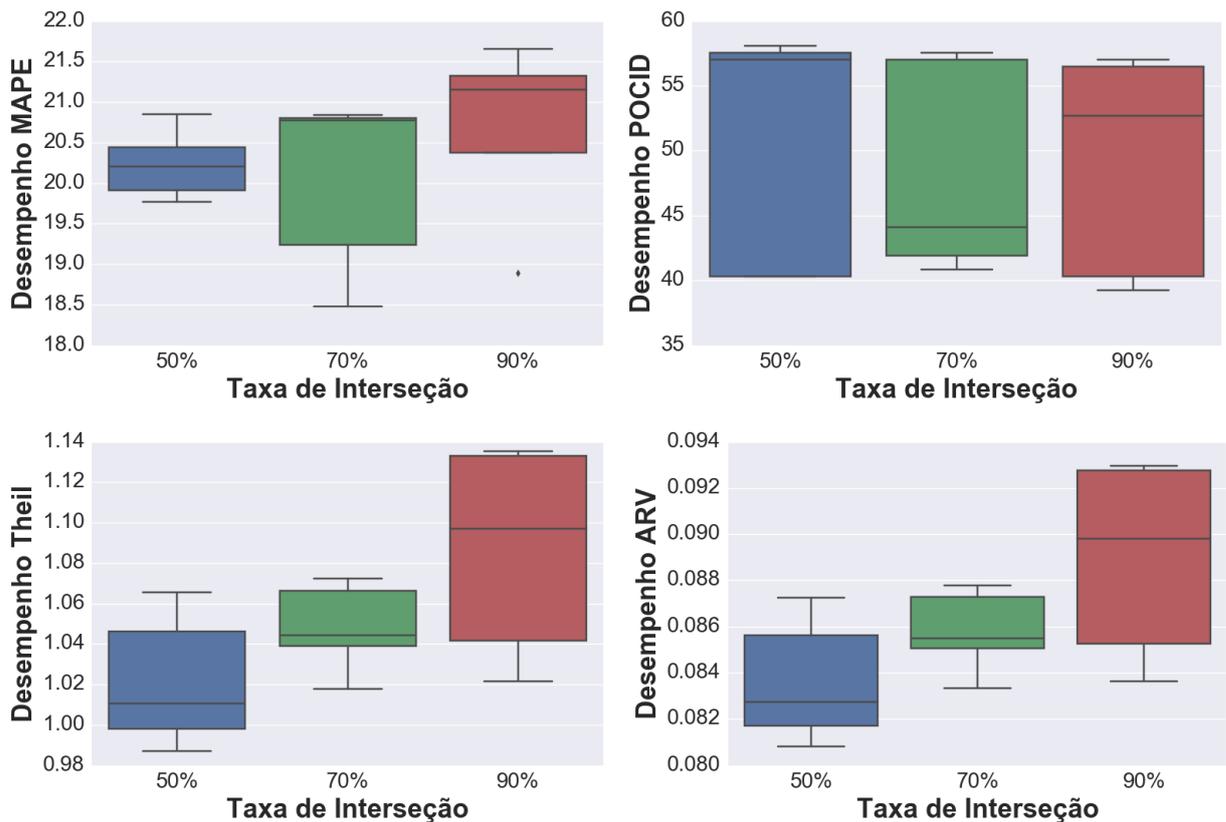


Figura 5.9: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Microsoft

50% de interseção. Dessa forma, os desempenhos da abordagem de combinação através da SVR com a taxa de 50% serão utilizados para comparações futuras.

Para a abordagem de combinação utilizando Regressão, como pode ser observado na Figura 5.10, os melhores desempenhos nas quatro métricas foram encontrados utilizando uma taxa de interseção de 50%. Com isso, para futuras comparações serão utilizados os desempenhos da abordagem de combinação por Regressão com taxa de 50%

Para a abordagem de combinação utilizando *Perceptron*, como pode ser observado nos *box-plots* da Figura 5.11, nas métricas MAPE e POCID obteve-se os melhores resultados utilizando a taxa de 70%. Enquanto que para as métricas *Theil* e ARV, os melhores resultados foram obtidos utilizando a taxa de 50%. Dessa forma, para comparações futuras serão utilizados os desempenhos da abordagem de combinação com a taxa de 50%, pois apresenta uma menor quantidade de modelos para serem combinados.

Em relação a abordagem de combinação utilizando a Média Simples, como mostra os *box-Plot* da Figura 5.12, a taxa de interseção de 70% apresentou melhores resultados nas quatro métricas. Dessa forma, para as próximas comparações utilizando a abordagem de combinação através da média, serão utilizados os desempenhos obtidos com a taxa de 70%.

Para a abordagem de combinação através da Mediana, como pode ser observado nos

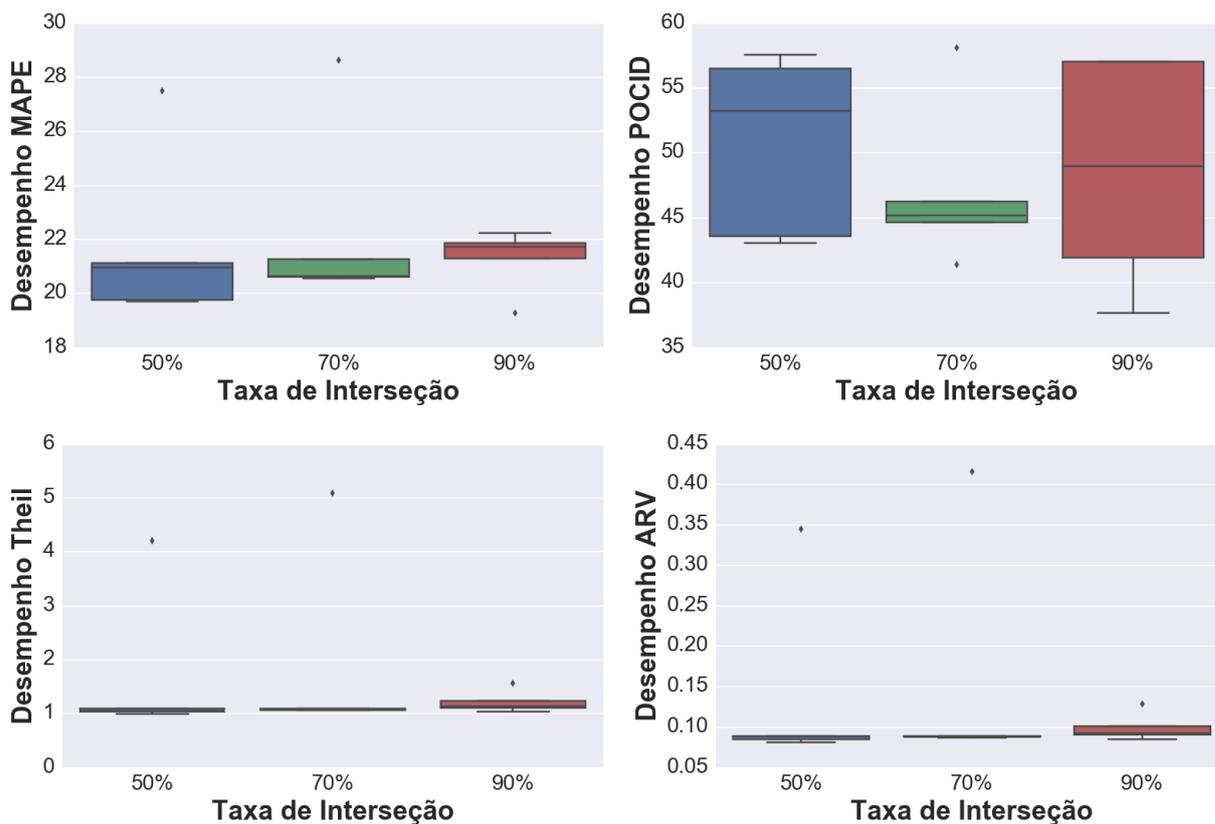


Figura 5.10: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Microsoft

Abordagem	Interseção (%)
MLP	70
SVR	50
RLM	50
Perceptron	50
Média	70
Mediana	90

Tabela 5.8: Melhores taxas para as abordagens de combinação Série Microsoft

box-plots da Figura 5.13, os melhores desempenhos foram obtidos através da taxa de 90%. Com isso, para futuras comparações com as abordagens de combinação com Mediana, serão utilizados os desempenhos com taxa de 90%.

A Tabela (5.8) mostra as taxas escolhidas para cada abordagem com base no desempenho médio obtido. Através dessa tabela, é possível observar que a maioria das abordagens tiveram seus melhores desempenhos utilizando as taxas 50% e 70% de interseção entre as partições de treinamento dos modelos de previsão. Esses resultados significam que as abordagens de combinação obtiveram desempenhos melhores quando a diversidade entre os dados de treinamento dos modelos de previsão foi maior, ou seja, compartilharam de uma quantidade menor de pontos da série em seu treinamento.

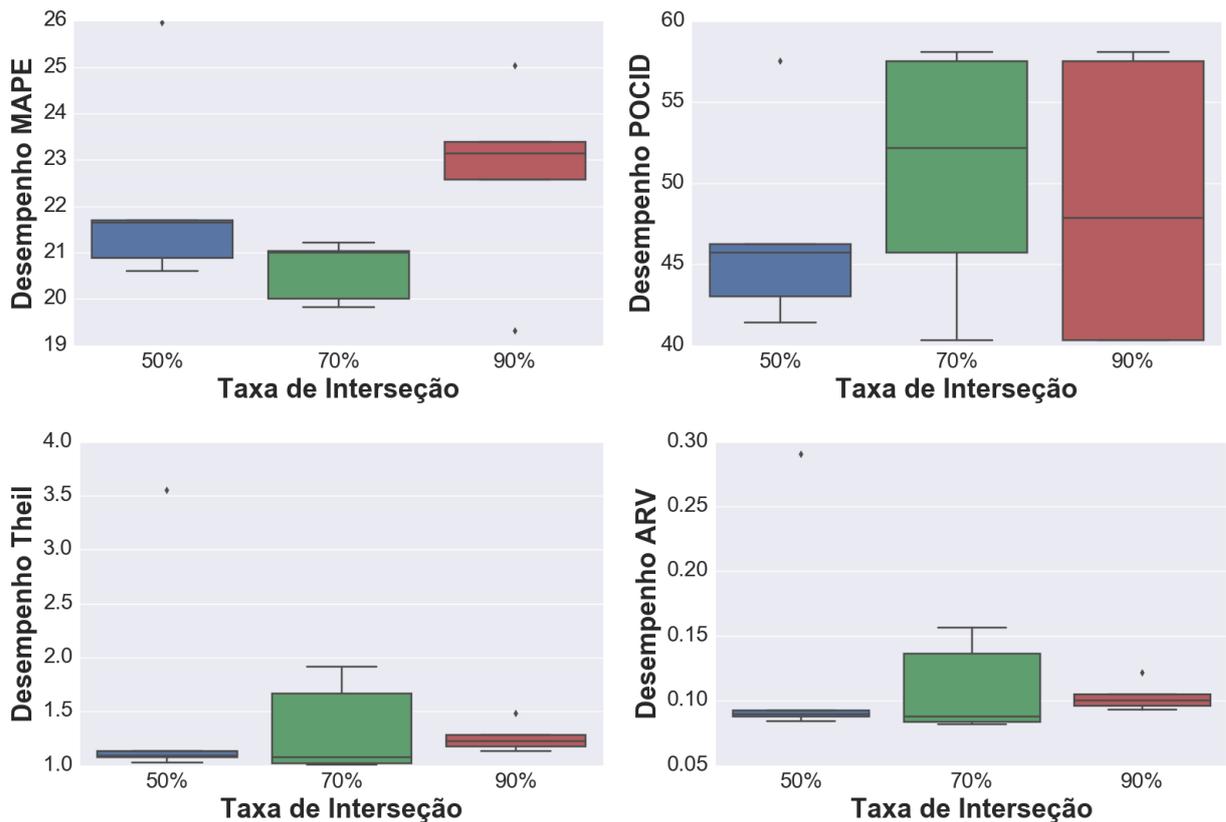


Figura 5.11: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com *Perceptron* na série Microsoft

5.2.2 Análise das Abordagens de Combinação

Após conhecer quais as melhores taxas de interseção, as abordagens de combinação têm os seus desempenhos comparados. Na Tabela (5.9) é possível observar os desempenhos médios de cada abordagem. Em relação ao MAPE e POCID as melhores abordagens foram através da Média e da MLP. Em relação ao *Theil* e ao ARV as melhores abordagens são SVR e Mediana. Através desses resultados é possível observar que nenhuma abordagem obteve o melhor desempenho em mais de duas métricas. No entanto, apenas as abordagens não lineares (MLP e SVR) e as abordagens lineares (Média Simples e Mediana) apresentaram os melhores resultados.

Para verificar estatisticamente se existem diferenças entre os resultados do *fitness* de cada abordagem, foram realizados os testes estatísticos. Primeiramente foi realizado o teste de normalidade, onde se verificou que apenas a amostra de desempenhos do *Perceptron* não apresentou normalidade, o restante das amostras são de distribuições normais. Com isso, para a análise comparativa em relação ao *Perceptron* foi utilizado o teste Wilcoxon, e para os demais foi utilizado o *t-Student*. A Hipótese nula H_0 assumiu que existe igualdade de desempenho entre as abordagens.

Através da Tabela (5.10) é possível observar que a abordagem de combinação através

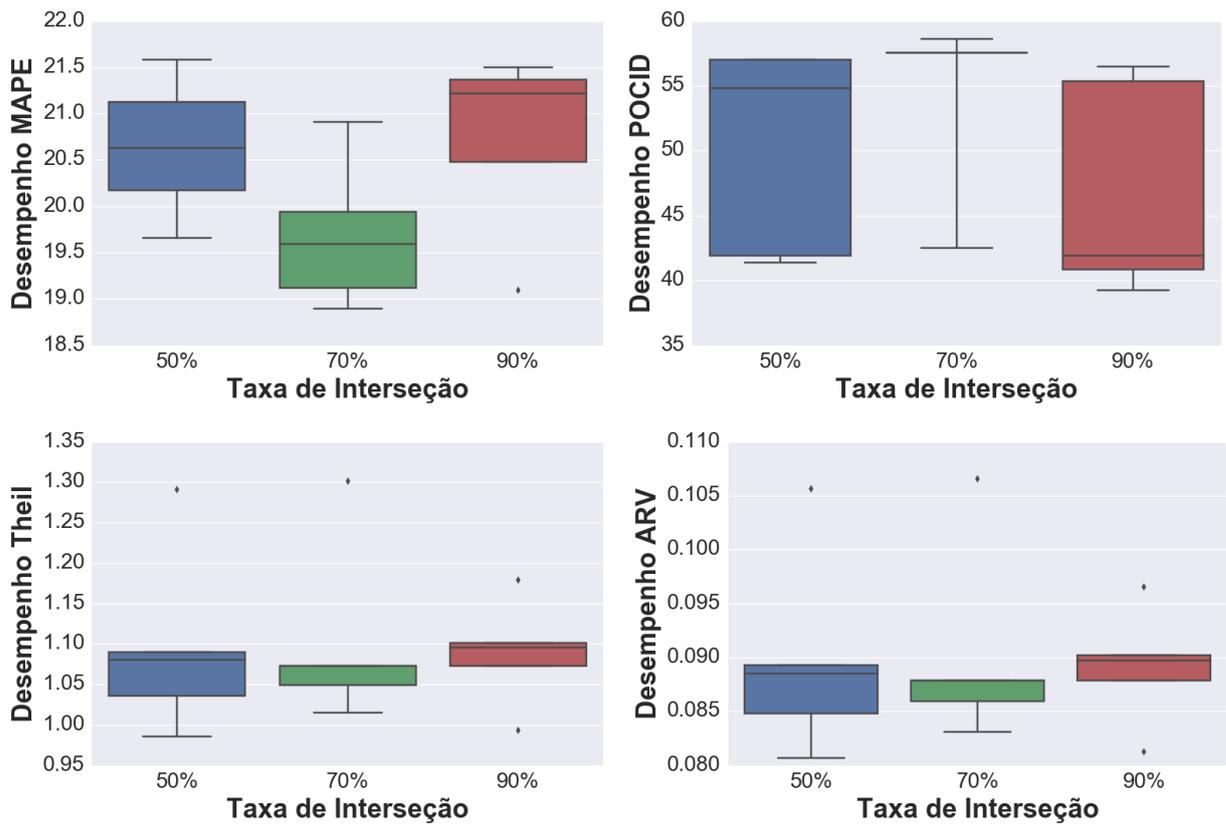


Figura 5.12: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Microsoft

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
MLP	19.97(0.72)	52.90(7.57)	1.069(0.112)	0.0875(0.0091)
SVR	20.23(0.43)	50.64(9.43)	1.021(0.033)	0.0836(0.0027)
RLM	21.79(3.24)	50.75(7.00)	1.676(1.411)	0.1372(0.1155)
Perceptron	22.15(2.17)	46.77(6.32)	1.572(1.103)	0.1287(0.0903)
Média	19.69(0.79)	54.73(6.86)	1.102(0.113)	0.0902(0.0093)
Mediana	20.71(0.63)	45.05(5.31)	1.059(0.035)	0.0867(0.0028)

Tabela 5.9: Comparação entre as diferentes abordagens de combinação para a série Microsoft

Abordagens	SVR	RLM	Perceptron	Média	Mediana
MLP	Não Rejeita H_0	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
SVR	–	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
RLM	–	–	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Perceptron	–	–	–	Não Rejeita H_0	Não Rejeita H_0
Média	–	–	–	–	Não Rejeita H_0

Tabela 5.10: Resultados dos Testes para o *fitness* na série Microsoft

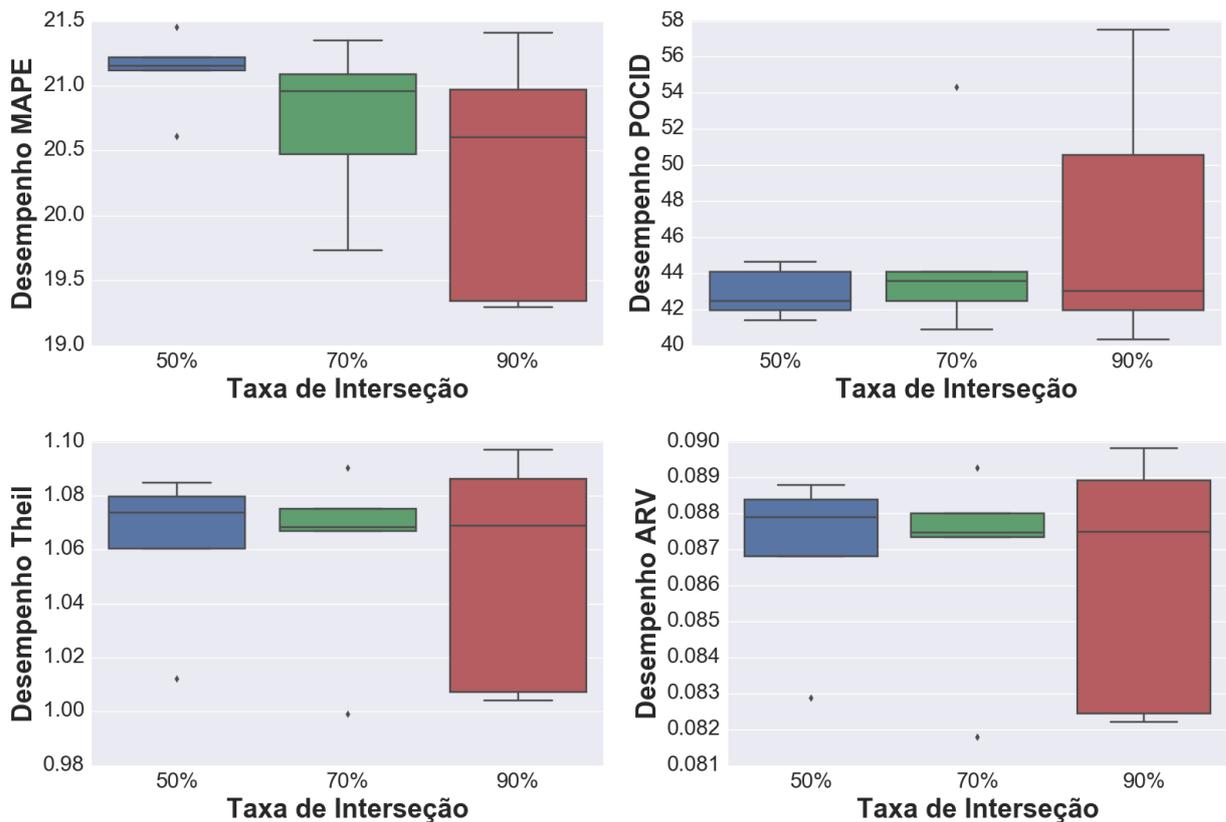


Figura 5.13: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Mediana na série Microsoft.

da MLP teve seu desempenho diferente apenas do *Perceptron*, a RLM apresentou resultados diferentes da Média Simples, SVR e Mediana apresentaram seus resultados estatisticamente semelhantes das demais abordagens.

5.2.3 Escolha da Arquitetura

A última análise para a série Microsoft consiste em verificar os desempenhos das abordagens de combinação e comparar com a abordagem de seleção. Assim como para a série anterior, os parâmetros da arquitetura de seleção foram os mesmos que apresentaram os melhores resultados no trabalho de (NOVA FILHO; PESSOA, 2015).

Através da Tabela (5.11) é possível observar os desempenhos das abordagens de combinação em relação a arquitetura de seleção. Para o MAPE, apenas as abordagens de combinação utilizando a Regressão e *Perceptron* não apresentaram melhores desempenhos que a abordagem de seleção. Em relação ao POCID, apenas as abordagens de combinação utilizando *Perceptron* e Mediana que não apresentaram desempenhos melhores que a abordagem de seleção. Em relação ao *Theil* e ao ARV, a abordagem de seleção foi melhor que as abordagens de combinação por MLP e por Mediana. Com base nesses resultados, é possível observar que para qualquer métrica de desempenho, a maioria das abordagens de combinação apresentam melhores resultados que a

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
Seleção	21.47(0.52)	48.81(5.16)	1.057(0.025)	0.0865(0.0020)
MLP	19.97(0.72)	52.90(7.57)	1.069(0.112)	0.0875(0.0091)
SVR	20.23(0.43)	50.64(9.43)	1.021(0.033)	0.0836(0.0027)
RLM	21.79(3.24)	50.75(7.00)	1.676(1.411)	0.1372(0.1155)
Perceptron	22.15(2.17)	46.77(6.32)	1.572(1.103)	0.1287(0.0903)
Média	19.69(0.79)	54.73(6.86)	1.102(0.113)	0.0902(0.0093)
Mediana	20.71(0.63)	45.05(5.31)	1.059(0.035)	0.0867(0.0028)

Tabela 5.11: Comparação entre as abordagens de combinação e a seleção para a série Microsoft

Rank	MAPE	POCID	Theil	ARV
1°	Comb. Média	Comb. Média	Comb. SVR	Comb. SVR
2°	Comb. MLP	Comb. MLP	Seleção	Seleção
3°	Comb. Mediana	Comb. Regressão	Comb. Mediana	Comb. Mediana
4°	Comb. SVR	Comb. SVR	Comb. MLP	Comb. MLP
5°	Seleção	Seleção	Comb. Média	Comb. Média
6°	Comb. Regressão	Comb. Perceptron	Comb. Perceptron	Comb. Perceptron
7°	Comb. Perceptron	Comb. Mediana	Comb. Regressão	Comb. Regressão

Tabela 5.12: Ranking dos melhores desempenhos médio para a série Microsoft

seleção, destacando a abordagem utilizando a Mediana que não apresentou melhor resultado que a seleção. A Tabela (5.12) mostra o *Ranking* das melhores abordagens resultante desta análise.

Para analisar se existem diferenças significativas nos desempenhos das abordagens, foram realizados os testes estatísticos. Primeiramente foi analisado se as amostras seguem uma distribuição normal, e como resultado as amostras são de distribuições normais. Em seguida foi realizado o teste *t-Student*, onde a hipótese nula h_0 é que não existem diferenças entre os resultados. Através da Tabela (5.13) é possível observar os resultados obtidos. Em relação ao MAPE, todas as abordagens de combinação obtiveram desempenhos diferentes da seleção. Em relação ao POCID e o *Fitness*, é possível observar que as abordagens não apresentaram diferenças significativas. Para as métricas *Theil* e *ARV* a abordagem de seleção teve seu desempenho diferente apenas em relação a abordagem de combinação utilizando SVR.

	Seleção x SVR	Seleção x MLP	Seleção x Média
MAPE	Rejeita H_0	Rejeita H_0	Rejeita H_0
POCID	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
Theil	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
ARV	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
Fitness	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0

Tabela 5.13: Resultados do Teste para a série Microsoft

A Tabela (5.14) mostra os melhores resultados encontrados nas abordagens implementa-

das nesse trabalho, através desses resultados é possível observar que a maioria das abordagens de combinação apresentaram resultados melhores que a seleção, com destaque para a abordagem utilizando SVR.

	MAPE	POCID	Theil	ARV
Seleção	20.45	56.45	1.0227	0.0837
Comb. MLP	18.83	58.60	0.9927	0.0812
Comb. SVR	20.85	58.06	0.9872	0.0808
Comb. RLM	19.69	57.52	0.9891	0.0809
Comb. Perceptron	20.59	57.52	1.0245	0.0838
Comb. Média	18.89	54.30	1.0150	0.0831
Comb. Mediana	19.29	57.52	1.0071	0.0821

Tabela 5.14: Comparação entre as diferentes abordagens para a série Microsoft

Contudo, com base nos resultados obtidos é possível concluir que para a série Microsoft, os melhores resultados foram encontrados utilizando as abordagens de combinação, com destaques para as combinações por Média, SVR e MLP, que apresentaram os melhores desempenhos. Porém, em relação aos testes de hipótese para o *Fitness*, os modelos não apresentam diferenças significativas em seus desempenhos.

Na Figura 5.14 é possível observar as previsões de teste da arquitetura de seleção e arquitetura de combinação utilizando a Média.

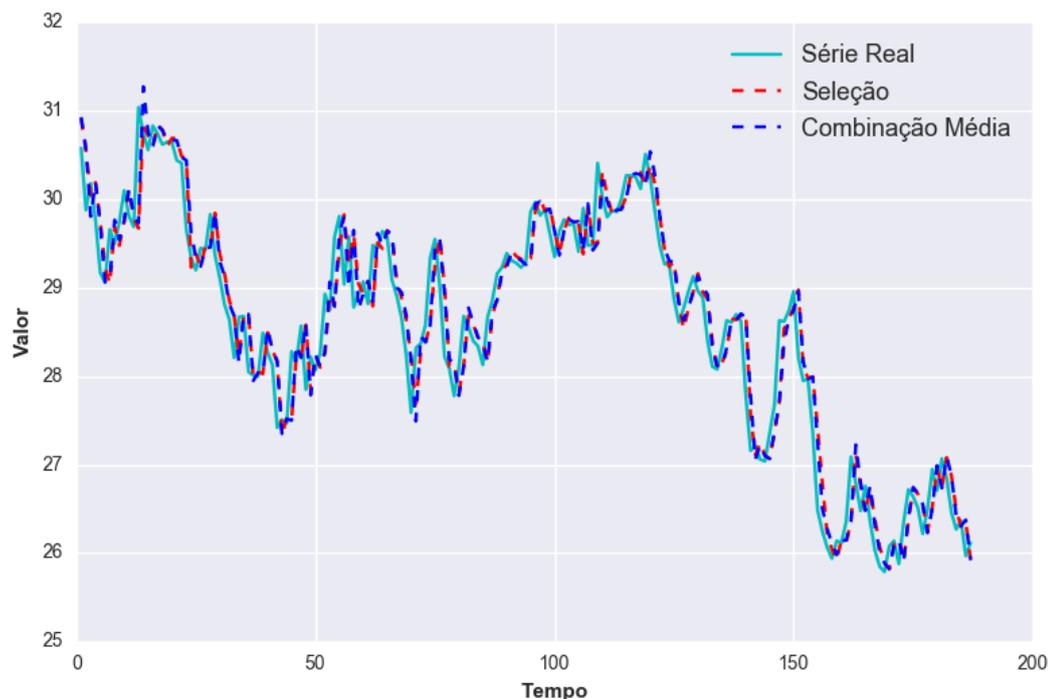


Figura 5.14: Previsões de Teste para a série Microsoft

5.3 Série Laser

A série é composta por 1000 pontos, onde os 750 primeiros pontos são utilizados para treinamento e validação dos modelos de previsões e os 250 últimos são utilizados para os testes.

5.3.1 Interseção entre as Partições

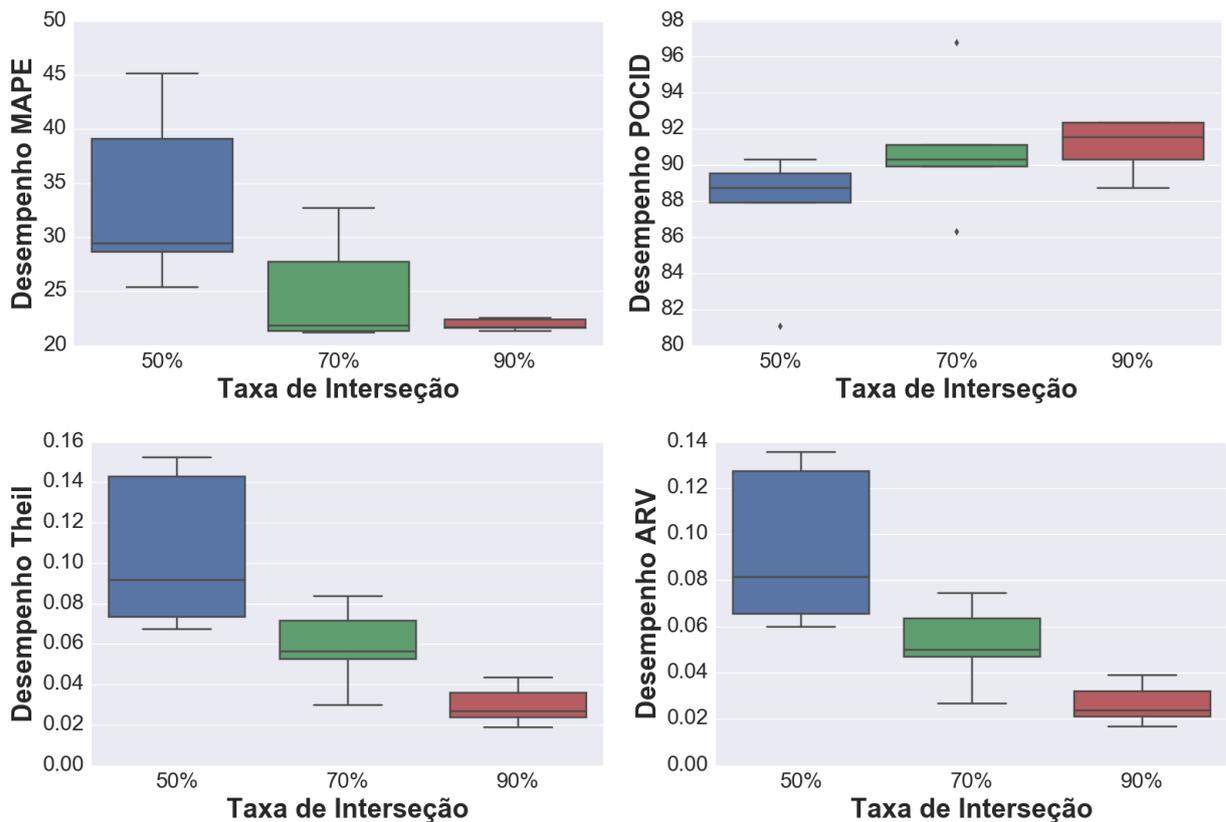


Figura 5.15: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Laser

Nesta etapa foi realizada uma análise comparativa entre as diferentes taxas de interseção para cada abordagem de combinação. Para tal, foi variado entre as taxas 50%, 70% e 90%, que correspondem a *Ensembles* compostos por 4, 5 e 13 modelos de previsão, onde cada modelo foi treinado por uma partição com 350 pontos e janelas com no máximo 20 *lags*.

Para a abordagem utilizando MLP, é possível observar na Figura 5.15 que em todas as métricas os melhores desempenhos foram obtidos utilizando a taxa de 90%. Com isso, para as próximas comparações utilizando a MLP como abordagem de combinação, será utilizado os desempenhos obtidos com a taxa de 90%.

Para a abordagem utilizando a SVR, é possível observar nos *box-plot* da Figura 5.16 que apenas para a métrica POCID que a taxa de 70% apresenta melhores resultados, para as demais

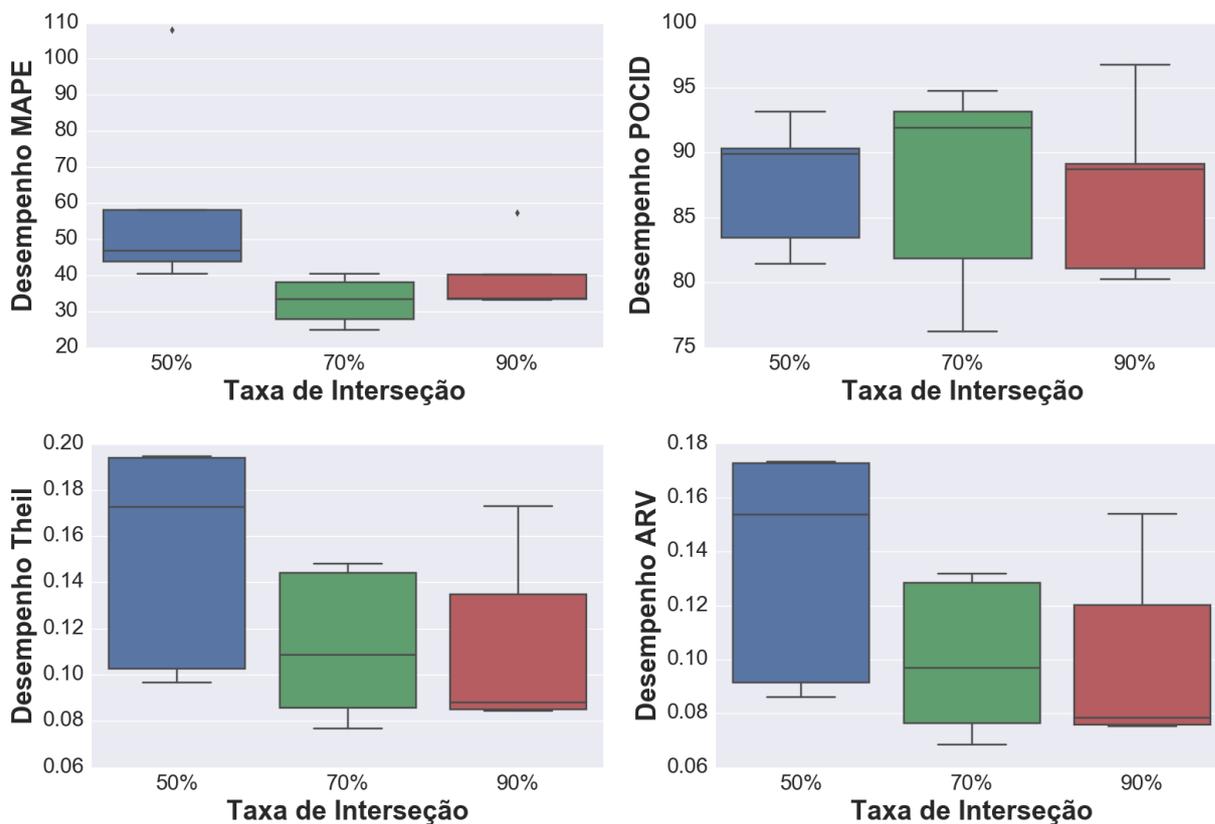


Figura 5.16: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Laser

métricas os melhores resultados foram obtidos através da taxa de 90%. Dessa forma, para as futuras comparações os desempenhos da abordagem de combinação através SVR com taxa de 90% serão utilizados.

Para a abordagem de combinação utilizando Regressão, é possível observar através da Figura 5.17 que em relação a métrica MAPE, os melhores desempenhos são utilizando a taxa de 70%. Para as demais métricas, os melhores desempenhos foram encontrados utilizando a taxa de 90%. Com isso, para as próximas comparações com a abordagem de combinação através da Regressão, serão utilizados os desempenhos relacionados a taxa de 90%.

Para a abordagem de combinação através do *Perceptron*, como pode ser observado na Figura 5.18 a taxa que apresentou os melhores desempenhos para o MAPE foi 70%, para as demais métricas foi a taxa de 50%. Dessa forma, em comparações futuras utilizando a abordagem de combinação através do *Perceptron* serão utilizados os desempenhos com a taxa de 50%.

Para a abordagem de combinação através da Média simples, é possível observar nos *box-plots* da Figura 5.19 que os melhores desempenhos foram obtidos utilizando as taxas 70% e 90%, sendo que, a taxa de 90% apresentou bons resultados apenas para a métrica POCID, enquanto que a taxa de 70% apresentou os melhores desempenho para as demais métricas. Com isso, em comparações futuras utilizando a abordagem de Média simples como combinação, serão

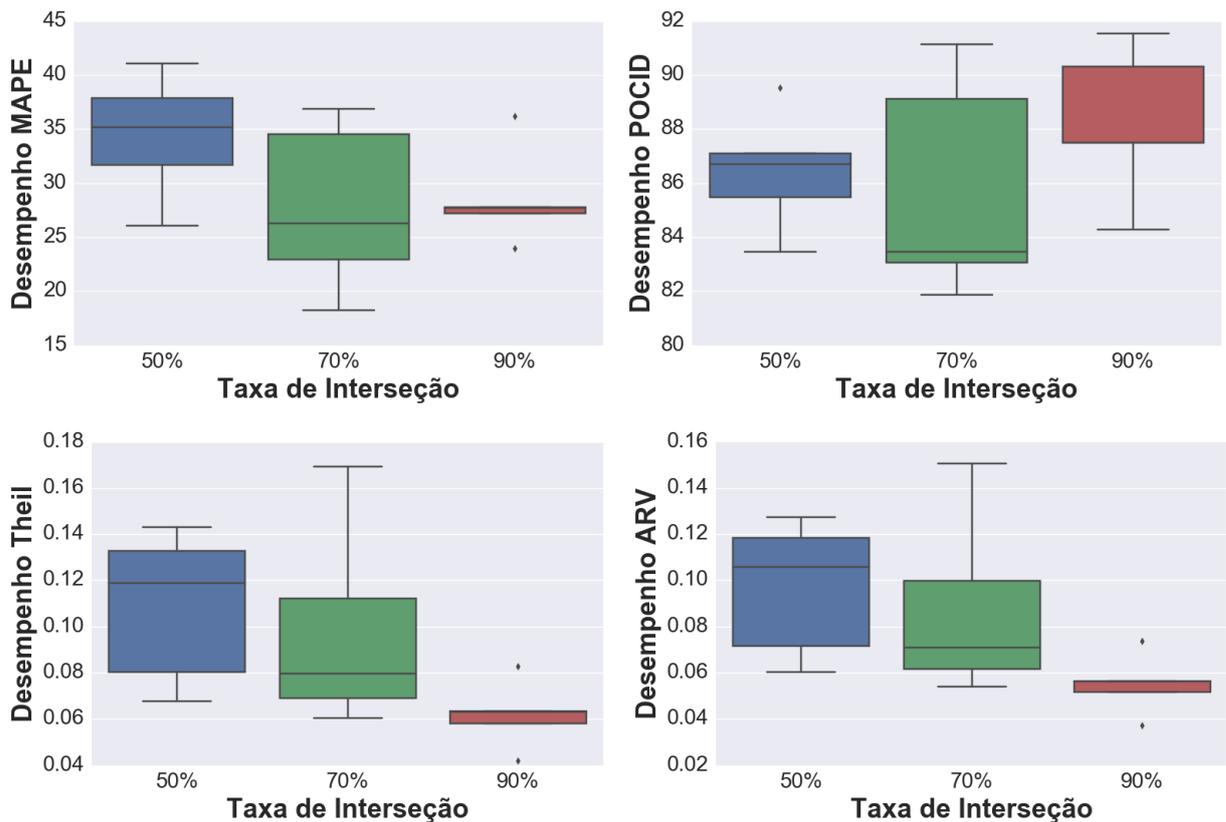


Figura 5.17: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Laser

utilizados os desempenhos relacionados a taxa de 70%.

Em relação a abordagem de combinação por Mediana, é possível analisar na Figura 5.20 que para a métrica POCID a melhor taxa foi 50%, enquanto que para as demais métricas a taxa de 70% apresentou os melhores desempenhos. Dessa forma, para as próximas comparações utilizando a abordagem de combinação por Mediana, serão utilizados os desempenhos referentes a taxa de 70%.

Abordagem	Interseção (%)
MLP	90
SVR	90
RLM	90
Perceptron	50
Média	70
Mediana	70

Tabela 5.15: Melhores taxas para as abordagens de combinação Laser

A Tabela (5.15) mostra as taxas escolhidas para cada abordagem com base no desempenho médio obtido. Através dessa tabela é possível observar as taxas que apresentaram os melhores desempenhos para as respectivas abordagens de combinação. Para as abordagens não lineares

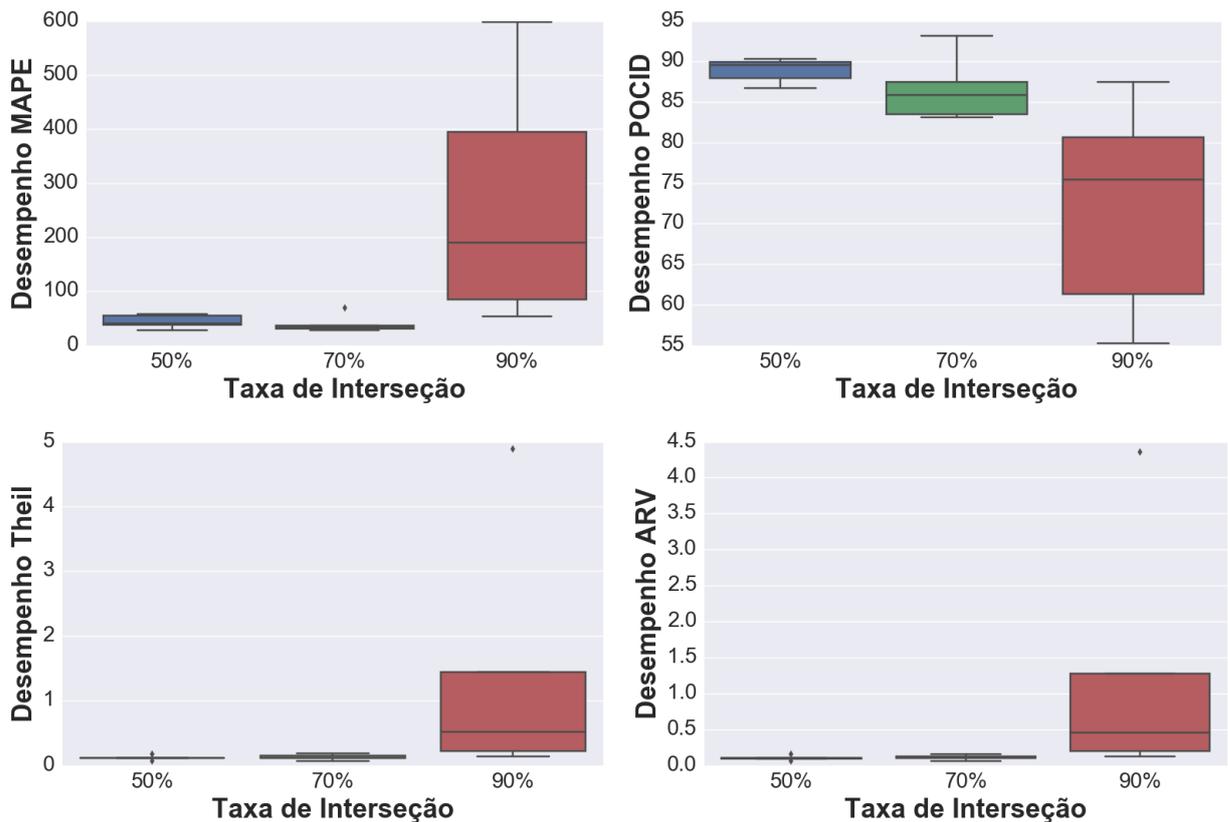


Figura 5.18: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com *Perceptron* na série Laser

(MLP e SVR) a melhor taxa foi de 90%, ou seja, apresentaram melhores resultados ao combinar um número maior de previsores e com menos diversidade. Para Regressão Linear e *Perceptron* as melhores taxas foram 90% e 50% respectivamente. E para a Média e Mediana a melhor taxa foi 70%.

5.3.2 Análise das Abordagens de Combinação

Após conhecer qual a melhor taxa de interseção para cada abordagem de combinação, é realizado uma comparação entre seus desempenhos. Através da Tabela (5.16) é possível realizar as comparações utilizando a média e o desvio padrão dos desempenhos. Em relação ao MAPE, as melhores abordagens de combinação foram utilizando a MLP, a Regressão e a Média Simples. Em relação ao POCID, as melhores abordagens foram utilizando a MLP, o *Perceptron*, e Regressão Linear. Em relação ao Theil e ao ARV, as melhores abordagens foram MLP, Regressão Linear e Mediana. Em geral é possível observar que a MLP foi a abordagem de combinação que apresentou os melhores resultados em todas as métricas.

Através da Tabela (5.17) é possível observar que a abordagem de combinação através da MLP apresentou um desempenho estatisticamente diferente das demais, ratificando os dados apresentado na Tabela (5.16). Contudo, para a série Laser a melhor abordagem de combinação

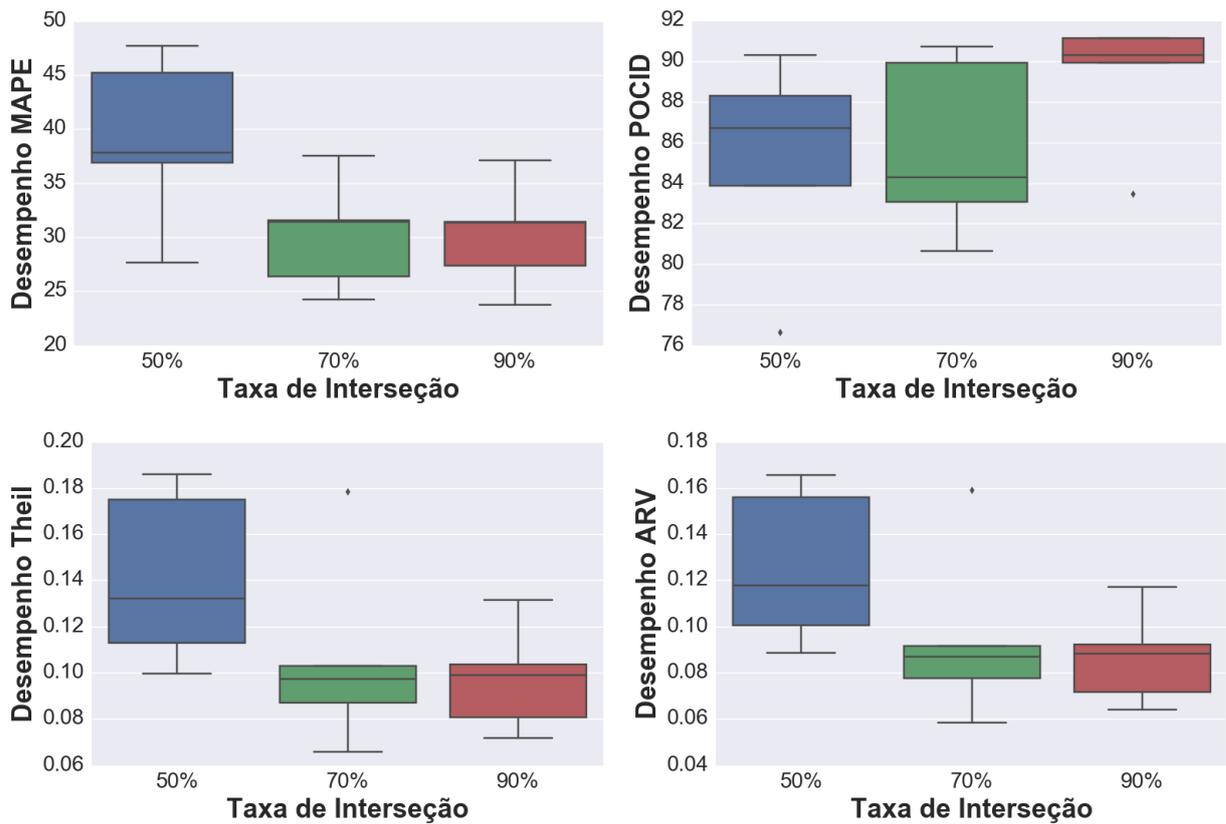


Figura 5.19: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Laser

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
MLP	21.90(0.55)	91.04(1.54)	0.0297(0.0099)	0.0265(0.0088)
SVR	39.49(10.3)	87.17(6.77)	0.1131(0.0396)	0.1007(0.0353)
RLM	28.55(4.53)	88.22(2.82)	0.0618(0.0145)	0.0550(0.0129)
Perceptron	43.71(12.2)	88.87(1.52)	0.1153(0.0388)	0.1026(0.0346)
Média	30.21(5.19)	85.72(4.40)	0.1063(0.0427)	0.0946(0.0380)
Mediana	30.67(6.70)	85.88(4.19)	0.1039(0.0291)	0.0925(0.0260)

Tabela 5.16: Comparação entre as diferentes abordagens de combinação para a série Laser

Abordagens	SVR	RLM	Perceptron	Média	Mediana
MLP	Rejeita H_0	Rejeita H_0	Rejeita H_0	Rejeita H_0	Rejeita H_0
SVR	–	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
RLM	–	–	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
Perceptron	–	–	–	Não Rejeita H_0	Não Rejeita H_0
Média	–	–	–	–	Não Rejeita H_0

Tabela 5.17: Resultados dos Testes para o *Fitness* na série Laser

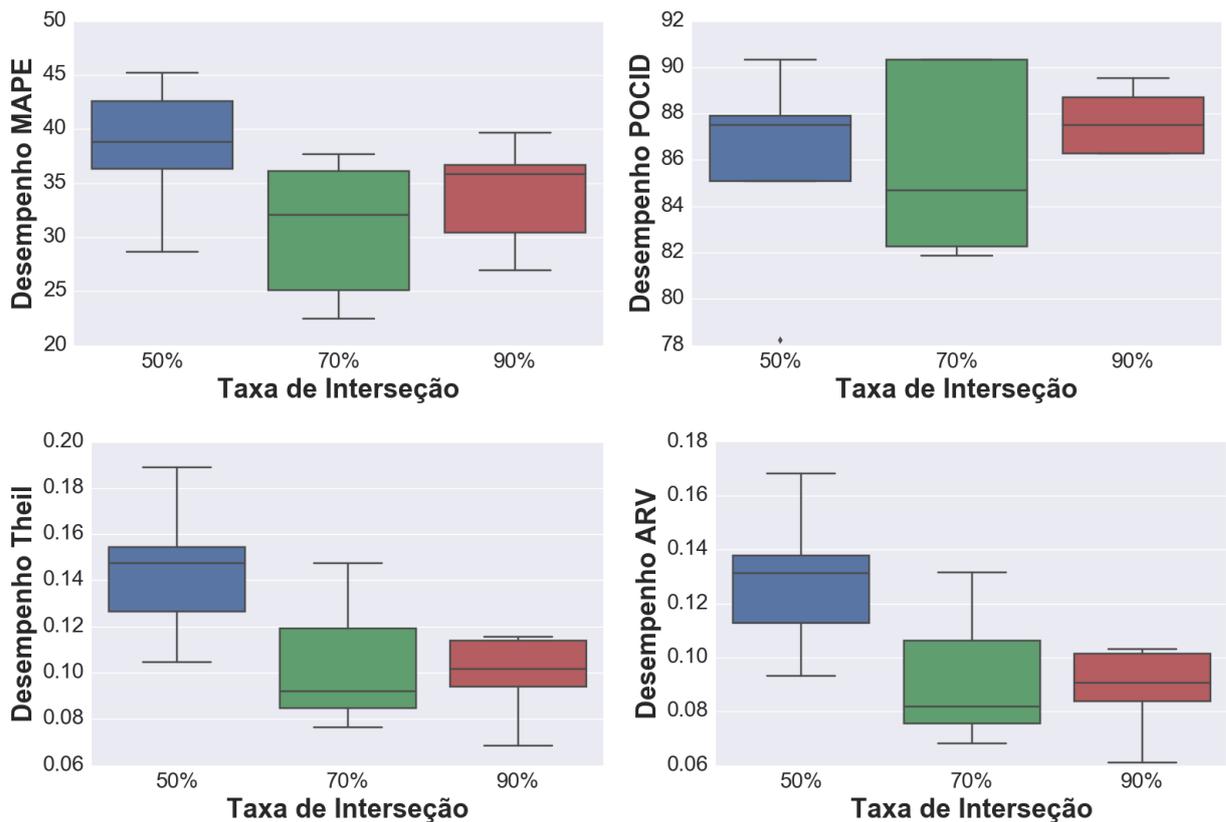


Figura 5.20: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Mediana na série Laser

foi através da MLP, combinando um *Ensemble* com 13 modelos de previsões.

5.3.3 Escolha da Arquitetura

A última etapa de análises comparativa para a série Laser, consiste em comparar os desempenhos das abordagens de combinação com a abordagem de seleção. Assim como para a série anterior, os parâmetros da arquitetura de seleção foram os mesmos que apresentaram os melhores resultados no trabalho de (NOVA FILHO; PESSOA, 2015).

Através da Tabela (5.18) é possível observar que o desempenho médio de todas as abordagens de combinação foi melhor que o desempenho da abordagem de seleção, em todas as métricas avaliadas. A Tabela (5.19) mostra o *Ranking* dos melhores modelos, resultante dessa análise.

Analisando através dos testes estatísticos, como mostra na Tabela (5.20), a hipótese nula h_0 de que as abordagens tem seus desempenhos iguais não foi rejeitada para as métricas POCID e *Fitness* em relação as abordagens de combinação por Regressão e por Média Simples. Com esses resultados, é possível inferir que para a série Laser as abordagens de combinação através de MLP, Regressão Linear ou Média Simples, apresentam melhores resultados que a abordagem de seleção. A Tabela (5.21) mostra os melhores resultados encontrados nas abordagens implementadas nesse

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
Seleção	56.06(19.3)	85.64(3.95)	0.2084(0.0479)	0.1856(0.0427)
MLP	21.90(0.55)	91.04(1.54)	0.0297(0.0099)	0.0265(0.0088)
SVR	39.49(10.3)	87.17(6.77)	0.1131(0.0396)	0.1007(0.0353)
RLM	28.55(4.53)	88.22(2.82)	0.0618(0.0145)	0.0550(0.0129)
Perceptron	43.71(12.2)	88.87(1.52)	0.1153(0.0388)	0.1026(0.0346)
Média	30.21(5.19)	85.72(4.40)	0.1063(0.0427)	0.0946(0.0380)
Mediana	30.67(6.70)	85.88(4.19)	0.1039(0.0291)	0.0925(0.0260)

Tabela 5.18: Comparação entre as abordagens de combinação e a seleção para a série Laser

Rank	MAPE	POCID	Theil	ARV
1°	Comb. MLP	Comb. MLP	Comb. MLP	Comb. MLP
2°	Comb. Regressão	Comb. Perceptron	Comb. Regressão	Comb. Regressão
3°	Comb. Média	Comb. Regressão	Comb. Mediana	Comb. Mediana
4°	Comb. Mediana	Comb. SVR	Comb. Média	Comb. Média
5°	Comb. SVR	Comb. Mediana	Comb. SVR	Comb. SVR
6°	Comb. Perceptron	Comb. Média	Comb. Perceptron	Comb. Perceptron
7°	Seleção	Seleção	Seleção	Seleção

Tabela 5.19: Ranking dos melhores desempenhos médio para a série Laser

	Seleção x MLP	Seleção x RLM	Seleção x Média
MAPE	Rejeita H_0	Rejeita H_0	Rejeita H_0
POCID	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
Theil	Rejeita H_0	Rejeita H_0	Rejeita H_0
ARV	Rejeita H_0	Rejeita H_0	Rejeita H_0
Fitness	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0

Tabela 5.20: Resultados do Teste para a série Laser

	MAPE	POCID	Theil	ARV
Seleção	28.31	91.93	0.0610	0.0546
Comb. MLP	21.28	92.33	0.0189	0.0168
Comb. SVR	33.21	96.77	0.0844	0.0752
Comb. RLM	23.89	90.32	0.0417	0.0371
Comb. Perceptron	28.55	90.32	0.0674	0.0600
Comb. Média	23.73	90.72	0.0656	0.0640
Comb. Mediana	22.42	90.32	0.0763	0.0680

Tabela 5.21: Comparação entre as diferentes abordagens para a série Laser

trabalho, onde é possível observar que a maioria das abordagens de combinação apresentaram resultados melhores que a seleção, com destaque para a abordagem utilizando MLP.

A Figura 5.21 mostra as séries de previsão utilizando a combinação por MLP e utilizando a abordagem de seleção.

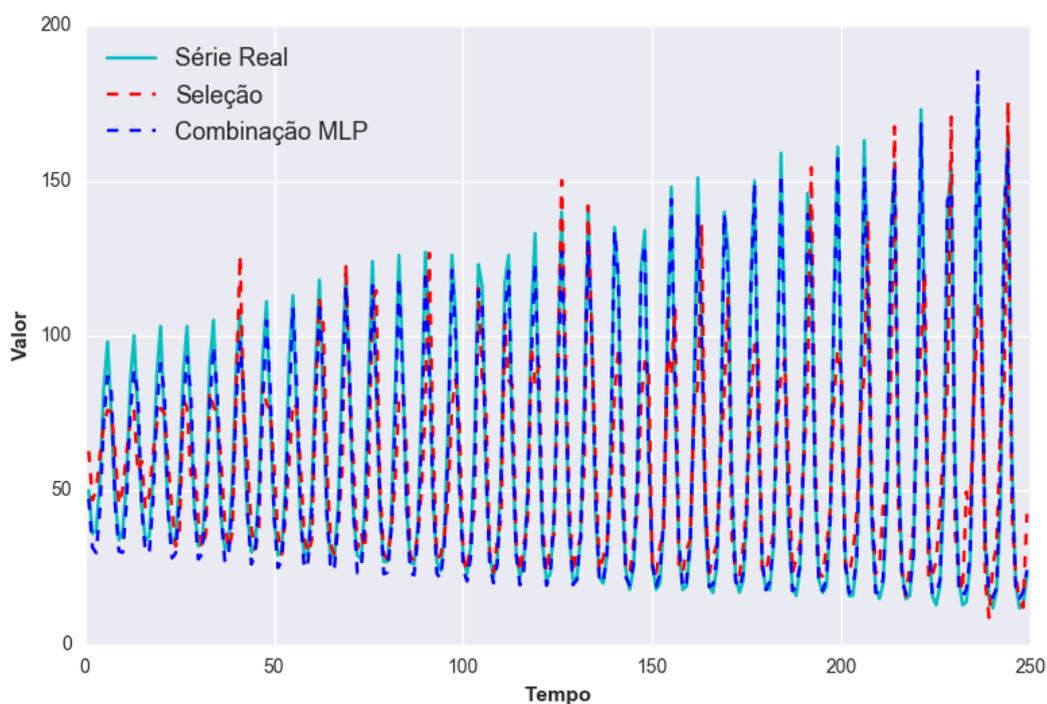


Figura 5.21: Previsões das Arquiteturas: Combinação x Seleção para a série Laser

5.4 Série Mackey Glass

A série é composta por 700 pontos, e foi dividida entre 525 para treinamento e validação dos modelos de previsões e 175 para testes.

5.4.1 Interseção entre as Partições

Nesta etapa foi realizada uma análise comparativa entre as diferentes taxas de interseção em relação a cada abordagem de combinação. As taxas utilizadas, 50%, 70% e 90%, correspondem a um *Ensemble* compostos por 2, 2, e 3 preditores. Cada modelo de previsão do *Ensemble* foi treinado utilizando uma partição com 450 pontos da série e cada janela teve um número máximo de 30 *lags*.

Para a abordagem de combinação utilizando MLP, é possível observar na Figura 5.22 que para a métrica MAPE os melhores desempenhos foram encontrados utilizando a taxa de 50%. Enquanto que para as métricas *Theil* e ARV, e POCID os melhores desempenhos foram obtidos com a taxa de 90%. Com isso, para as próximas comparações utilizando essa abordagem, serão utilizados os desempenhos relacionados a taxa de 50%.

Para a abordagem de combinação utilizando SVR, como pode ser observado na Figura 5.23, a taxa de 50% apresentou os melhores desempenhos para as métricas MAPE e POCID. E para as métricas *Theil* e ARV, os melhores resultados obtidos foram através da taxa de 90% de interseção. Dessa forma, para futuras comparações com essa abordagem serão utilizados os

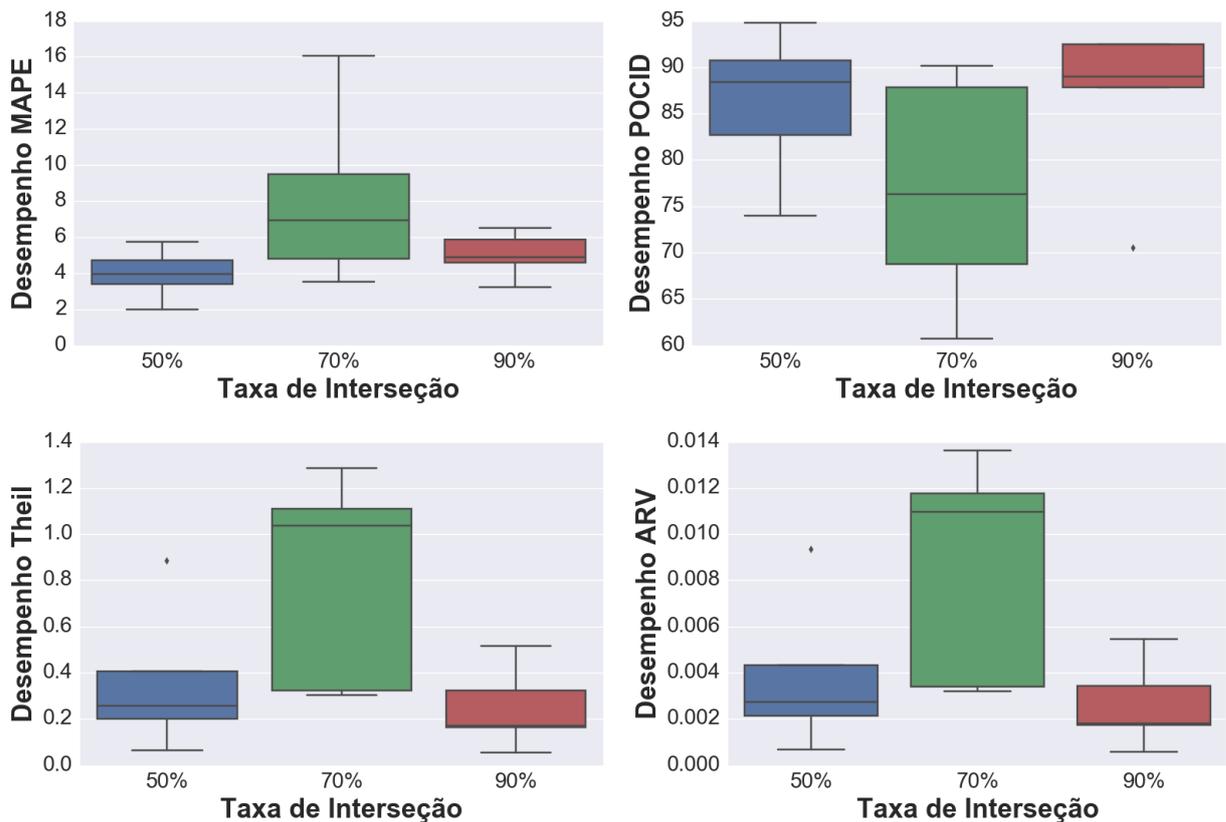


Figura 5.22: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com MLP na série Mackey Glass

resultados obtidos com a taxa de 50%.

Em relação a abordagem de combinação através da Regressão, é possível observar através dos *box-plots* da Figura 5.24 que em todas as métricas utilizadas (MAPE, POCID, *Theil* e ARV), a taxa que apresentou os melhores desempenhos foi a de 50% de interseção entre as partições. Com isso, para futuras comparações utilizando a abordagem de combinação através da Regressão, serão utilizados os desempenhos obtido com a taxa de 50%.

Para a abordagem de combinação utilizando o *Perceptron*, como pode ser observado na Figura 5.25, apenas para a métrica POCID em que a taxa com melhor desempenho foi 70%, para as demais métricas a taxa de 50% apresentou os melhores desempenhos. Com isso, para as próximas comparações utilizando a abordagem de combinação através do *Perceptron*, serão utilizados os desempenhos relacionados a taxa de 50%.

Para as abordagens de combinação utilizando a Média Simples ou a Mediana, como pode ser observados na Figura 5.26 e Figura 5.27, a taxa de 50% de interseção apresentou os melhores resultados em todas as métricas. Com isso, para as próximas comparações utilizando as abordagens de combinação através da Média Simples ou da Mediana, serão utilizados os desempenhos relacionados a taxa de 50%.

A Tabela (5.22) mostra as taxas escolhidas para cada abordagem com base no desempenho

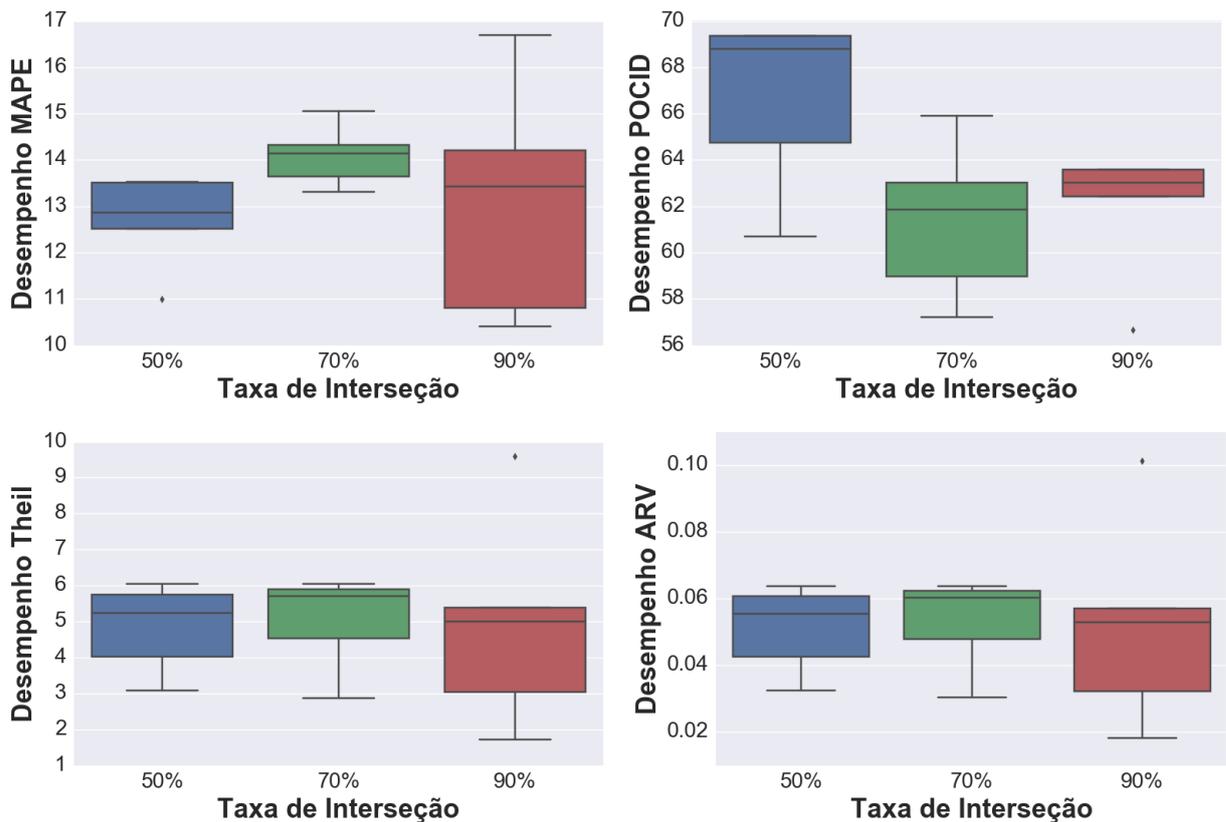


Figura 5.23: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com SVR na série Mackey Glass

Abordagem	Interseção (%)
MLP	50
SVR	50
RLM	50
Perceptron	50
Média	50
Mediana	50

Tabela 5.22: Melhores taxas para as abordagens de combinação Série Mackey-Glass

médio obtido. Através dela é possível perceber que em todas as abordagens a taxa de 50% de interseção entre as partições apresentou os melhores desempenhos.

5.4.2 Análise das Abordagens de Combinação

Após encontrar as melhores taxas de interseção entre as partições de treinamento dos modelos de previsão que serão combinados por cada abordagem de combinação, é realizada uma comparação entre essas abordagens a fim de encontrar as melhores formas de combinação.

Na Tabela (5.25) é possível observar a média e o desvio padrão dos desempenhos em relação as métricas. Em relação ao MAPE as melhores abordagens foram: a combinação

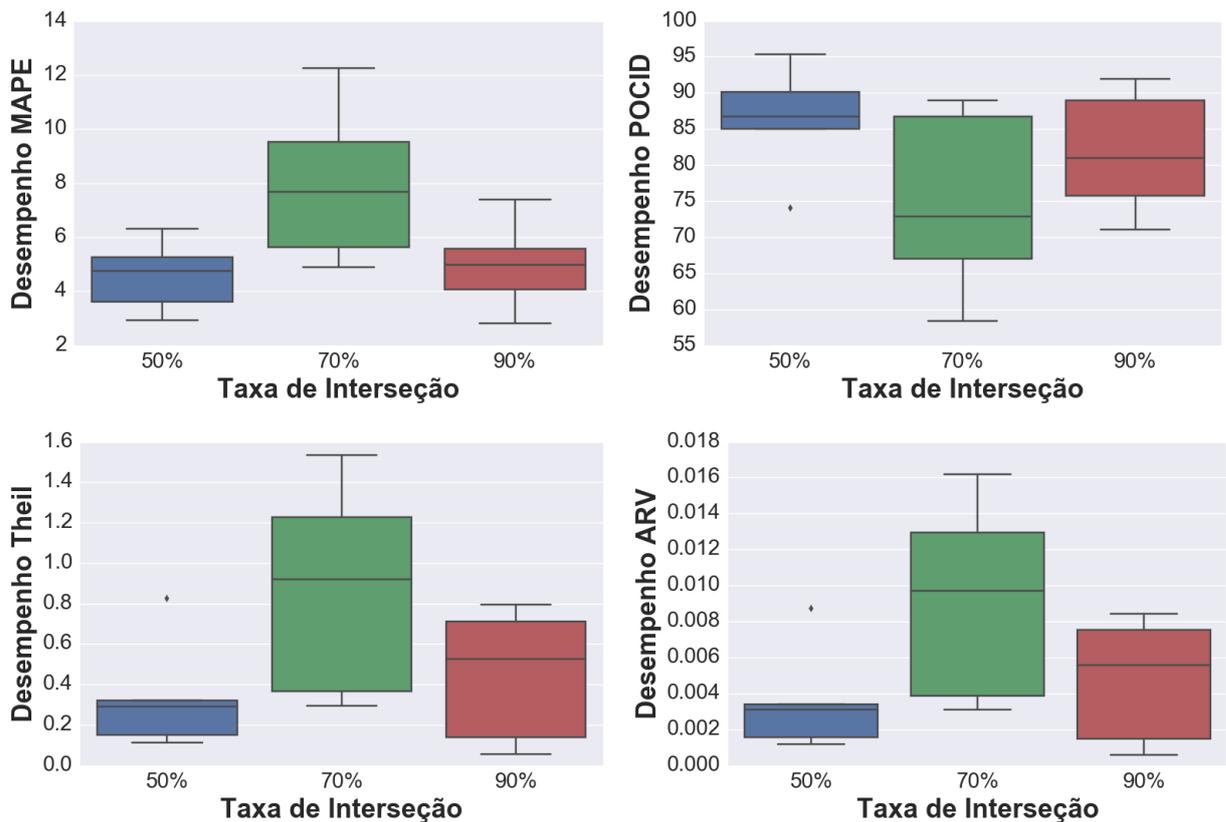


Figura 5.24: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Regressão na série Mackey Glass

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
MLP	3.95(1.41)	86.12(8.08)	0.363(0.315)	0.0038(0.0033)
SVR	12.67(1.03)	66.58(3.82)	4.823(1.246)	0.0509(0.0131)
RLM	4.56(1.33)	86.24(7.91)	0.339(0.285)	0.0035(0.0030)
Perceptron	26.46(7.19)	50.98(3.99)	15.73(2.82)	0.1663(0.0298)
Média	5.72(2.55)	80.57(7.76)	0.756(0.481)	0.0079(0.0050)
Mediana	5.72(2.55)	80.57(7.76)	0.756(0.481)	0.0079(0.0050)

Tabela 5.23: Comparação entre as diferentes abordagens de combinação para a série Mackey Glass

Abordagens	SVR	RLM	Perceptron	Média	Mediana
MLP	Rejeita H_0	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
SVR	–	Rejeita H_0	Rejeita H_0	Rejeita H_0	Rejeita H_0
RLM	–	–	Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
Perceptron	–	–	–	Rejeita H_0	Rejeita H_0
Média	–	–	–	–	Não Rejeita H_0

Tabela 5.24: Resultados dos Testes para o Fitness em relação a série Mackey Glass

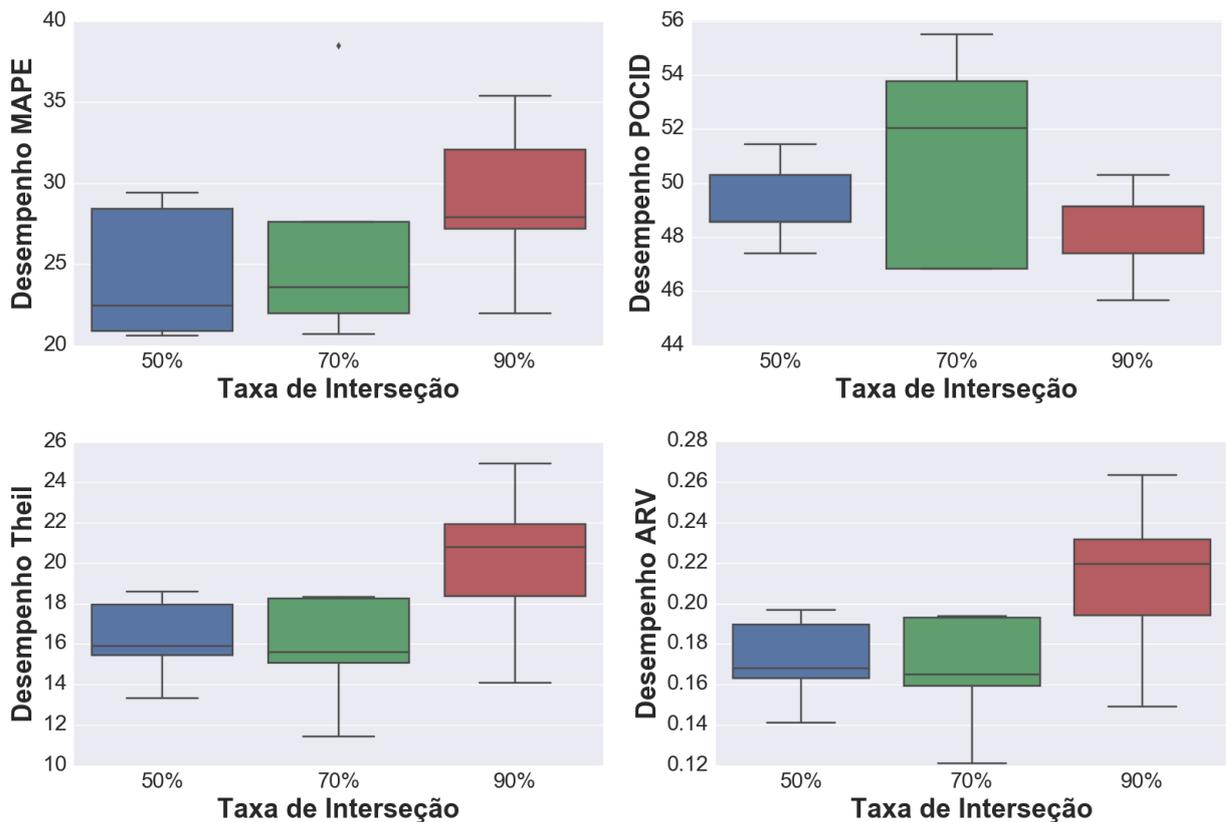


Figura 5.25: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com *Perceptron* na série Mackey Glass

através da MLP, da Regressão Linear, da Média Simples e da Mediana, destacando a MLP que obteve o melhor desempenho. Para as demais métricas (POCID, *Theil* e ARV) as melhores abordagens foram utilizando a Regressão Linear e MLP, com destaque para a Regressão Linear que apresentou os melhores desempenhos.

Na Tabela (5.24), é possível observar os resultados obtidos através do teste estatístico *t-Student*, onde a hipótese nula H_0 assume que não existem diferenças entre os desempenhos das abordagens para a métrica *Fitness*. Esses resultados mostram que a Regressão Linear que apresentou os melhores desempenhos médios, teve seu desempenho no *Fitness* estatisticamente semelhante aos desempenhos obtidos através das abordagens com Média Simples e com Mediana.

Contudo, para a série Mackey-Glass as melhores formas de combinação são através da Regressão e MLP.

5.4.3 Escolha da Arquitetura

A última análise realizada na série Mackey-Glass, consiste em comparar os desempenhos das abordagens de combinação com a abordagem de seleção. Assim como para as séries anteriores, os parâmetros da abordagem de seleção utilizada foram os mesmos que apresentaram os melhores resultados no trabalho de (NOVA FILHO; PESSOA, 2015).

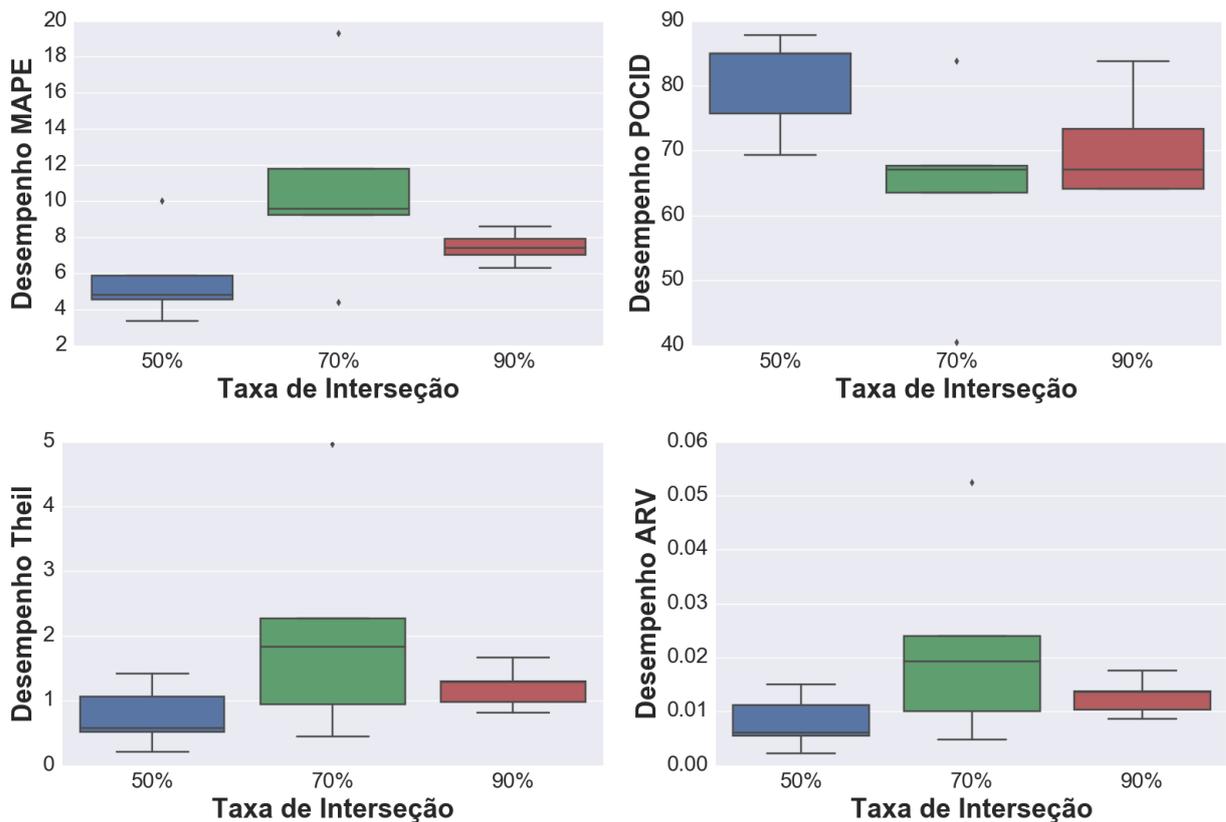


Figura 5.26: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação com Média Simples na série Mackey Glass

	MAPE	POCID	Theil	ARV
	Média(desv)	Média(desv)	Média(desv)	Média(desv)
Seleção	5.16(1.16)	79.76(8.34)	1.230(1.013)	0.0130(0.0107)
MLP	3.95(1.41)	86.12(8.08)	0.363(0.315)	0.0038(0.0033)
SVR	12.67(1.03)	66.58(3.82)	4.823(1.246)	0.0509(0.0131)
RLM	4.56(1.33)	86.24(7.91)	0.339(0.285)	0.0035(0.0030)
Perceptron	26.46(7.19)	50.98(3.99)	15.73(2.82)	0.1663(0.0298)
Média	5.72(2.55)	80.57(7.76)	0.756(0.481)	0.0079(0.0050)
Mediana	5.72(2.55)	80.57(7.76)	0.756(0.481)	0.0079(0.0050)

Tabela 5.25: Comparação entre as abordagens de combinação e seleção para a série Mackey Glass

Com base nos resultados apresentados na Tabela (5.25) é possível realizar a comparação dos desempenhos da abordagem de seleção com as abordagens de combinação. Em relação ao MAPE, a abordagem de seleção foi melhor que a maioria das abordagens de combinação, perdendo apenas para as abordagens com MLP ou Regressão. Em relação as demais métricas (POCID, *Theil* e ARV) a abordagem de seleção foi melhor apenas que as abordagens de combinação utilizando o SVR ou *Perceptron*. A Tabela (5.26) mostra o *Ranking* das melhores abordagens, resultante dessa análise.

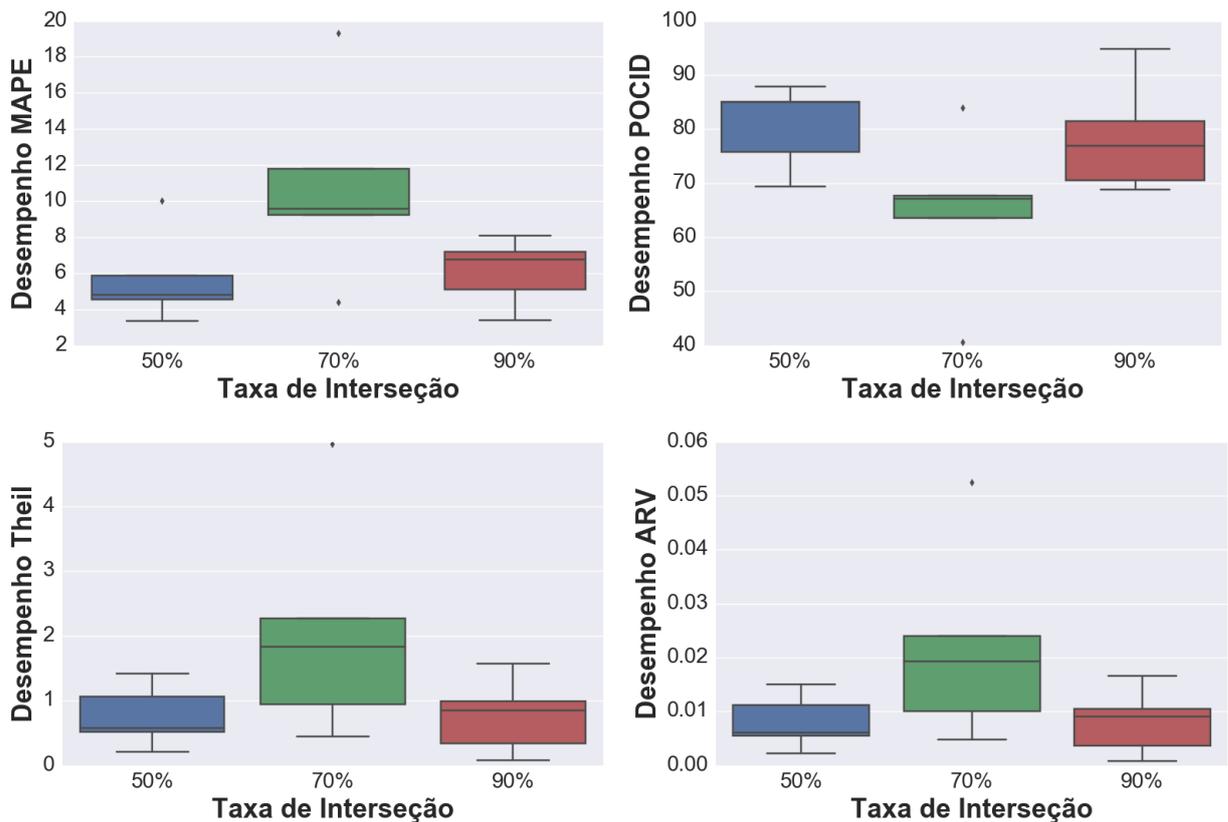


Figura 5.27: Os *Box-Plots* representam as distribuições dos desempenhos obtidos utilizando diferentes taxas de interseção para a abordagem de combinação por Mediana na série Mackey Glass

Rank	MAPE	POCID	Theil	ARV
1°	Comb. MLP	Comb. Regressão	Comb. Regressão	Comb. Regressão
2°	Comb. Regressão	Comb. MLP	Comb. MLP	Comb. MLP
3°	Seleção	Comb. Média	Comb. Mediana	Comb. Mediana
4°	Comb. Mediana	Comb. Mediana	Comb. Média	Comb. Média
5°	Comb. Média	Seleção	Seleção	Seleção
6°	Comb. SVR	Comb. SVR	Comb. SVR	Comb. SVR
7°	Comb. Perceptron	Comb. Perceptron	Comb. Perceptron	Comb. Perceptron

Tabela 5.26: Ranking dos melhores desempenhos médio para a série Mackey Glass

	Seleção x MLP	Seleção x RLM	Seleção x Média
MAPE	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
POCID	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0
Theil	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
ARV	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Fitness	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0

Tabela 5.27: Resultados do Teste para a série Mackey-Glass

Através da Tabela (5.27) é possível analisar as comparações entre as principais abordagens de combinação em relação com a abordagem de seleção, utilizando o teste *t-Student* para

verificar se a hipótese nula h_0 de que os desempenhos são iguais. Em relação com as abordagens de combinação por MLP, Média ou RLM, existe a possibilidade de que tenham estatisticamente os mesmos desempenhos.

A Tabela (5.28) mostra os melhores resultados encontrados nas abordagens implementadas nesse trabalho onde é possível observar que a abordagem de seleção apresentou resultados melhores que as abordagens de combinação por SVR e *Perceptron*. Em relação aos modelos de combinação, a abordagem por Regressão apresentou os melhores desempenhos, com exceção para a métrica MAPE.

	MAPE	POCID	Theil	ARV
Seleção	3.65	90.75	0.2103	0.0022
Comb. MLP	1.98	94.79	0.0655	6.9E-04
Comb. SVR	10.98	69.36	3.0711	0.0324
Comb. RLM	2.92	95.37	0.1109	0.0011
Comb. Perceptron	20.61	51.44	13.33	0.1409
Comb. Média	3.36	87.86	0.2049	0.0021
Comb. Mediana	3.36	87.86	0.2049	0.0021

Tabela 5.28: Comparação entre as diferentes abordagens para a série Mackey-Glass

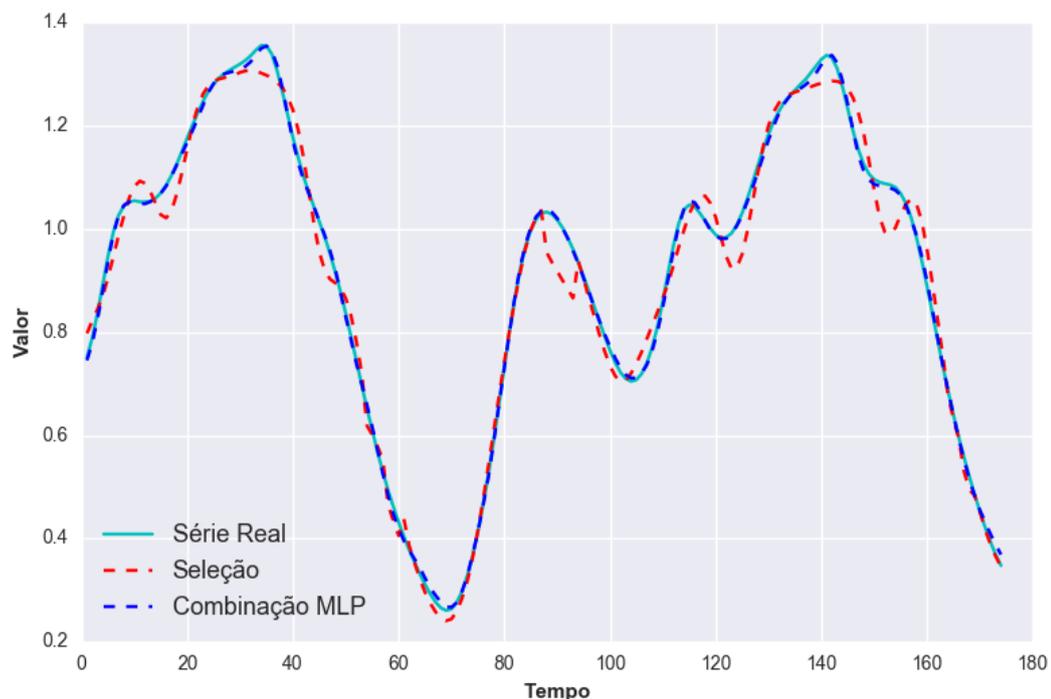


Figura 5.28: Previsões para a Série Mackey-Glass

Com base nessas análises, é possível inferir que para a série Mackey-Glass utilizar as principais abordagens de combinação, por Regressão ou MLP, pode apresentar resultados em média melhores que a utilização da abordagem de seleção. A Figura 5.28 mostra as previsões

para o conjunto de teste utilizando a abordagem de combinação por MLP e a abordagem de seleção.

5.5 Considerações Finais

Neste capítulo foram discutidos os resultados obtidos através das diferentes abordagens utilizadas. E foram explanados de acordo com cada série temporal utilizada. Para a série Goldman Sachs, a abordagem de combinação com MLP apresentou os melhores resultados. Para a série Microsoft as abordagens de combinação com SVR e Média, apresentaram os melhores resultados. Para a série Laser, a abordagem de combinação com MLP apresentou os melhores resultados. E para a série Mackey Glass os melhores resultados foram obtidos através da abordagem de combinação com Regressão Linear e MLP. Em geral, a maioria das abordagens de combinação utilizadas apresentaram os melhores desempenhos em relação a abordagem de seleção utilizada.

6

CONCLUSÃO E TRABALHOS FUTUROS

Neste capítulo estão as respostas para as questões de pesquisa que guiaram esta dissertação. Assim como, o resumo do trabalho realizado, as contribuições, limitações e os trabalhos futuros.

6.1 Resumo

Nesta dissertação o problema de previsão de séries temporais foi abordado através do contexto de combinação e seleção de preditores. A combinação de preditores consiste em juntar as previsões obtidas de mais de um preditor através de alguma técnica. No presente trabalho foram utilizadas técnicas já existentes na literatura, como a Regressão Linear, o MLP, a Média Simples e a Mediana, SVR e o *Perceptron*. A seleção de preditores consiste em selecionar, dentre um conjunto de preditores, apenas um modelo para realizar a previsão final. No presente trabalho foi utilizado a arquitetura de seleção proposta por (NOVA FILHO; PESSOA, 2015), que consiste em selecionar o modelo com base na distância entre a série de treinamento de cada modelo e a janela que antecede o ponto a ser previsto, utilizando a métrica de distância DTW.

A utilização da combinação de preditores é motivada como alternativa para juntar diferentes modelos especialistas, e com isso obter uma previsão final melhor que a previsão de seus indivíduos. Já a seleção de preditores é motivada por reduzir o custo computacional da utilização de mais de um preditor e utilizar um modelo específico para cada comportamento da série temporal, podendo apresentar uma previsão com maior precisão.

Para verificar a viabilidade entre a arquitetura de seleção e a arquitetura de combinação, na presente dissertação foram realizadas diversas análises comparativas, a fim de encontrar em quais casos cada arquitetura apresenta um melhor desempenho. Para tal, foram utilizadas quatro séries: Mackey Glass, Laser, Microsoft e Goldman Sachs. Para cada série os desempenhos das arquiteturas foram avaliados através de diversas métricas junto com testes estatísticos para verificar se existem diferenças estatisticamente significantes entre os desempenhos.

Além da análise comparativa entre combinação e seleção, foram realizadas comparações entre as diferentes abordagens de combinação. Primeiramente, foram realizados os experimentos

para verificar qual a taxa de interseção entre as partições que apresentam melhores resultados para cada abordagem de combinação, variando entre as taxas de 50%, 70% e 90%. Com base no resultado obtido entre as diferentes taxas, foi realizada a comparação entre os diferentes tipos de abordagens de combinação utilizadas: lineares (*Perceptron*, Regressão Linear, Média Simples e Mediana) e não Linear (MLP e SVR) .

Os resultados obtidos nos experimentos realizados mostraram que nas séries utilizadas, a arquitetura de combinação apresenta melhores resultados que a arquitetura de seleção. Porém, nem todas as abordagens de combinação apresentam desempenhos superiores a abordagem de seleção. A abordagem não linear (MLP) e as abordagens lineares (Média Simples, Mediana e Regressão) se mostraram com melhores desempenhos em relação ao conjunto de métricas utilizadas. Para explicar quais fatores influenciaram no desempenho dessas abordagens é necessário a realização de experimentos mais detalhados, porém, a priori é possível inferir que um dos motivos se dar ao comportamento da relação entre os modelos de previsão serem lineares e por não apresentarem previsões aberrantes, facilitando para as abordagens de combinação lineares.

Contudo, os resultados obtidos para as métricas e séries utilizadas nesta dissertação mostram que a arquitetura de combinação apresentam em média, resultados melhores que a arquitetura de seleção, destacando as abordagens de combinação através da MLP, Média Simples, Mediana e Regressão Linear. Tornando assim, essas abordagens de combinação uma alternativa viável para futuras pesquisas científicas ou aplicações práticas.

6.2 Questões de Pesquisa

Nesta seção as questões que guiaram o desenvolvimento dessa pesquisa serão respondidas com base nos resultados apresentados no capítulo anterior, levando em consideração o conjunto de métricas utilizadas neste trabalho.

6.2.1 Quando a combinação é melhor que a seleção?

Com base nos resultados obtidos, é possível observar que a relação entre as abordagens de combinação e a abordagem de seleção apresentaram comportamentos diferentes de acordo com cada série.

Para a série Goldman Sachs e Laser, a abordagem mais adequada é a combinação de preditores através do modelo não linear MLP. Para a série Microsoft, a abordagem de combinação por SVR apresentou os melhores resultados, tornando-se a abordagem mais adequada. Para a série Mackey Glass, a Regressão Linear apresentou em média os melhores desempenhos.

De forma geral, para as séries utilizadas é possível concluir que a maioria das abordagens de combinações utilizadas apresentam melhores desempenhos que a abordagem de seleção, principalmente em séries temporais com comportamentos caóticos.

6.2.2 Qual a relação entre a diversidade nos dados e as abordagens de combinação?

Com base nesses resultados obtidos, é possível inferir que as menores taxas de interseção apresentaram na maioria dos casos, os melhores resultados. Ou seja, o desenvolvimento de partições que compartilham uma menor quantidade de dados, resulta na composição de modelos de previsão mais diversos e conseqüentemente em previsões combinadas mais acuradas. Além disso, não existe uma relação entre o tipo de abordagem de combinação, seja linear ou não linear, e a taxa de interseção, assim como não existe uma relação entre o tipo de série, caótica ou financeira.

6.2.3 Quais as melhores abordagens de combinação?

Com base nos resultados obtidos, é possível observar que não existe uma abordagem, seja ela não linear ou linear, que apresente os melhores resultados em todas as métricas e em qualquer série temporal. No entanto, algumas abordagens de combinação apresentam melhores desempenhos que outras. Em relação ao MAPE, *Theil* e ARV as abordagens não lineares (MLP e SVR) e as lineares (Média Simples e Mediana) tiveram os melhores desempenhos. Em relação ao POCID, a abordagem não linear MLP e as abordagens lineares: Regressão Linear, Média e Perceptron apresentaram os melhores resultados.

6.3 Contribuições

Esta dissertação apresenta algumas contribuições para o campo de pesquisa na área de previsão de séries temporais e aprendizado de máquina.

Os experimentos realizados em busca da taxa de interseção entre as abordagens de combinação mostram o comportamento de cada abordagem de combinação em relação a diversidade nos dados. Através desses comportamentos é possível observar que as taxas são diferentes para cada modelo utilizado. Em seguida, as comparações entre as diversas abordagens de combinações mostram quais são as melhores para determinada série temporal. No presente trabalho foram apresentadas os desempenhos de diferentes abordagens que exigem um treinamento como MLP, Regressão Linear, SVR e *Perceptron* e abordagens que não necessitam de um treinamento como: Média Simples e Mediana. E por fim, as análises entre as abordagens de combinação e a arquitetura de seleção proposta por (NOVA FILHO; PESSOA, 2015), mostram que na maioria dos casos, combinar é melhor que selecionar. Principalmente, quando se a abordagem de combinação é através de uma MLP.

6.4 Limitações

A pesquisa realizada nesta dissertação tem como limitação a quantidade de séries temporais utilizadas, onde, por questão de tempo de execução e de análise dos resultados foram utilizadas apenas quatro séries temporais, e com isso, o comportamento obtido nessas séries podem não refletir em outras séries.

Em relação a análise comparativa entre as abordagens de combinação e a seleção, a limitação dá pelo o fato de que o presente trabalho não teve o objetivo de analisar o custo computacional e de tempo de execução para cada arquitetura.

6.5 Trabalhos Futuros

Após a realização desta dissertação, novos trabalhos podem dar continuidade em busca de novas abordagens de combinação que apresentem melhores resultados que as abordagens utilizadas nesse trabalho. Ou através da comparação de outras abordagens de combinação como a Média Bayesiana, Lasso, Filtro de Kalman (NÓBREGA; OLIVEIRA, 2013), ou cópulas (OLIVEIRA et al., 2017).

Além disso, por conta da quantidade de parâmetros utilizados tanto na arquitetura de seleção como de combinação, é possível realizar uma otimização através de algoritmos de inteligência coletiva, com o objetivo de ajustar esses parâmetros e encontrar melhores resultados, ou que possam encontrar dentro do conjunto de preditores, quais os modelos mais significantes para serem combinados.

Outra proposta, é o desenvolvimento de uma arquitetura de combinação de previsores, com modelos de previsão mais simples que a RNA utilizada, de forma que possa avaliar a capacidade das abordagens de combinação em melhorar a previsão final para previsores simples.

E por fim, o desenvolvimento de novas abordagens para a criação de conjuntos de preditores com mais diversidade, ou comparação com outras técnicas como *Stacked* (WOLPERT, 1992) e *Blending* (YU et al., 2015). Além da possibilidade da utilização de técnicas para encontrar diferentes padrões na série, por exemplo técnicas de *Concept Drift* (WIDMER; KUBAT, 1996).

REFERÊNCIAS

- ADHIKARI, R. A neural network based linear ensemble framework for time series forecasting. **Neurocomputing**, [S.l.], v.157, p.231–242, 2015.
- ADHIKARI, R.; AGRAWAL, R. Combining multiple time series models through a robust weighted mechanism. In: RECENT ADVANCES IN INFORMATION TECHNOLOGY (RAIT), 2012 1ST INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2012. p.455–460.
- ADHIKARI, R.; VERMA, G.; KHANDELWAL, I. A Model Ranking Based Selective Ensemble Approach for Time Series Forecasting. **Procedia Computer Science**, [S.l.], v.48, p.14–21, 2015.
- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: **Selected Papers of Hirotugu Akaike**. [S.l.]: Springer, 1998. p.199–213.
- AKSU, C.; GUNTER, S. I. An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. **International Journal of Forecasting**, [S.l.], v.8, n.1, p.27–43, 1992.
- ALAM, J.; LJUNGEHED, J. **A comparative study of hybrid artificial neural network models for one-day stock price prediction**. 2015.
- ALWEE, R.; HJ SHAMSUDDIN, S. M.; SALLEHUDDIN, R. Hybrid support vector regression and autoregressive integrated moving average models improved by particle swarm optimization for property crime rates forecasting with economic indicators. **The Scientific World Journal**, [S.l.], v.2013, 2013.
- BATES, J. M.; GRANGER, C. W. The combination of forecasts. **Journal of the Operational Research Society**, [S.l.], v.20, n.4, p.451–468, 1969.
- BERGMAN, S. **The kernel function and conformal mapping**. [S.l.]: American Mathematical Soc., 1970. n.5.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of econometrics**, [S.l.], v.31, n.3, p.307–327, 1986.
- BOX, G. E. et al. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BOX, G. E.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American statistical Association**, [S.l.], v.65, n.332, p.1509–1526, 1970.
- BROCKWELL, P. J.; DAVIS, R. A. **Time series: theory and methods**. [S.l.]: Springer Science & Business Media, 2013.
- BURGES, C. J. A tutorial on support vector machines for pattern recognition. **Data mining and knowledge discovery**, [S.l.], v.2, n.2, p.121–167, 1998.
- CHEN, T.-T.; LEE, S.-J. A weighted LS-SVM based learning system for time series forecasting. **Information Sciences**, [S.l.], v.299, p.99–116, 2015.

- COULIBALY, P.; ANCTIL, F.; BOBEE, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. **Journal of Hydrology**, [S.l.], v.230, n.3, p.244–257, 2000.
- COWPERTWAIT, P. S.; METCALFE, A. V. **Introductory time series with R**. [S.l.]: Springer Science & Business Media, 2009.
- DAS, P. P.; BISOI, R.; DASH, P. Time series forecasting using Fuzzy Functional link neural network trained by improved second order Levenberg-Marquardt algorithm. In: IEEE POWER, COMMUNICATION AND INFORMATION TECHNOLOGY CONFERENCE (PCITC), 2015. **Anais...** [S.l.: s.n.], 2015. p.827–833.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American statistical association**, [S.l.], v.74, n.366a, p.427–431, 1979.
- DONALDSON, R. G.; KAMSTRA, M. Forecast combining with neural networks. **Journal of Forecasting**, [S.l.], v.15, n.1, p.49–61, 1996.
- FIORDALISO, A. A nonlinear forecasts combination method based on Takagi–Sugeno fuzzy systems. **International journal of forecasting**, [S.l.], v.14, n.3, p.367–379, 1998.
- FLETCHER, R. **Practical methods of optimization**. [S.l.]: John Wiley & Sons, 2013.
- FRANKEL, J. A. Effects of speculation and interest rates in a “carry trade” model of commodity prices. **Journal of International Money and Finance**, [S.l.], v.42, p.88–112, 2014.
- GAXIOLA, F. et al. Interval type-2 fuzzy weight adjustment for backpropagation neural networks with application in time series prediction. **Information Sciences**, [S.l.], v.260, p.1–14, 2014.
- GRANGER, C. W.; RAMANATHAN, R. Improved methods of combining forecasts. **Journal of forecasting**, [S.l.], v.3, n.2, p.197–204, 1984.
- GUO, L.; LIEW, A. W.-C. Online-Offline Extreme Learning Machine with Concept Drift Tracking for Time Series Data. In: DIGITAL IMAGE COMPUTING: TECHNIQUES AND APPLICATIONS (DICTA), 2016 INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2016. p.1–6.
- GUO, T. et al. Robust Online Time Series Prediction with Recurrent Neural Networks. In: DATA SCIENCE AND ADVANCED ANALYTICS (DSAA), 2016 IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2016. p.816–825.
- HARRINGTON, P. d. B. Sigmoid transfer functions in backpropagation neural networks. **Analytical Chemistry**, [S.l.], v.65, n.15, p.2167–2168, 1993.
- HARVEY, A. C. **Forecasting, structural time series models and the Kalman filter**. [S.l.]: Cambridge university press, 1990.
- HAWKINS, D. M. The problem of overfitting. **Journal of chemical information and computer sciences**, [S.l.], v.44, n.1, p.1–12, 2004.
- HAYKIN, S.; NETWORK, N. A comprehensive foundation. **Neural Networks**, [S.l.], v.2, n.2004, 2004.

- HE, C.; XU, X. Combination of forecasts using self-organizing algorithms. **Journal of forecasting**, [S.l.], v.24, n.4, p.269, 2005.
- HIBON, M.; EVGENIOU, T. To combine or not to combine: selecting among forecasts and their combinations. **International Journal of Forecasting**, [S.l.], v.21, n.1, p.15–24, 2005.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural networks**, [S.l.], v.2, n.5, p.359–366, 1989.
- JIA, S. et al. Similarity measurement based on cloud models for time series prediction. In: CONTROL AND DECISION CONFERENCE (CCDC), 2016 CHINESE. **Anais...** [S.l.: s.n.], 2016. p.5138–5142.
- JOSE, V. R. R.; WINKLER, R. L. Simple robust averages of forecasts: some empirical results. **International Journal of Forecasting**, [S.l.], v.24, n.1, p.163–169, 2008.
- KARTHIKEYAN, L.; KUMAR, D. N. Predictability of nonstationary time series using wavelet and EMD based ARMA models. **Journal of Hydrology**, [S.l.], v.502, p.103–119, 2013.
- KAVOUSI-FARD, A.; SAMET, H.; MARZBANI, F. A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. **Expert systems with applications**, [S.l.], v.41, n.13, p.6047–6056, 2014.
- KAZEM, A. et al. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. **Applied soft computing**, [S.l.], v.13, n.2, p.947–958, 2013.
- KHASHEI, M.; BIJARI, M. An artificial neural network (p, d, q) model for timeseries forecasting. **Expert Systems with applications**, [S.l.], v.37, n.1, p.479–489, 2010.
- KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI. **Anais...** [S.l.: s.n.], 1995. v.14, n.2, p.1137–1145.
- KOHZADI, N. et al. A comparison of artificial neural network and time series models for forecasting commodity prices. **Neurocomputing**, [S.l.], v.10, n.2, p.169–181, 1996.
- KOURENTZES, N.; BARROW, D. K.; CRONE, S. F. Neural network ensemble operators for time series forecasting. **Expert Systems with Applications**, [S.l.], v.41, n.9, p.4235–4244, 2014.
- LEMKE, C. **Combinations of Time Series Forecasts: when and why are they beneficial?** 2010. Tese (Doutorado em Ciência da Computação) — Bournemouth University.
- LI, H.; PI, D.; JIANG, M. Forecasting chaotic time series with wavelet neural network improved by particle swarm optimization. **International Journal of Hybrid Information Technology**, [S.l.], v.6, n.6, p.369–380, 2013.
- LIU, H. et al. Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. **Energy Conversion and Management**, [S.l.], v.92, p.67–81, 2015.
- LIU, Y. et al. Demand forecasting for footwear products using wavelet transform and Artificial Bee Colony algorithm optimized Polynomial Fitting. In: NATURAL COMPUTATION (ICNC), 2015 11TH INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2015. p.1146–1150.

- LORENZ, E. N. Deterministic nonperiodic flow. **Journal of the atmospheric sciences**, [S.l.], v.20, n.2, p.130–141, 1963.
- MAHALAKSHMI, G.; SRIDEVI, S.; RAJARAM, S. A survey on forecasting of time series data. In: COMPUTING TECHNOLOGIES AND INTELLIGENT DATA ENGINEERING (ICCTIDE), INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2016. p.1–8.
- MATTOS NETO, P. S. de et al. Hybrid intelligent system for air quality forecasting using phase adjustment. **Engineering Applications of Artificial Intelligence**, [S.l.], v.32, p.185–191, 2014.
- MATTOS NETO, P. S. de et al. An approach to improve the performance of PM forecasters. **PloS one**, [S.l.], v.10, n.9, p.e0138507, 2015.
- MATTOS NETO, P. Salgado Gomes de. Teoria da perturbação em sistemas híbridos inteligentes para a previsão de séries temporais. , [S.l.], 2008.
- MENDES-MOREIRA, J. et al. Ensemble approaches for regression: a survey. **ACM Computing Surveys (CSUR)**, [S.l.], v.45, n.1, p.10, 2012.
- MIRANIAN, A.; ABDOLLAHZADE, M. Developing a local least-squares support vector machines-based neuro-fuzzy model for nonlinear and chaotic time series prediction. **IEEE Transactions on Neural Networks and Learning Systems**, [S.l.], v.24, n.2, p.207–218, 2013.
- MORETTI, F. et al. Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. **Neurocomputing**, [S.l.], v.167, p.3–7, 2015.
- NEUMAN, S. P. Maximum likelihood Bayesian averaging of uncertain model predictions. **Stochastic Environmental Research and Risk Assessment**, [S.l.], v.17, n.5, p.291–305, 2003.
- NGUYEN, T. H.; SHIRAI, K.; VELCIN, J. Sentiment analysis on social media for stock movement prediction. **Expert Systems with Applications**, [S.l.], v.42, n.24, p.9603–9611, 2015.
- NÓBREGA, J. P.; OLIVEIRA, A. L. Improving the statistical arbitrage strategy in intraday trading by combining extreme learning machine and support vector regression with linear regression models. In: TOOLS WITH ARTIFICIAL INTELLIGENCE (ICTAI), 2013 IEEE 25TH INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2013. p.182–188.
- NOVA FILHO, V.; PESSOA, S. R. Previsão de séries temporais utilizando pools de preditores criados a partir do particionamento da série e da divisão da tarefa de previsão. , [S.l.], 2015.
- NOWICKA-ZAGRAJEK, J.; WERON, R. Modeling electricity loads in California: arma models with hyperbolic noise. **Signal Processing**, [S.l.], v.82, n.12, p.1903–1915, 2002.
- OLIVEIRA, J. F. de; LUDERMIR, T. B. A hybrid evolutionary decomposition system for time series forecasting. **Neurocomputing**, [S.l.], v.180, p.27–34, 2016.
- OLIVEIRA, J. F. L. de; LUDERMIR, T. B. A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting. In: INTELLIGENT SYSTEMS (BRACIS), 2014 BRAZILIAN CONFERENCE ON. **Anais...** [S.l.: s.n.], 2014. p.73–78.
- OLIVEIRA, M.; TORGO, L. Ensembles for Time Series Forecasting. In: ACML. **Anais...** [S.l.: s.n.], 2014.

- OLIVEIRA, R. T. de et al. Copulas-based time series combined forecasters. **Information Sciences**, [S.l.], v.376, p.110–124, 2017.
- REIF, F. **Fundamentals of statistical and thermal physics**. [S.l.]: Waveland Press, 2009.
- RIBEIRO, G. H. et al. Lag selection for time series forecasting using particle swarm optimization. In: NEURAL NETWORKS (IJCNN), THE 2011 INTERNATIONAL JOINT CONFERENCE ON. **Anais...** [S.l.: s.n.], 2011. p.2437–2444.
- ROLI, F.; GIACINTO, G.; VERNAZZA, G. Methods for designing multiple classifier systems. In: INTERNATIONAL WORKSHOP ON MULTIPLE CLASSIFIER SYSTEMS. **Anais...** [S.l.: s.n.], 2001. p.78–87.
- ROSA, J. H. da; BARBOSA, J. L.; RIBEIRO, G. D. ORACON: an adaptive model for context prediction. **Expert Systems with Applications**, [S.l.], v.45, p.56–70, 2016.
- RUMELHART, D. E. et al. Learning representations by back-propagating errors. **Cognitive modeling**, [S.l.], v.5, n.3, p.1, 1988.
- SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. **IEEE transactions on acoustics, speech, and signal processing**, [S.l.], v.26, n.1, p.43–49, 1978.
- SAMARAWICKRAMA, N.; HEMAPALA, K.; JAYASEKARA, A. Support Vector Machine Regression for forecasting electricity demand for large commercial buildings by using kernel parameter and storage effect. In: MORATUWA ENGINEERING RESEARCH CONFERENCE (MERCEN), 2016. **Anais...** [S.l.: s.n.], 2016. p.162–167.
- SANTAMARÍA-BONFIL, G.; REYES-BALLESTEROS, A.; GERSHENSON, C. Wind speed forecasting for wind farms: a method based on support vector regression. **Renewable Energy**, [S.l.], v.85, p.790–809, 2016.
- SCHMIDT, W. F.; KRAAIJVELD, M. A.; DUIN, R. P. Feedforward neural networks with random weights. In: PATTERN RECOGNITION, 1992. VOL. II. CONFERENCE B: PATTERN RECOGNITION METHODOLOGY AND SYSTEMS, PROCEEDINGS., 11TH IAPR INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 1992. p.1–4.
- SCHWARZ, G. et al. Estimating the dimension of a model. **The annals of statistics**, [S.l.], v.6, n.2, p.461–464, 1978.
- SERGIO, A. T.; LIMA, T. P. de; LUDERMIR, T. B. Dynamic selection of forecast combiners. **Neurocomputing**, [S.l.], v.218, p.37–50, 2016.
- SHI, S. M.; DA XU, L.; LIU, B. Improving the accuracy of nonlinear combined forecasting using neural networks. **Expert Systems with Applications**, [S.l.], v.16, n.1, p.49–54, 1999.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and computing**, [S.l.], v.14, n.3, p.199–222, 2004.
- TAKEYASU, K.; HIGUCHI, Y.; TSUCHIDA, Y. The Method to Improve Forecasting Accuracy under the Utilization of Neural Network. In: SOFT COMPUTING AND INTELLIGENT SYSTEMS (SCIS) AND 17TH INTERNATIONAL SYMPOSIUM ON ADVANCED INTELLIGENT SYSTEMS, 2016 JOINT 8TH INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2016. p.208–213.

- TANG, L.; DU, S.; JI, S. Forecasting network traffic at large time scale by using dual-related method. In: NATURAL COMPUTATION, FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY (ICNC-FSKD), 2016 12TH INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2016. p.1336–1340.
- TICKNOR, J. L. A Bayesian regularized artificial neural network for stock market forecasting. **Expert Systems with Applications**, [S.l.], v.40, n.14, p.5501–5506, 2013.
- TIMMERMANN, A. Forecast combinations. **Handbook of economic forecasting**, [S.l.], v.1, p.135–196, 2006.
- TSOUMAKAS, G.; PARTALAS, I.; VLAHAVAS, I. A taxonomy and short review of ensemble selection. In: ECAI 2008, WORKSHOP ON SUPERVISED AND UNSUPERVISED ENSEMBLE METHODS AND THEIR APPLICATIONS. **Anais...** [S.l.: s.n.], 2008. p.41–46.
- VAPNIK, V. Pattern recognition using generalized portrait method. **Automation and remote control**, [S.l.], v.24, p.774–780, 1963.
- VAPNIK, V. N.; VAPNIK, V. **Statistical learning theory**. [S.l.]: Wiley New York, 1998. v.1.
- WANG, L.; ZENG, Y.; CHEN, T. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. **Expert Systems with Applications**, [S.l.], v.42, n.2, p.855–863, 2015.
- WICHARD, J. D.; OGORZALEK, M. Time series prediction with ensemble models. In: NEURAL NETWORKS, 2004. PROCEEDINGS. 2004 IEEE INTERNATIONAL JOINT CONFERENCE ON. **Anais...** [S.l.: s.n.], 2004. v.2, p.1625–1630.
- WIDMER, G.; KUBAT, M. Learning in the presence of concept drift and hidden contexts. **Machine learning**, [S.l.], v.23, n.1, p.69–101, 1996.
- WIDODO, A.; BUDI, I. Combination of time series forecasts using neural network. In: ELECTRICAL ENGINEERING AND INFORMATICS (ICEEI), 2011 INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2011. p.1–6.
- WOLPERT, D. H. Stacked generalization. **Neural networks**, [S.l.], v.5, n.2, p.241–259, 1992.
- WU, J.; CHENG, E. A novel hybrid particle swarm optimization for feature selection and kernel optimization in support vector regression. In: COMPUTATIONAL INTELLIGENCE AND SECURITY (CIS), 2010 INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2010. p.189–194.
- WU, S.-F.; LEE, S.-J. Employing local modeling in machine learning based methods for time-series prediction. **Expert Systems with Applications**, [S.l.], v.42, n.1, p.341–354, 2015.
- XIN, N. et al. Application of genetic algorithm-support vector regression (GA-SVR) for quantitative analysis of herbal medicines. **Journal of Chemometrics**, [S.l.], v.26, n.7, p.353–360, 2012.
- YOLCU, U.; EGRIOGLU, E.; ALADAG, C. H. A new linear & nonlinear artificial neural network model for time series forecasting. **Decision support systems**, [S.l.], v.54, n.3, p.1340–1347, 2013.

- YU, L. et al. Time series forecasting with multiple candidate models: selecting or combining. **Journal of Systems Science and Complexity**, [S.l.], v.18, n.1, p.1–18, 2005.
- YU, W. et al. Improvement of rainfall and flood forecasts by blending ensemble NWP rainfall with radar prediction considering orographic rainfall. **Journal of Hydrology**, [S.l.], v.531, p.494–507, 2015.
- ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, [S.l.], v.50, p.159–175, 2003.
- ZHANG, G. P.; QI, M. Neural network forecasting for seasonal and trend time series. **European journal of operational research**, [S.l.], v.160, n.2, p.501–514, 2005.
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks:: the state of the art. **International journal of forecasting**, [S.l.], v.14, n.1, p.35–62, 1998.
- ZHANG, L. et al. Forecasting price shocks with social attention and sentiment analysis. In: IEEE/ACM INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING (ASONAM), 2016. **Anais...** [S.l.: s.n.], 2016. p.559–566.
- ZHANG, R. et al. A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts. In: POWER AND ENERGY SOCIETY GENERAL MEETING (PESGM), 2016. **Anais...** [S.l.: s.n.], 2016. p.1–5.
- ZHAO, J. et al. Levenberg-Marquardt algorithm for Mackey-Glass chaotic time series prediction. **Discrete Dynamics in Nature and Society**, [S.l.], v.2014, 2014.
- ZOU, H.; YANG, Y. Combining time series models for forecasting. **International journal of Forecasting**, [S.l.], v.20, n.1, p.69–84, 2004.