



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

ERAYLSON GALDINO DA SILVA

**UMA ABORDAGEM DE SELEÇÃO DINÂMICA DE PREDITORES BASEADA  
NAS JANELAS TEMPORAIS MAIS RECENTES**

Recife

2021

ERAYLSON GALDINO DA SILVA

**UMA ABORDAGEM DE SELEÇÃO DINÂMICA DE PREDITORES BASEADA  
NAS JANELAS TEMPORAIS MAIS RECENTES**

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática (CIn) da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Doutor em Ciência da Computação.

**Área de Concentração:** Inteligência Computacional

**Orientador (a):** Paulo Salgado Gomes de Mattos Neto

**Coorientador (a):** George Darmiton da Cunha Cavalcanti

Recife

2021

Catálogo na fonte  
Bibliotecária Mônica Uchôa, CRB4-1010

S586a Silva, Eraylson Galdino da.  
Uma abordagem de seleção dinâmica de preditores baseada nas janelas temporais mais recentes / Eraylson Galdino da Silva. – 2021.  
86 f.: il., fig., tab.

Orientador: Paulo Salgado Gomes de Mattos Neto.  
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2021.  
Inclui referências.

1. Seleção dinâmica. 2. Sistema de múltiplos preditores. 3. Séries temporais. 4. Ensemble. I. Mattos Neto, Paulo Salgado Gomes de (Orientador). II. Título.

005.1 CDD (23. ed.) UFPE - CCEN 2022 – 002

**Eraylson Galdino da Silva**

**“Uma Abordagem de Seleção Dinâmica de Preditores Baseada nas Janelas Temporais mais Recentes”**

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

Aprovado em: 16/08/2021.

---

**Orientador: Prof. Dr. Paulo Salgado Gomes de Mattos Neto**

**BANCA EXAMINADORA**

---

Prof. Dr. Tsang Ing Ren  
Centro de Informática / UFPE

---

Prof. Dr. Paulo Renato Alves Firmino  
Centro de Ciências e Tecnologia / UFCA

---

Prof. Dr. Tiago Alessandro Espínola Ferreira  
Departamento de Estatística e Informática / UFRPE

---

Prof. Dr. Hugo Valadares Siqueira UPE  
Departamento Acadêmico de Engenharia Elétrica / UTFPR

---

Prof. Dr. João Fausto Lorenzato de Oliveira  
Escola Politécnica de Pernambuco

Dedico esta Tese 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, aos meus sobrinhos e a minha companheira, 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 e por todo o ensinamento que levarei para toda a minha jornada como pesquisador.

Agradeço aos Professores do Centro de Informática da UFPE, por todo o conhecimento compartilhando, que foram fundamentais para minha formação como Doutor. E aos professores do curso de Licenciatura em Computação na UPE, que me iniciaram no 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 Doutorado.

## RESUMO

O desenvolvimento de sistemas de previsão acurados para modelagem de séries temporais do mundo real é uma tarefa desafiadora. Devido à presença de padrões temporais que mudam ao longo do tempo, a adoção de um modelo único pode levar a previsões de baixa acurácia. Nesse cenário, Sistemas de Múltiplo Preditores (SMP) surgem como uma alternativa promissora. A seleção dinâmica de preditores é um caso especial de SMP, onde cada modelo é um especialista em padrões específicos da série temporal. Na seleção dinâmica os modelos mais competentes são selecionados para cada novo padrão de teste. Um critério frequentemente utilizado é avaliar o desempenho dos modelos na região de competência, formada pelos padrões (presentes nos conjuntos de treinamento ou validação) mais similares ao padrão de teste. A qualidade da região de competência é um fator chave na precisão do SMP. Porém, definir adequadamente a medida de similaridade e o tamanho da região de competência é um desafio. Além disso, não há garantia de que existam padrões similares no conjunto de treinamento ou validação. Esta tese propõe uma abordagem de seleção dinâmica intitulada Dynamic Selection based on the Nearest Windows (DSNAW) que escolhe um ou mais modelos de acordo com seu desempenho na região de competência composta pelas janelas que antecedem à nova janela da observação de teste. Essa estratégia assume que as janelas temporais mais próximas do novo padrão de teste têm um comportamento mais similar à observação a ser predita do que os dados de treinamento. O estudo experimental utilizando dez séries temporais mostrou que a abordagem proposta supera as abordagens da literatura, consolidando-a como uma nova abordagem para seleção dinâmica de preditores.

**Palavras-chave:** seleção dinâmica; sistema de múltiplos preditores; séries temporais; ensemble.

## ABSTRACT

The development of accurate forecasting systems for real-world time series modeling is a challenging task. Due to the presence of temporal patterns that change over time, the adoption of a single model can lead to underperformed forecasts. In this scenario, Multiple Predictor Systems (MPS) emerge as an alternative to adopting single models since they struggle to learn in the presence of temporal patterns that change over time. Dynamic prediction/ensemble selection is a special case of MPS where each model is an expert in the time series's specific patterns. In dynamic selection, instead of combining all models, the most competent models per test pattern are selected. A criterion commonly used is to evaluate the models' performance in the region of competence, formed by the patterns present in the in-sample set (training or validation sets) more similar to the test pattern. Thus, the region of competence's quality is a key factor in the precision of the MPS. However, adequately defining the similarity criterion and the size of the region of competence is challenging and problem-dependent. Furthermore, there is no guarantee that similar data exist in the in-sample set. This thesis proposes a dynamic selection approach entitled Dynamic Selection based on the Nearest Windows (DSNAW) that selects one or more competent models according to their performance in the region of competence composed of the nearest antecedent windows to the new target time window. This strategy assumes that the temporal windows closer to a test pattern have a behavior more similar to the target than in-sample data. The experimental study using ten well-known time series showed that the DSNAW outperforms the literature approaches.

**Keywords:** dynamic selection; multiple predictor system; time series; ensemble.



## LISTA DE FIGURAS

Figura 1 – Preço de fechamento das ações da Petrobras em 2019, obtido em <i>Yahoo Finance</i> ( <a href="http://finance.yahoo.com/quote/PETR4.SA/">http://finance.yahoo.com/quote/PETR4.SA/</a> ). . . . .	23
Figura 2 – Correlograma dos 20 primeiros pontos da série apresentada na Figura 1. . .	24
Figura 3 – A figura ilustra duas regiões de competência: "Literatura RoC", formada por janelas de treinamento mais similares a $w_t$ , "Proposta RoC", região de competência formada por janelas que antecedem $w_t$ . . . . .	36
Figura 4 – Percentual médio de janelas em cada RoC que apresenta a mesma distribuição das respectivas janelas do <i>out-of-sample</i> . As barras em cinza e preto corresponde as abordagens "Literatura RoC" e "Proposta RoC", respectivamente. . . . .	38
Figura 5 – Fluxograma do SMP proposto utilizando seleção dinâmica. $\phi$ é o conjunto de treinamento; $\gamma$ é o conjunto de teste; $P$ é o <i>Pool</i> de preditores; $w_t$ é a nova janela de teste $\gamma$ ; $W$ é o histórico de janelas; $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$ são as previsões dos preditores selecionados pelo DSNAW, e $\hat{z}_{t+1}$ é a previsão final do SMP. . . . .	39
Figura 6 – Fluxograma da fase de seleção dinâmica utilizando a abordagem proposta DSNAW. $W$ é o histórico de janelas, que vai sendo atualizado de acordo com as novas observações da série temporal; $\Delta_t^{PW}$ é a região de competência; $P$ é o <i>Pool</i> de preditores; $P'$ é conjunto de modelos selecionados e $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$ são as previsões obtidas de cada preditor selecionado. . . . .	41
Figura 7 – Série Amazon . . . . .	45
Figura 8 – Série Apple . . . . .	46
Figura 9 – Série Electricity . . . . .	47
Figura 10 – Série Goldman Sachs . . . . .	48
Figura 11 – Série Microsoft . . . . .	49
Figura 12 – Série Pollution . . . . .	50
Figura 13 – Série Star . . . . .	51
Figura 14 – Série Sunspot . . . . .	52
Figura 15 – Série Vehicle . . . . .	53
Figura 16 – Série Wine . . . . .	54

Figura 17 – Previsões para o intervalo de 100 até 200 observações do conjunto de teste da série *Amazon*, utilizando o DSNAW e as abordagens da literatura que apresentaram os melhores resultados. . . . . 72

Figura 18 – Previsões para as 100 primeiras observações do conjunto de teste da série *Star*, utilizando o DSNAW e as abordagens da literatura que apresentaram os melhores resultados. . . . . 73

## LISTA DE TABELAS

Tabela 1 – Valores para os parâmetros da SVR . . . . .	57
Tabela 2 – Valores dos parâmetros do DSNAW selecionados por meio do conjunto de validação em comparação aos parâmetros que apresentam o melhor desempenho no conjunto de teste. . . . .	61
Tabela 3 – Comparação do desempenho do método proposto em relação aos métodos monolíticos. Em negrito estão os valores com o maior desempenho. . . . .	63
Tabela 4 – Comparação do desempenho do método proposto em relação aos métodos de seleção dinâmica de preditores. Em negrito estão os valores com o maior desempenho. . . . .	65
Tabela 5 – Comparação do desempenho do método proposto em relação aos métodos que combinam todas as previsões do <i>Pool</i> . Em negrito estão os valores com o maior desempenho. . . . .	67
Tabela 6 – Comparação do desempenho do método proposto em relação aos métodos de seleção dinâmica de <i>Ensemble</i> . Em negrito estão os valores com o maior desempenho. . . . .	69
Tabela 7 – Comparação do desempenho do método proposto em relação aos métodos da literatura. Em negrito estão os valores com o maior desempenho. . . . .	70
Tabela 8 – Diferença Percentual (%) entre o DSNAW e as abordagens da literatura calculado por meio da equação 4.5 em relação aos resultados do MSE . . . . .	71
Tabela 9 – Resultado do teste estatístico DM ( valor crítico $\alpha = 0.05$ ) que compara o desempenho do DSNAW em relação aos métodos da literatura na perspectiva da métrica MSE. Os símbolos “+”, “–” e “ $\approx$ ” corresponde ao desempenho superior, pior e igual, respectivamente. . . . .	72

## LISTA DE ABREVIATURAS E SIGLAS

<b>AdaBoosting</b>	<i>AdaBoost Regression and Threshold</i>
<b>AIC</b>	<i>Akaike's Information Criterion</i>
<b>ARIMA</b>	Auto Regressivo Integrado de Média Móveis
<b>ARV</b>	<i>Average Relative Prediction Error Variation</i>
<b>Bagging</b>	<i>Bootstrap Aggregating</i>
<b>DCS-LA</b>	<i>Dynamic Classifier Selection by Local Accuracy</i>
<b>DES</b>	<i>Dynamic Ensemble Selection</i>
<b>DM</b>	<i>Diebold-Mariano Test</i>
<b>DS</b>	<i>Dynamic Selection</i>
<b>DTW</b>	<i>Dynamic Time Warping</i>
<b>ETS</b>	<i>Exponential Smoothing</i>
<b>LSTM</b>	<i>Long Short-Term Memory</i>
<b>MAE</b>	<i>Mean Absolute Error</i>
<b>MLP</b>	<i>Multilayer Perceptron</i>
<b>MSE</b>	<i>Mean Squared Error</i>
<b>RBF</b>	<i>Radial Basis Function</i>
<b>SMAPE</b>	<i>Symmetric Mean Absolute Percentage Error</i>
<b>SMP</b>	Sistema de Múltiplos Preditores
<b>SVR</b>	<i>Support Vector Regression</i>
<b>TWF</b>	<i>Temporal-Window Framework</i>

## LISTA DE SÍMBOLOS

$Z$	Conjunto de observações
$z_t$	Observação no tempo $t$
$w_t$	Janela temporal no tempo $t$
$\beta$	Parâmetro do modelo AR
$\theta$	Parâmetro do modelo MA
$\Delta_t$	Região de Competência
$\phi$	Conjunto de Treinamento
$\gamma$	Conjunto de Teste

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
1.1	MOTIVAÇÃO	16
1.2	OBJETIVOS	19
1.3	QUESTÕES DE PESQUISA	19
1.4	HIPÓTESES DA PESQUISA	20
1.5	ESTRUTURA DA TESE	20
<b>2</b>	<b>CONCEITOS E TÉCNICAS</b>	<b>22</b>
2.1	PREVISÃO DE SÉRIES TEMPORAIS	22
2.2	MODELOS DE PREVISÃO PARA SÉRIES TEMPORAIS	25
<b>2.2.1</b>	<b>Modelos estatísticos</b>	<b>25</b>
<b>2.2.2</b>	<b>Modelos de Aprendizado de Máquina</b>	<b>26</b>
2.3	SISTEMA DE MÚLTIPLOS PREDITORES	28
<b>2.3.1</b>	<b>Geração</b>	<b>28</b>
<b>2.3.2</b>	<b>Seleção</b>	<b>29</b>
<b>2.3.3</b>	<b>Integração</b>	<b>30</b>
2.4	TRABALHOS RELACIONADOS	31
2.5	CONSIDERAÇÕES DO CAPÍTULO	33
<b>3</b>	<b>MÉTODO PROPOSTO</b>	<b>35</b>
3.1	DEFINIÇÃO DO PROBLEMA	35
3.2	ARQUITETURA PROPOSTA	39
<b>3.2.1</b>	<b>Geração</b>	<b>39</b>
<b>3.2.2</b>	<b>Seleção Dinâmica</b>	<b>40</b>
<b>3.2.3</b>	<b>Integração</b>	<b>42</b>
3.3	CONSIDERAÇÕES DO CAPÍTULO	42
<b>4</b>	<b>METODOLOGIA DOS EXPERIMENTOS</b>	<b>44</b>
4.1	BASE DE DADOS	44
<b>4.1.1</b>	<b>Amazon</b>	<b>44</b>
<b>4.1.2</b>	<b>Apple</b>	<b>45</b>
<b>4.1.3</b>	<b>Electricity</b>	<b>46</b>
<b>4.1.4</b>	<b>Goldman Sachs</b>	<b>47</b>

4.1.5	Microsoft . . . . .	48
4.1.6	Pollution . . . . .	49
4.1.7	Star . . . . .	50
4.1.8	Sunspot . . . . .	51
4.1.9	Vehicle . . . . .	52
4.1.10	Wine . . . . .	53
4.2	MÉTRICAS DE AVALIAÇÃO . . . . .	54
4.3	TESTE ESTATÍSTICO . . . . .	55
4.4	CONFIGURAÇÃO DO SISTEMA DE MÚLTIPLOS PREDITORES . . . . .	56
4.5	COMPARAÇÃO COM A LITERATURA . . . . .	57
4.5.1	Modelos Monolíticos . . . . .	57
4.5.2	Seleção Dinâmica de Preditores . . . . .	58
4.5.3	<i>Ensemble</i> . . . . .	59
4.5.4	Seleção Dinâmica de <i>Ensemble</i> . . . . .	59
4.6	CONSIDERAÇÕES DO CAPÍTULO . . . . .	60
5	<b>ANÁLISE E RESULTADOS . . . . .</b>	<b>61</b>
5.1	ANÁLISE DOS PARÂMETROS SELECIONADOS . . . . .	61
5.2	COMPARAÇÃO COM OS MODELOS MONOLÍTICOS . . . . .	62
5.3	COMPARAÇÃO COM OS MÉTODOS DE SELEÇÃO DINÂMICA DE PRE- DITORES . . . . .	64
5.4	COMPARAÇÃO COM OS MÉTODOS <i>ENSEMBLE</i> . . . . .	66
5.5	COMPARAÇÃO COM OS MÉTODOS DE SELEÇÃO DINÂMICA DE <i>EN- SEMBLE</i> . . . . .	68
5.6	ANÁLISE GERAL . . . . .	70
5.7	CONSIDERAÇÕES FINAIS . . . . .	75
6	<b>CONCLUSÃO . . . . .</b>	<b>76</b>
6.1	RESUMO DO TRABALHO . . . . .	76
6.2	QUESTÕES DE PESQUISA . . . . .	77
6.2.1	Uma região de competência composta por janelas provenientes do conjunto de treinamento possui a mesma distribuição de um dado novo padrão de teste? . . . . .	77

6.2.2	Uma região de competência composta por padrões mais recentes possui uma maior quantidade de janelas com a mesma distribuição do padrão de teste que regiões de competência usadas na literatura? 77
6.3	PRINCIPAIS CONTRIBUIÇÕES DA TESE . . . . . 77
6.4	LIMITAÇÕES DA PESQUISA . . . . . 78
6.5	TRABALHOS FUTUROS . . . . . 78
6.6	ARTIGOS PUBLICADOS . . . . . 79
	<b>REFERÊNCIAS . . . . . 80</b>



## 1 INTRODUÇÃO

Este capítulo introduz o problema e as questões que norteiam a pesquisa. As cinco seções estão organizadas da seguinte forma: Seção 1.1 apresenta a motivação do trabalho, Seção 1.2 lista os objetivos, Seção 1.3 descreve as questões de pesquisa, Seção 1.4 introduz as hipóteses e, por fim, a estrutura da tese é apresentada.

### 1.1 MOTIVAÇÃO

Existem diferentes modelos estatísticos e de aprendizado de máquina que podem ser utilizados na previsão de séries temporais, tais como: *Multilayer Perceptron* (MLP) (SAMARA; NATSHEH, 2018), *Support Vector Regression* (SVR) (EHSAN; SIMON; VENKATESWARAN, 2017), *Radial Basis Function* (RBF) (RUBIO et al., 2017) e *Long Short-Term Memory* (LSTM) (MA et al., 2015). Cada modelo tem sua particularidade e pode ser o mais acurado para modelar um comportamento específico da série temporal. De acordo com o teorema do *No-free-lunch* (WOLPERT, 1996), não existe um modelo que apresente a maior acurácia em todas as situações. Logo, a variedade de possíveis modelos para realizar a previsão e o fato de que nenhum modelo é o melhor para todos os casos tornam incerta a atividade de selecionar um modelo para realizar a previsão. Além disso, é difícil definir um critério de seleção de preditores que possa ser aplicado a todas as séries temporais (WANG et al., 2018; ZOU; YANG, 2004). Essa dificuldade está associada ao comportamento da série temporal, que pode se alterar no decorrer do tempo (CAO; GU, 2002).

Os Sistema de Múltiplos Preditores (SMP), também conhecidos como *Ensembles*, vêm sendo cada vez mais utilizados em previsão de séries temporais (ZHANG et al., 2021; NETO et al., 2020; WANG; ZHANG; LU, 2019; YAO; DAI; SONG, 2018; OLIVEIRA et al., 2017). A utilização de um SMP permite obter maior acurácia na previsão (QIU et al., 2017) e reduzir a incerteza da escolha do modelo selecionar (KOURENTZES; BARROW; CRONE, 2014). O objetivo de um SMP é obter um sistema com previsões mais acuradas que preditores utilizados de forma individual, conhecidos como monolíticos, por meio da combinação e/ou seleção de preditores com diferentes características.

Um SMP é composto por três fases (YAO; DAI; SONG, 2018): (1) geração; (2) seleção; (3) integração. A primeira fase consiste em gerar um conjunto de preditores, também conhecido

como *Pool* de preditores, a partir dos dados de treinamento. Nesse momento é importante que sejam gerados preditores que apresentem diferentes previsões para uma mesma entrada, ou seja, que exista diversidade entre os modelos. Dessa forma, é possível que o SMP resulte em previsões mais acuradas que as abordagens que utilizam apenas um preditor (RAHMAN et al., 2015). Para tal, existem duas formas principais de se obter um conjunto de preditores com diversidade: (I) por meio do treinamento do conjunto de modelos utilizando dados diferentes da série, mas utilizando um único algoritmo de aprendizado de máquina, resultando em um *Pool* homogêneo; (II) usando modelos com características diferentes, resultando em um *Pool* heterogêneo. Para o primeiro caso, são utilizadas técnicas de re-amostragem, como o *Bootstrap Aggregating* (Bagging) (MORETTI et al., 2015) e *AdaBoost Regression and Threshold* (AdaBoosting) (BARROW; CRONE, 2016). Já no segundo caso, são utilizados diferentes algoritmos, como MLP (SAMARA; NATSHEH, 2018), SVR (EHSAN; SIMON; VENKATESWARAN, 2017), RBF (RUBIO et al., 2017) e LSTM (HOCHREITER; SCHMIDHUBER, 1997).

Na segunda fase, seleciona-se um ou mais preditores do *Pool* com base em algum critério. Finalmente, na última etapa do SMP é obtida a previsão final por meio da combinação das previsões dos modelos selecionados na etapa anterior.

A seleção possibilita aumentar a acurácia do sistema (LIANG; JIA, 2018) e reduzir o número de preditores que serão considerados para a previsão do SMP, reduzindo a complexidade computacional (RAFTERY et al., 2005). Na literatura, diversas abordagens utilizam seleção com o objetivo de encontrar os melhores preditores (PRUDÊNCIO; LUDERMIR, 2004; CANG; YU, 2014; MA; DAI; LIU, 2015; MA; DAI, 2016; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; BUDI; AJI; WIDODO, 2014; KRIKUNOV; KOVALCHUK, 2015; YAO; DAI; SONG, 2018). A maioria dessas abordagens realiza a seleção de forma estática, em que os modelos são selecionados apenas uma vez e aplicados em todos os novos dados. Porém, sabe-se que séries temporais do mundo real apresentam um comportamento dinâmico (CAO; GU, 2002) e, por conta disso, abordagens de seleção dinâmica tendem a apresentar desempenho superior às abordagens de seleção estática (YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; BUDI; AJI; WIDODO, 2014; KRIKUNOV; KOVALCHUK, 2015; YAO; DAI; SONG, 2018).

A seleção dinâmica tem como suposição que cada modelo é especialista em um comportamento específico dos dados (ZHU; WU; YANG, 2004) e, por essa razão, se faz necessária a mudança de modelos de acordo com a mudança de comportamento. Dessa forma, para cada novo padrão, a abordagem de seleção dinâmica escolhe um ou mais modelos seguindo algum critério de seleção.

Definir qual critério de seleção utilizar para escolher os modelos é uma etapa fundamental para o desempenho da abordagem de seleção (CRUZ et al., 2015). A abordagem de seleção mais utilizada em séries temporais, *Dynamic Selection* (DS) (ROONEY et al., 2004), é baseada na técnica de seleção dinâmica aplicada na tarefa de classificação, *Dynamic Classifier Selection by Local Accuracy* (DCS-LA) (WOODS; KEGELMEYER; BOWYER, 1997) e seleciona os modelos de acordo com seu desempenho em um conjunto de padrões (região de competência) que foram selecionados baseados na similaridade entre os padrões de treinamento e o novo padrão de teste.

Em (SILVA et al., 2020) foi investigada a influência dos parâmetros utilizados na abordagem DS, tamanho da região de competência e medida de similaridade. Dez diferentes medidas de similaridades foram avaliadas, sendo que para o tamanho da região de competência foram avaliados valores de 1 até 10. Os resultados obtidos mostram que esses parâmetros influenciam diretamente a abordagem do DS, porém, não foi encontrada uma combinação de parâmetros que apresentasse o melhor desempenho em todas as séries temporais avaliadas. Outra contribuição importante do trabalho (SILVA et al., 2020) é a comparação do desempenho do DS utilizando os melhores parâmetros em relação ao desempenho do *Oracle* (abordagem hipotética que seleciona o melhor preditor para cada ponto da amostra de teste). Essa comparação mostra que a abordagem de seleção dinâmica de preditores mais utilizada na literatura (NADEEM et al., 2019; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011), o DS, apresenta um desempenho inferior ao do *Oracle*, ou seja, essa abordagem não foi capaz de selecionar os melhores preditores do *Pool*.

A acurácia das abordagens de seleção dinâmica de preditores está diretamente associada à qualidade da região de competência. Porém, como mostra a literatura (SILVA et al., 2020; CRUZ et al., 2015; OLIVEIRA; SILVA; NETO, 2021), a escolha dos melhores parâmetros para a construção da região de competência depende da base de dados utilizada. Além disso, não é possível garantir a existência de padrões no conjunto de treinamento com comportamentos similares ao padrão de teste, devido ao ruído nos dados, ao tamanho da amostra, ou à ausência de tais padrões (CRUZ; SABOURIN; CAVALCANTI, 2018). Já no contexto de séries temporais, a distribuição dos dados pode mudar no decorrer do tempo (ELWELL; POLIKAR, 2011; ADHIKARI, 2015), consequentemente, o melhor modelo para prever a nova observação também pode mudar (MEADE, 2000).

Nesse contexto, é importante destacar que a seleção dinâmica de preditores pode ser uma abordagem utilizada para aumentar a acurácia da previsão e que definir a região de compe-

tência é uma tarefa importante na escolha dos melhores modelos. Portanto, é necessária uma abordagem para construir a região de competência levando em consideração a mudança de comportamento da série temporal. Essa necessidade é a principal motivação para o desenvolvimento deste trabalho, que se pautará no estudo e implementação de abordagens de seleção dinâmica de preditores.

## 1.2 OBJETIVOS

O objetivo principal deste trabalho é a proposição de um método para a construção de Sistema de Múltiplos Preditores (SMP) utilizando seleção dinâmica de preditores. O método proposto visa selecionar preditores em uma região de competência formada por padrões com o mesmo comportamento da observação a ser predita.

O objetivo geral deste trabalho pode ser decomposto nos seguintes objetivos específicos:

1. Implementar e investigar a abordagem de construção da região de competência mais utilizada na literatura;
2. Investigar alternativas para a construção da região de competência;
3. Aplicar a abordagem desenvolvida em séries temporais reais;
4. Analisar o comportamento da abordagem proposta a partir da variação de seus parâmetros e de diferentes séries temporais;
5. Analisar o desempenho da abordagem proposta e comparar com abordagens já existentes na literatura;

## 1.3 QUESTÕES DE PESQUISA

Esta pesquisa é guiada pelas seguintes questões de pesquisa:

- **Uma região de competência composta por janelas provenientes do conjunto de treinamento possui a mesma distribuição de um dado novo padrão de teste?**

Janelas do conjunto de treinamento podem não apresentar o mesmo comportamento (distribuição) das novas janelas. Por isso, é importante avaliar se a região de competência

formada por essas janelas é capaz de apresentar o mesmo comportamento do novo padrão de teste.

- **Uma região de competência composta por padrões mais recentes possui uma maior quantidade de janelas com a mesma distribuição do padrão de teste que regiões de competência usadas na literatura?**

É possível considerar que o modelo com maior acurácia para prever  $z_{t+1}$  mude no decorrer do tempo de acordo com a distribuição. A partir disso, é esperado que uma região de competência formada por  $k$  janelas que antecedem a nova janela de teste e o padrão a ser predito tenha alta probabilidade de apresentarem a mesma distribuição. Dessa forma, o modelo que apresentar maior acurácia na região de competência é um candidato promissor para realizar a previsão com acurácia para o ponto  $z_{t+1}$ .

## 1.4 HIPÓTESES DA PESQUISA

As seguintes hipóteses integram esta pesquisa:

- Uma região de competência formada por janelas de treinamento ou validação só apresenta o mesmo comportamento do novo padrão de teste em séries com comportamento estacionário.
- Uma região de competência formada por janelas que antecedem o ponto a ser predito tem uma maior chance de ter o mesmo comportamento do novo padrão de teste e, por isso, usada com maior sucesso para selecionar preditores com maior acurácia.

## 1.5 ESTRUTURA DA TESE

A estrutura deste trabalho é composta por 6 capítulos, conforme descritos a seguir:

**Capítulo 1 - Introdução:** apresenta de forma introdutória o contexto de previsões de séries temporais, bem como a motivação e os objetivos da realização deste trabalho;

**Capítulo 2 - Conceitos e técnicas:** apresenta os conceitos e as técnicas utilizados na pesquisa, assim como os principais trabalhos relacionados;

**Capítulo 3 - Método Proposto:** apresenta a metodologia e o resultado do experimento realizado para avaliar diferentes alternativas para a construção da região de competência. O resultado descreve as etapas do método proposto;

**Capítulo 4 - Metodologia dos Experimentos:** apresenta as séries temporais utilizadas, destacando suas características e comportamentos, os parâmetros empregados em cada modelo, as métricas de avaliação e o processo de execução dos experimentos;

**Capítulo 5 - Resultados Experimentais:** os experimentos realizados são descritos e seus resultados são apresentados e discutidos;

**Capítulo 6 - Conclusões e Trabalhos Futuros:** apresenta um resumo deste trabalho juntamente às conclusões obtidas com base nos experimentos, além das limitações e propostas para trabalhos futuros;

## 2 CONCEITOS E TÉCNICAS

Este capítulo apresenta os conceitos, as técnicas utilizadas e os principais trabalhos relacionados à pesquisa. A Seção 2.1 apresenta a definição e as características de séries temporais e as etapas para realizar a previsão. A Seção 2.2 descreve os principais modelos de aprendizado de máquina utilizados na modelagem de séries temporais. A Seção 2.3 traz o conceito e as etapas de um Sistema de Múltiplos Preditores. Já a Seção 2.4 introduz os trabalhos relacionados. Finalmente, a Seção 2.5 apresenta as considerações finais do capítulo.

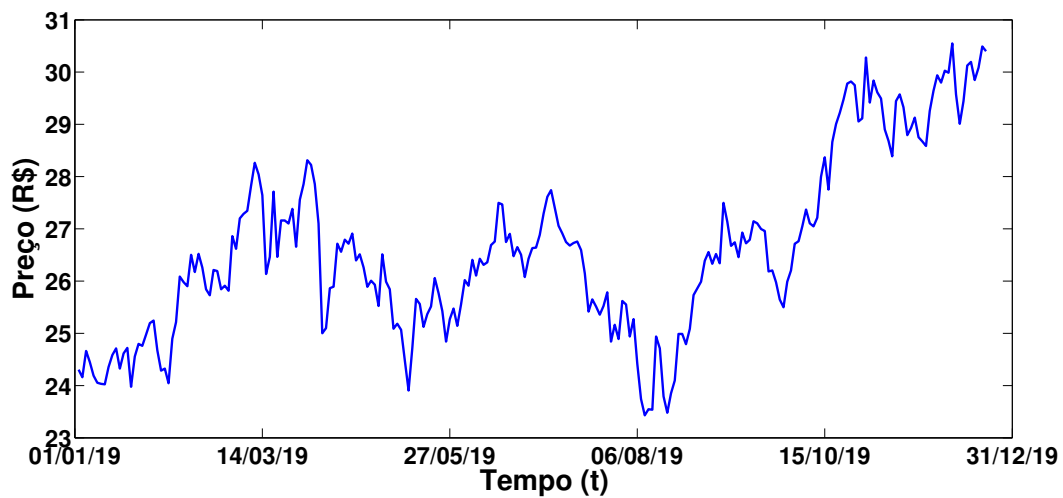
### 2.1 PREVISÃO DE SÉRIES TEMPORAIS

Uma série temporal  $Z$  é um conjunto de observações  $z_t$  ordenadas (BOX et al., 2015) capturadas em um intervalo específico do tempo  $t$  (BROCKWELL; DAVIS, 2016). Dessa forma, temos que uma série temporal com  $s$  pontos tem a seguinte expressão matemática:  $Z = (z_t, z_{t+1}, \dots, z_{t+s-1})$ . Séries temporais podem ser compostas por valores discretos ou contínuos (BROCKWELL; DAVIS; FIENBERG, 1991) e são associadas a um fenômeno gerador. Alguns exemplos de séries temporais discretas são: valores de fechamento diário de uma determinada ação no mercado financeiro, valores diários de passageiros em uma companhia aérea, taxa de desemprego anual em um determinado país, valores diários de poluentes na atmosfera, produção mensal de soja no Brasil e quantidade de chuva em uma determinada cidade.

Uma série temporal univariada pode ser representada em um gráfico de duas dimensões, no qual o eixo  $x$  determina o momento ( $t$ ) em que os valores do eixo  $y$  foram observados. A Figura 1 mostra um exemplo de uma série temporal financeira. A série é composta por valores diários do preço de fechamento das ações da Petrobras durante o ano de 2019.

Alguns dos comportamentos encontrados nas séries temporais são sazonalidade, cíclico, tendência, e estacionariedade (BROCKWELL; DAVIS, 2016). Sazonalidade é um padrão periódico e recorrente (ZHANG; QI, 2005), diferente do comportamento cíclico, que pode acontecer em diferentes frequências. Já uma tendência é uma mudança sistemática e não periódica de crescimento ou de queda dos valores em uma série temporal (METCALFE; COWPERTWAIT, 2009). Por sua vez, estacionariedade é o comportamento no qual a série segue um equilíbrio estatístico com propriedades probabilísticas que não se alteram ao longo do tempo, seguindo média e variação constantes (BOX et al., 2015).

Figura 1 – Preço de fechamento das ações da Petrobras em 2019, obtido em *Yahoo Finance* (<http://finance.yahoo.com/quote/PETR4.SA/> ).



**Fonte:** o autor (2021)

Séries temporais em que a média se altera no decorrer do tempo, seja por resultado do comportamento de tendência ou por mudanças no fenômeno gerador da série, são chamadas de séries temporais não estacionárias. Segundo (CAO; GU, 2002), uma série não estacionária é uma série temporal em que a distribuição estatística muda no decorrer do tempo, tornando-se um dos principais problema para a atividade de previsão.

A análise de séries temporais pode ser aplicada com os seguintes objetivos: descrição, explicação, previsão e controle do comportamento estudado (CHATFIELD, 2003). A descrição consiste em pontuar as principais propriedades ou comportamento da série. Já a explicação implica obter um entendimento do mecanismo gerador da série temporal de forma a possibilitar que as variações ocorridas sejam explicadas. Por sua vez, a predição consiste em prever os novos valores da série por meio de valores passados. Finalmente, o controle está associado à tomada de decisão e pode ser desempenhado a partir do resultado da análise da série temporal.

A atividade de previsão (campo de estudo deste trabalho) tem aplicações nas mais diferentes áreas, tais como: astronomia (JUNIOR et al., 2018), mercado financeiro (LAHMIRI, 2016), meteorologia (SUN et al., 2018), sismologia (KHAN; MITTNIK, 2017), entre outras (GONZÁLEZ; ROQUE; PÉREZ, 2018; SUDHEER et al., 2014; SACCHI et al., 2007). No caso de uma série temporal  $Z$ , a atividade de previsão consiste em estimar seu valor futuro  $z_{t+h}$ , em que  $h$  é o horizonte da previsão. A atividade de previsão pode ser representada matematicamente por

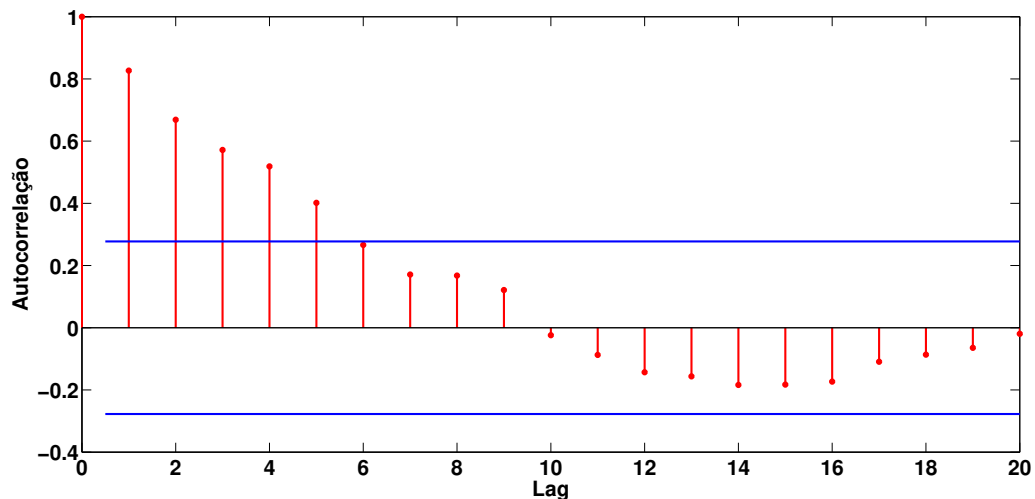


meio da Equação 2.1:

$$\hat{z}_{t+h} = f(w_t), \quad (2.1)$$

em que  $f(\cdot)$  é um modelo de previsão que recebe como entrada uma janela temporal ( $w_t$ ) e retorna o valor predito ( $\hat{z}_{t+h}$ ). Uma janela temporal é composta por valores passados da série. A representação matemática  $w_t = (z_t, z_{t-1}, z_{t-2} \dots, z_{t-n-1})$  pode ser utilizada para definir uma janela formada por  $n$  retardos temporais (*lags*) sequenciais ao ponto a ser predito. Porém, uma janela temporal pode ser formada por *lags* de qualquer instante da série temporal, de acordo com a correlação entre os *lags* e o ponto  $t$  por meio, por exemplo, da Função de Autocorreção (FAC) (BOX et al., 2015) ou algoritmos de otimização (OLIVEIRA; LUDERMIR, 2014).

Figura 2 – Correlograma dos 20 primeiros pontos da série apresentada na Figura 1.



Fonte: o autor (2021)

A figura 2 demonstra um exemplo de correlograma para a série do preço da ação da Petrobras apresentada na figura 1. Essa análise pode ser utilizada para visualizar quais *lags* têm correlação com o ponto  $t$ . No eixo x do correlograma estão os retardos (*lags*) representados no intervalo de 0 (ponto  $t$ ) ao tamanho máximo da janela ( $n$ ). No exemplo ilustrado na figura 2, a janela tem tamanho máximo igual a 20. No eixo y se encontra a taxa de correlação de cada ponto do eixo x em relação ao ponto 0, seguida do intervalo  $[+1$  e  $-1]$ , em que 1 é a correlação máxima positiva, 0 é a correlação nula e -1 é a correlação máxima negativa. As linhas em azul definem o intervalo em que a correlação não é significativa (METCALFE; COWPERTWAIT, 2009). Dessa forma, o correlograma apresentado na figura 2 permite a seguinte interpretação: em um intervalo de 20 pontos, apenas os pontos ( $z_{t-1}, z_{t-2}, z_{t-3}, z_{t-4}, z_{t-5}$ ) têm correlação

significativa com o ponto  $z_t$ , logo, estes pontos poderão ser utilizados para compor a janela  $w_t$ , que será a entrada para o modelo de previsão. Porém, outras abordagens podem ser aplicadas para seleção dos *lags* que irão compor a janela, tais como: *Wrapper* (KARASU et al., 2020); Algoritmos de Otimização (OLIVEIRA; LUDERMIR, 2014).

## 2.2 MODELOS DE PREVISÃO PARA SÉRIES TEMPORAIS

Um modelo de série temporal é utilizado para compreender o comportamento dos dados (METCALFE; COWPERTWAIT, 2009) e pode ser empregado para realizar atividades de simulação, previsão e reconhecimento de padrões (PARZEN, 1974). Para a atividade de previsão, dois tipos de modelos são frequentemente utilizados: os estatísticos e os de aprendizado de máquina.

### 2.2.1 Modelos estatísticos

Os modelos estatísticos aplicam técnicas de estimação de parâmetros, regressão e análise de ruído para mapear o comportamento da série temporal. Os modelos mais conhecidos são: o modelo Auto Regressivo (AR), o modelo Média Móveis (MA), o modelo ARMA, que é a junção dos modelos AR e o MA. O modelo Auto Regressivo Integrado de Média Móveis (ARIMA), que assim como o ARMA é uma junção de AR e MA, pode ser aplicado para séries não estacionárias, pois é adicionado de uma técnica de diferenciação da série, aplicada para transformar a série em estacionária (caso já não seja). O modelo ARIMA tem a seguinte representação matemática da Equação 2.2:

$$z'_t = \beta_1 z'_{t-1} + \beta_2 z'_{t-2} + \dots + \beta_p z'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.2)$$

em que  $\beta_1, \dots, \beta_p$  são os parâmetros do modelo AR de ordem  $p$ ,  $\theta_1, \dots, \theta_q$  são os parâmetros do modelo MA de ordem  $q$  e  $z'$  são os valores da série após aplicar a técnica de diferenciação, e  $\varepsilon_t$  é o componente aleatório. Para a modelagem de um ARIMA, utiliza-se a metodologia de Box e Jenkins (BOX et al., 2015), composta por três etapas: identificação, estimação e diagnóstico.

A etapa de identificação envolve a escolha dos valores para os hiper-parâmetros do ARIMA( $p, d, q$ ) por meio da aplicação de funções de autocorrelação (FAC) e autocorreção parcial (FACP).

Ambas as funções são utilizadas para encontrar os lags correlacionados da série, sendo que por meio da FACP é definido o parâmetro  $p$  do modelo AR, enquanto que a FAC define o parâmetro  $q$ . O parâmetro  $d$  do ARIMA é resultado do número de diferenciação que é aplicado na série até ser transformada em uma série temporal estacionária.

Na estimação, os parâmetros são estimados por meio de técnicas como mínimos quadrados ou máxima-verossimilhança (BOX et al., 2015). Por fim, é realizada a etapa de diagnóstico com o objetivo de avaliar se o modelo estimado representa o comportamento da série e se pode ser aplicado para realizar previsões. Para tal, é realizada uma análise da série de ruído do modelo, caso seja um ruído branco, o modelo é adequado para a série temporal, podendo então ser utilizado para previsão.

Modelos estatísticos lineares têm metodologias bem definidas para sua utilização, porém, segundo (ZHANG, 2003), são modelos com desempenho limitado para previsão de séries temporais do mundo real, que são compostas por comportamentos não lineares.

### 2.2.2 Modelos de Aprendizado de Máquina

Nas últimas décadas, algoritmos de aprendizado de máquina vêm sendo utilizado com frequência na atividade de previsão de séries temporais (GONZÁLEZ; ROQUE; PÉREZ, 2018; SUDHEER et al., 2014; SACCHI et al., 2007). São considerados modelos dirigidos aos dados (MITCHELL et al., 1997), uma vez que uma série temporal possibilita ao modelo aprender a dependência entre passado e futuro dos valores da série (BONTEMPI; TAIEB; BORGNE, 2012) sem precisar entender os dados nos quais o modelo será aplicado, podendo assim obter previsões com boa acurácia. Redes MLP (SAMARA; NATSHEH, 2018), SVR (EHSAN; SIMON; VENKATESWARAN, 2017) e Redes RBF (RUBIO et al., 2017) e LSTM (MA et al., 2015) são exemplos de modelos de aprendizado de máquina utilizados em séries temporais.

Entre esses modelos, a SVR tem um destaque importante, pois apresenta estabilidade e robustez (DRUCKER et al., 1997), com capacidade de realizar previsões acuradas (OLIVEIRA; LUDERMIR, 2016; SILVA et al., 2018). A característica de estabilidade da SVR está associada à natureza determinística do treinamento do modelo (DRUCKER et al., 1997), de forma que para o mesmo conjunto de dados e o mesmo conjunto de valores para os hiper-parâmetros, o algoritmo resultará em um mesmo modelo, diferentemente de modelos como a MLP, que sofrem influência da inicialização aleatória (MITCHELL et al., 1997), podendo resultar em diferentes modelos.

A SVR é um modelo proposto por (DRUCKER et al., 1997) baseado na utilização do espaço de característica em alta dimensão e que penaliza a complexidade por meio de um termo adicionado à função de erro (AHMED et al., 2010). Um modelo SVR para previsão de séries temporais pode ser representado na Equação 2.3:

$$f(x) = (\mathbf{w}^t \cdot \phi(\mathbf{x})) + b, \quad (2.3)$$

em que  $\mathbf{w}$  é o conjunto de pesos,  $b$  é o limiar,  $\phi(\cdot)$  é uma função kernel e  $\mathbf{x}$  é a janela temporal. As estimações de pesos  $\mathbf{w}^t$  e limiar  $b$  são realizadas por meio da minimização da função de risco regularizada apresentada na Equação 2.4:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l L(y_i, f(\mathbf{x})), \quad (2.4)$$

em que  $C > 0$  é o fator de regularização e  $L(\cdot, \cdot)$  é a função de custo apresentada na Equação 2.5:

$$L(y, f(\mathbf{x})) = \begin{cases} 0 & |f(\mathbf{x}) - y| < \varepsilon \\ |f(\mathbf{x}) - y| & \text{caso contrário} \end{cases}, \quad (2.5)$$

em que  $\varepsilon$  define uma margem de tolerância na qual nenhuma penalidade é aplicada.

Dessa forma, um modelo SVR pode ser definido aplicando-se a seguinte equação:

$$\text{minimizar } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \quad (2.6)$$

obedecendo as seguintes restrições:

$$(\mathbf{w}^t \cdot \phi(\mathbf{x})) + b \leq \varepsilon + \xi_i \quad (2.7)$$

$$y_i - (\mathbf{w}^t \cdot \phi(\mathbf{x})) + b \leq \varepsilon + \xi_i^* \quad (2.8)$$

$$\xi_i, \xi_i^* \geq 0, i = 0 \dots l \quad (2.9)$$

em que  $\xi_i^*$  e  $\xi_i$  são variáveis de folga utilizadas para medir o custo dos erros de previsão. Já o parâmetro  $C$  se relaciona à complexidade da função, logo, quanto maior o valor de  $C$ , maior é a complexidade, e à suavização da função, logo, quanto menor o valor de  $C$ , mais suave será a função.

A função kernel  $\phi(\cdot)$  é utilizada para que o modelo possa ser aplicado a dados não lineares. Em previsão de séries temporais, a função kernel mais utilizada é a gaussiana, também conhecida como função RBF (WU; CHENG, 2010), pois apresenta um treinamento rápido e é bastante flexível (XIN et al., 2012). A função RBF gaussiana tem a seguinte representação:

$$\phi(x) = \exp\left(-\frac{1}{2\sigma^2}|x - x_i|^2\right), \quad (2.10)$$

em que  $\sigma$  é o desvio da função gaussiana.

A SVR tem alta sensibilidade à escolha dos valores para os hiper-parâmetros, podendo acarretar na construção de um modelo que não seja capaz de realizar o mapeamento entre a janela temporal (entrada) e o valor a ser predito (saída), caracterizando um problema de *underfitting*, ou na construção de um modelo super ajustado para os dados de treinamento e sem capacidade de generalizar, caracterizando um problema de *overfitting* (WU; XIE, 2019).

## 2.3 SISTEMA DE MÚLTIPLOS PREDITORES

O Sistema de Múltiplos Preditores (SMP), também conhecido como *Ensemble*, vem sendo cada vez mais utilizado em previsão de séries temporais (ZHANG et al., 2021; NETO et al., 2020; WANG; ZHANG; LU, 2019; YAO; DAI; SONG, 2018; OLIVEIRA et al., 2017). O uso de um SMP possibilita uma maior acurácia na previsão (QIU et al., 2017) e redução da incerteza em relação à escolha de qual modelo utilizar (KOURENTZES; BARROW; CRONE, 2014). O objetivo de um SMP é obter um sistema de previsão mais robusto que preditores utilizados de forma individual (conhecidos como Monolíticos) por meio de um conjunto de preditores (conhecido como *Pool*) com diferentes características. Um SMP pode ser composto por três fases: geração, seleção e integração.

### 2.3.1 Geração

A etapa de geração tem o objetivo de criar um conjunto  $M = \{M_1, M_2 \dots, M_N\}$  com  $N > 1$  preditores, também conhecido como *Pool* de preditores. Nessa etapa, é importante que o conjunto seja composto por modelos acurados e diversos (VALENTINI; DIETTERICH, 2004). Preditores diversos são modelos que apresentam previsões diferentes para o mesmo ponto da série, sendo que é esperado que sejam especialistas em comportamentos específicos da série

temporal.

Para a criação de um *Pool* com diversidade, duas abordagens podem ser aplicadas. A primeira consiste em utilizar o mesmo modelo de aprendizado de máquina e variar o conjunto de treinamento desses modelos, criando assim um *Pool* de preditores homogêneos. Algoritmos como o *Bagging* (MORETTI et al., 2015) e o *Adaboosting* (BARROW; CRONE, 2016) são aplicados para a construção de um *Pool* homogêneo. A segunda abordagem consiste em utilizar diferentes modelos de aprendizado de máquina, criando assim um *Pool* heterogêneo. Para tal, são utilizados modelos como MLP, SVR, RBF e LSTM, treinados no mesmo conjunto de dados. Por possuírem algoritmos de treinamento diferentes, é esperado que apresentem previsões diferentes que sejam especialistas em comportamentos específicos da série temporal, resultando assim em um *Pool* com diversidade.

### 2.3.2 Seleção

A etapa de seleção tem o objetivo de selecionar um conjunto  $M' \subset M$  com um ou mais ( $n > 0$ ) preditores por meio de um critério de seleção. Caso seja selecionado apenas um preditor ( $n = 1$ ), defini-se como seleção de preditor. Caso seja selecionado mais de um preditor, define-se como seleção de *Ensemble*. A seleção pode acontecer de forma estática ou dinâmica. No caso da seleção estática, o conjunto de preditores selecionado será utilizado em todos os novos dados. Já no caso da seleção dinâmica, um conjunto de preditores será selecionado para cada novo ponto da série temporal.

Para a seleção estática, diversas abordagens são encontradas na literatura (TSOUMAKAS; PARTALAS; VLAHAVAS, 2008). Por exemplo, por meio de Ranking (ADHIKARI; VERMA; KHANDELWAL, 2015), que consiste em selecionar os  $n$  modelos com maior acurácia em um conjunto de validação, de *Meta-Learning* (LEMKE; GABRYS, 2010), que utiliza um modelo para aprender em quais características da série cada modelo tem maior acurácia, e a partir de algoritmo de otimização, que seleciona os  $n$  preditores que minimizam uma dada métrica de erro (SIN; WANG, 2017).

Para a seleção dinâmica, são utilizadas abordagens baseadas na técnica de seleção aplicada na tarefa de classificação, *Dynamic Classifier Selection by Local Accuracy* (OLA) (WOODS; KEGELMEYER; BOWYER, 1997) que foi proposta por (ROONEY et al., 2004) com o nome de *Dynamic Selection* (DS) para problemas de regressão, aqui designado para *Dynamic Selection by Local Accuracy* (DS-LA). O DS-LA já foi aplicado em vários trabalhos de previsão de séries

temporais (NADEEM et al., 2019; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; YAO; DAI; SONG, 2018). O método DS-LA consiste em selecionar  $n$  modelos com maior acurácia em uma região de competência formada por  $k$  padrões de treinamento ou validação mais similares ao novo padrão de teste. Dessa forma, o desempenho do método pode estar associado aos seus parâmetros: medida de similaridade, tamanho da região de competência e métrica de desempenho. No trabalho de (NETO et al., 2020), é apresentado o SMP *Temporal-Window Framework* (TWF), que particiona a amostra de treinamento em  $p$  partições, sendo que para cada partição, um modelo é treinado com o objetivo de criar um *Pool* de especialistas em diferentes comportamentos da série temporal. A previsão é obtida por meio do modelo respectivo à partição de treinamento que apresenta a menor distância, calculada pelo *Dynamic Time Warping* (DTW) em relação à nova janela de teste.

A etapa de seleção é considerada importante por dois motivos principais: eficiência e desempenho (TSOUMAKAS; PARTALAS; VLAHAVAS, 2008). A seleção possibilita diminuir o número de modelos utilizados, reduzindo assim o custo computacional (RAFTERY et al., 2005). Além disso, a seleção de preditores evita que preditores com baixa acurácia sejam utilizados, resultando em uma otimização da previsão do SMP (LIANG; JIA, 2018). Ainda, segundo (MA; DAI; LIU, 2015), um subconjunto de modelos selecionados  $M'$  tende a ter maior capacidade de generalização que o conjunto completo  $M$  e, conseqüentemente, a uma maior acurácia em padrões novos, pois reduz a influência de modelos que são especialistas em comportamentos do conjunto de treinamento.

### 2.3.3 Integração

A etapa de integração tem o objetivo de obter a previsão do SMP por meio de uma técnica de combinação. As técnicas podem ser treináveis ou não treináveis. As técnicas treináveis criam modelos que mapeiam o relacionamento entre as previsões dos preditores do *Ensemble* e a previsão desejada. Entre as técnicas treináveis, temos combinação linear ponderada (MENDES-MOREIRA et al., 2012), SVR e MLP (ADHIKARI; AGRAWAL, 2012), já entre as técnicas não treináveis temos média simples, média ponderada e mediana.

## 2.4 TRABALHOS RELACIONADOS

Nos anos mais recentes, o problema de seleção de modelos para previsão de séries temporais vem sendo estudado com frequência (AKÇA; YOZGATLIGIL, 2020), sendo que os principais trabalhos relacionados são apresentados nesta seção.

O trabalho de (PRUDÊNCIO; LUDERMIR, 2004) apresenta uma abordagem de seleção de preditores por meio de *Meta-Learning*. Os autores reforçam que uma forma eficiente de resolver o problema se dá por meio do conhecimento especializado em relação aos modelos e à série temporal. Para tal, a abordagem proposta recebe um conjunto de características e retorna um *ranking* dos melhores modelos para realizar a previsão da série temporal.

Por sua vez, o trabalho de (CANG; YU, 2014) descreve uma abordagem para uma seleção de *Ensemble* de preditores baseada na teoria da informação. Os resultados desse trabalho mostram que realizar a seleção de modelos antes da combinação resulta em um ganho de acurácia na previsão. Outra abordagem para seleção de *Ensemble* é proposta em (MA; DAI; LIU, 2015), consistindo na composição de preditores do *Ensemble* com base na acurácia da sua combinação, sendo selecionado o conjunto de preditores que apresente o menor erro de previsão em uma região de competência. O trabalho reforça a importância de se realizar a seleção de preditores anteriormente à combinação, além de destacar que a seleção possibilita reduzir o custo computacional da utilização de *Ensemble*. O trabalho de (MA; DAI, 2016) realiza seleção de *Ensemble* baseado no desempenho dos preditores em uma região de seleção, além disso, a abordagem consiste em combinar as previsões dos modelos selecionados por meio de um modelo de aprendizado de máquina. Os resultados obtidos reforçam a importância de se selecionar os preditores antes de se aplicar a combinação.

Dentre esses trabalhos, é possível destacar que as abordagens se utilizam de uma região de competência, ou de características da série temporal, para avaliar um conjunto  $M$  de preditores e selecionar um ou mais preditores. Os resultados apresentados mostram que a seleção é uma etapa importante na acurácia do *Ensemble* para previsão de séries temporais. Apesar desses trabalhos apresentarem diferentes propostas para a seleção de preditores, as abordagens são aplicadas no contexto de seleção estática. Porém, é sabido que as séries temporais podem apresentar um comportamento dinâmico e, por essa razão, a escolha do modelo de previsão deve ser dinâmica.

No contexto de seleção dinâmica de preditores, poucas abordagens são encontradas na literatura. O trabalho de (YANKOV; DECOSTE; KEOGH, 2006) apresenta uma abordagem de



seleção de preditor por meio de medida de similaridade e acurácia dos preditores em um conjunto de validação. A abordagem é comparada à seleção estática e os autores mostram que o uso de seleção dinâmica possibilita aumentar a capacidade de generalização do *Ensemble*.

O trabalho de (WIDODO; BUDI, 2011) segue a hipótese de que o modelo com maior acurácia em prever dados de treinamento e validação similar aos da região do problema será o modelo ideal para prever o próximo ponto da série temporal. Para tal, é apresentada uma abordagem de seleção dinâmica que consiste em construir uma região de competência utilizando distância Euclidiana ou DTW para selecionar as janelas de validação mais similares à nova janela de teste. As previsões dos preditores selecionados são combinadas por média simples e ponderada. Os resultados apresentados mostram que essa abordagem de seleção possibilita obter previsões combinadas mais acuradas. O trabalho de (BUDI; AJI; WIDODO, 2014) apresenta a mesma abordagem de seleção dinâmica para um conjunto de séries temporais diferentes do trabalho anterior.

O trabalho de (KRIKUNOV; KOVALCHUK, 2015) mostra que abordagens para seleção dinâmica de classificadores não são aplicáveis para seleção dinâmica de preditores, pois são problemas com características diferentes. A partir dessa motivação, os autores apresentam uma abordagem de seleção dinâmica de preditores baseada na correlação entre a previsão do erro de cada composição de *Ensemble* e o valor atual da série temporal. Os resultados apresentados no trabalho mostram um pequeno ganho de acurácia em comparação à não utilização de seleção, porém, os autores argumentam que o ganho pode aumentar com a utilização de medidas de distância na abordagem de seleção.

Por sua vez, o trabalho de (YAO; DAI; SONG, 2018) realiza um estudo com diferentes abordagens para o desenvolvimento da região de competência com o objetivo de selecionar preditores em uma região de competência similar à nova janela de teste. A partir dos resultados obtidos, os autores reforçam a importância da seleção dinâmica e destacam que dentre as abordagens avaliadas, a seleção de preditores em uma região de competência selecionada por meio de distância Euclidiana pode apresentar boa acurácia, além de ser uma abordagem simples.

Já o trabalho de (NETO et al., 2020) mostra que desenvolver modelos especialistas em partições da série e depois selecionar de forma dinâmica cada preditor de acordo com a similaridade entre as partições e a nova janela de teste é uma estratégia factível, pois é capaz de selecionar preditores mais acurados que outras abordagens presentes na literatura.

Em relação aos principais trabalhos para a seleção dinâmica, é possível destacar que as

abordagens apresentadas em (YANKOV; DECOSTE; KEOGH, 2006; ??; YAO; DAI; SONG, 2018) têm propostas semelhantes, pois selecionam os preditores a partir de uma região de competência formada por janelas de treinamento ou validação que foram selecionadas com base na similaridade com a nova janela de teste. Esse tipo de abordagem de seleção foi inicialmente apresentado no contexto de classificação no trabalho de (WOODS; KEGELMEYER; BOWYER, 1997) e então adaptada ao problema de regressão por (ROONEY et al., 2004), a qual os autores nomearam de *Dynamic Selection* (DS). Neste trabalho, designaremos tal abordagem como *Dynamic Selection by Local Accuracy* (DS-LA), pois melhor descreve o método.

Em resumo, os trabalhos encontrados na literatura relacionados à pesquisa demonstram a importância da seleção de preditores. Além disso, é possível destacar que as abordagens mais indicadas para realizar a previsão para a nova observação de testes são as propostas para a seleção dinâmica que seguem a hipótese de que os preditores que apresentam a maior acurácia em padrões de treinamento ou validação similares à janela de teste. Porém, é possível que tais abordagens não sejam capazes de selecionar os preditores mais acurados, pois não têm como garantir a existência de padrões no conjunto de treinamento com comportamentos similares ao padrão de teste devido ao ruído nos dados, ao tamanho da amostra ou à ausência de tais padrões (CRUZ; SABOURIN; CAVALCANTI, 2018). Além disso, a distribuição dos dados pode mudar no decorrer do tempo (ELWELL; POLIKAR, 2011; ADHIKARI, 2015), consequentemente, o melhor modelo para prever a nova observação também pode mudar (MEADE, 2000).

## 2.5 CONSIDERAÇÕES DO CAPÍTULO

Neste segundo capítulo foram definidos alguns conceitos e características a respeito da atividade de previsão de séries temporais, por exemplo: tendência, sazonalidade, séries estacionárias e retardos temporais relevantes (*Lags*). Também foram abordados diferentes tipos de modelos de previsão de séries temporais. O modelo estatístico ARIMA e os modelos de aprendizado de máquina, com ênfase no SVR, se destaca por ser considerado um modelo com estabilidade e capacidade de realizar previsões acuradas.

Ainda neste capítulo, foram apresentadas as três etapas para construção de um Sistema de Múltiplos Preditores, com destaque para a etapa de seleção, que tem o papel de reduzir o número de modelos utilizados, podendo aumentar a acurácia da previsão do SMP e reduzir o custo computacional. E, por fim, são apresentados os principais trabalhos da literatura que propõem ou utilizam abordagens de seleção. Os trabalhos podem ser organizados de acordo

com o tipo de abordagem, estática e dinâmica, sendo que aqueles para a seleção estática mostram a importância de se realizar a seleção, porém, não se aplicam no contexto de seleção dinâmica. Já os trabalhos de seleção dinâmica reforçam a importância de se aplicar a seleção dinâmica, porém, a maioria deles utiliza como base uma abordagem que envolve selecionar o preditor em uma região de competência formada por janelas de treinamento ou validação que são mais similares à nova janela de teste. Outra limitação dos trabalhos apresentados é a não realização de uma análise dos parâmetros responsáveis para a seleção do preditor.

### 3 MÉTODO PROPOSTO

Este capítulo apresenta o problema para definir regiões de competência e as etapas do método proposto. O capítulo está dividido nas seguintes seções: definição do problema 3.1, arquitetura proposta 3.2 e considerações do capítulo 3.3.

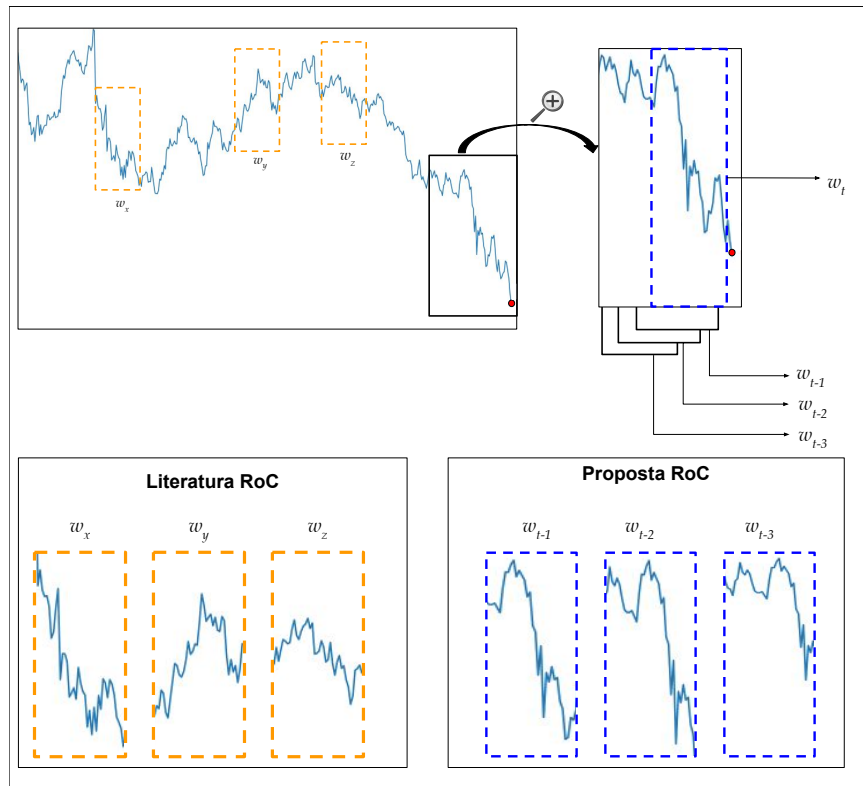
#### 3.1 DEFINIÇÃO DO PROBLEMA

Abordagens de seleção dinâmica para previsão de séries temporais têm sido aplicadas com o objetivo de melhorar a precisão do Sistema de Múltiplos Preditores (SMP) (NADEEM et al., 2019; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; YAO; DAI; SONG, 2018). Essas abordagens são baseadas em (ROONEY et al., 2004), que propôs a utilização da seleção dinâmica para a regressão por meio da *Dynamic Selection* (DS) (neste trabalho renomeada de *Dynamic Selection by Local Accuracy* (DS-LA) para evitar conflito com o nome da área de seleção dinâmica). No contexto de previsão de séries temporais, as abordagens de seleção dinâmica supõem que o modelo mais acurado para prever a próxima observação ( $z_{t+1}$ ) dado uma janela temporal  $w_t = \{z_t, z_{t-1}, z_{t-2}, \dots, z_{t-n-1}\}$  com  $n$  lags é o modelo com maior acurácia na Região de Competência (RoC) formada apenas por  $k$  janelas de treinamento ou validação similares a nova janela de teste  $w_t$ .

A construção da RoC desempenha um papel importante nas abordagens de seleção dinâmica, uma vez que sua qualidade é um fator que influencia a seleção dos preditores mais acurados. Da mesma forma que para os problemas de classificação não é possível garantir que a região de competência seja composta por padrões realmente semelhantes ao da amostra de teste, pois depende da distribuição dos dados utilizados (CRUZ; SABOURIN; CAVALCANTI, 2018). Além disso, a escolha da medida de similaridade mais adequada para definir a região de competência depende da série temporal (SILVA et al., 2020).

A Figura 3 mostra uma nova janela temporal  $w_t$  e uma RoC formada por três janelas de treinamento  $w_x$ ,  $w_y$ , e  $w_z$ . Essa RoC foi definida utilizando o pressuposto adotado pela literatura, por isso, foi nomeada como "Literatura RoC". Por meio de uma medida de similaridade, essa abordagem busca no conjunto de treinamento e/ou validação por janelas que sejam similares à nova janela de teste  $w_t$ . Porém, é possível que essa estratégia não seja adequada à previsão de séries temporais, pois pode selecionar modelo com baixa acurácia para prever a

Figura 3 – A figura ilustra duas regiões de competência: "Literatura RoC", formada por janelas de treinamento mais similares a  $w_t$ , "Proposta RoC", região de competência formada por janelas que antecedem  $w_t$ .



Fonte: o autor (2021)

nova observação. Em contraste, temos a suposição de que selecionar as janelas temporais que antecedem a nova janela de teste pode aumentar a chance de selecionar preditores acurados para prever a próxima observação. Essa nova região de competência é formada por  $w_{t-1}$ ,  $w_{t-2}$ , e  $w_{t-3}$ , como apresentado na Figura 3 no quadro "Proposta Roc".

Temos como hipótese que as  $k$  janelas  $(w_{t-1}, w_{t-2}, \dots, w_{t-k})$  que antecedem  $w_t$  e a nova janela  $w_t$  compartilham de uma mesma distribuição. Assim, é esperado que o modelo mais acurado para prever as  $k$  janelas seja um forte candidato para ser o modelo com maior acurácia ao prever  $z_{t+1}$ . Uma vez que o comportamento da série temporal pode ser descrita por meio de uma Função Densidade de Probabilidade (FDP) em relação ao tempo (KOWALSKI et al., 2012), definida como  $F(w_t)$  e que descreve a probabilidade de uma observação da série temporal cair em um intervalo de valores. De forma que qualquer alteração em  $F(w_t)$  influencie em  $F(w_{t+1}|w_t)$ , em que  $w_t$  e  $w_{t+1}$  tendem a ser descritas por uma mesma distribuição. Além disso, o modelo candidato pode mudar para cada nova janela de teste, pois é esperada uma mudança de distribuição no decorrer do tempo em series temporais do mundo real (ELWELL; POLIKAR, 2011).

Foi realizado um experimento para avaliar a suposição de que uma região de competência formada por janelas que antecedem a nova janela de teste tem uma maior ocorrência de janelas com a mesma distribuição em comparação com a região de competência formada por janelas selecionadas por meio de uma medida de similaridade. Para tal, foram utilizadas 10 séries temporais com diferentes comportamentos (Descritas na Seção 4.1). Cada série temporal foi dividida em três conjuntos: as primeiras 50% observações da série temporal para treinamento, as próximas 25% para validação e as últimas 25% para o conjunto de teste. Para evitar o desenvolvimento de uma abordagem enviesada, nesse experimento foram consideradas apenas as amostras de treinamento e validação e renomeamos para *in-sample* e *out-of-sample*, respectivamente. As amostras foram organizadas em janelas temporais deslizantes ( $w_t$ ) contendo 20 observações cada, dessa forma, em uma amostra com 100 observações, 80 janelas são construídas. Importante destacar que entre duas janelas vizinhas ( $w_t$  e  $w_{t-1}$ ), a diferença é apenas uma observação.

Para cada janela  $w_t$  do *out-of-sample* são construídas duas regiões de competência  $\Delta_t^{LT}$  e  $\Delta_t^{PR}$ , cada uma composta por 10 janelas ( $k = 10$ ).  $\Delta_t^{LT} = (w_{m1}, w_{n2}, \dots, w_{ik})$  representa a RoC que segue a suposição da literatura (NADEEM et al., 2019; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; YAO; DAI; SONG, 2018) e é gerada por  $k$  janelas de *in-sample* que são mais similares (mensuradas por meio da distância Euclidiana (YOSHIDA; CHAKRABORTY, 2015)) em relação a  $w_t$ . A região de competência  $\Delta_t^{PR} = (w_{t-1}, \dots, w_{t-k})$  representa a RoC composta por  $k$  janelas que antecedem  $w_t$  seguindo a suposição proposta neste trabalho.

O teste estatístico *Kolmogorov Smirnov Two Sample* (KS2) (HONG-ZHI; BING, 1991) foi utilizado para realizar a comparação de distribuição. O teste KS2 é utilizado para indentificar padrões similares nos dados e detectar diferenças em relação à posição, à dispersão ou ao formato da distribuição de duas amostras (MORA-LÓPEZ; MORA, 2015). Esse teste é apropriado para amostras que variam de tamanho entre 10 e 50 observações (LILLIEFORS, 1969). A hipótese nula indica que as amostras têm a mesma distribuição, independentemente de qual distribuição. O teste KS2 foi aplicado para comparar  $w_t$  com cada janela nas diferentes RoC  $\Delta_t^{LT}$  e  $\Delta_t^{PR}$ , separadamente. Por exemplo, para uma janela do *out-of-sample*  $w_t$  e a uma região de competência  $\Delta_t^{LT}$ ,  $w_t$  é comparado a cada janela  $w_j$  de  $\Delta_t^{LT}$ , e é calculada a média de janelas que têm a mesma distribuição de  $w_t$  pela Equação 3.1:

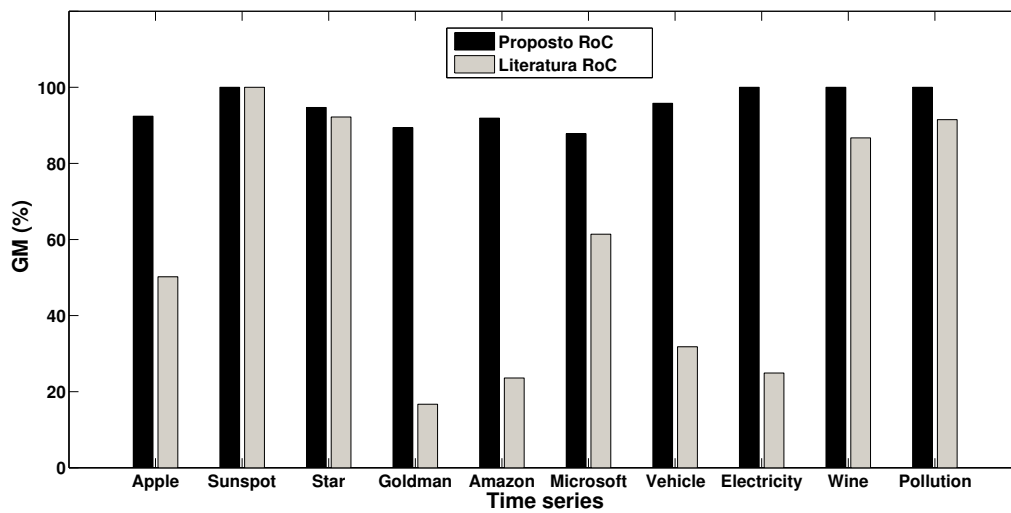
$$Q_t(w_t, \Delta_t) = \frac{1}{k} \sum_{w_j \in \Delta_t} KS2(w_t, w_j), \quad (3.1)$$

em que  $\Delta_t$  pode ser  $\Delta_t^{PR}$  ou  $\Delta_t^{LT}$  e a função  $KS2(w_t, w_j)$  retorna 1 se  $w_t$  e  $w_j$  têm a mesma distribuição, caso contrário, o valor é 0. Aplicando a Equação 3.1 para todas as janelas  $w_t$  do *out-of-sample*, o resultado será um vetor  $Q$  com  $n$  médias para cada abordagem ( $\Delta_t^{PR}$  ou  $\Delta_t^{LT}$ ). Por fim, é calculado o percentual médio para cada abordagem por meio da Equação 3.2:

$$GM(Q) = \frac{100}{n} \sum_{t \in n} Q_t, \quad (3.2)$$

em que *Grand Mean* ( $GM$ ) é a média das médias em percentual.

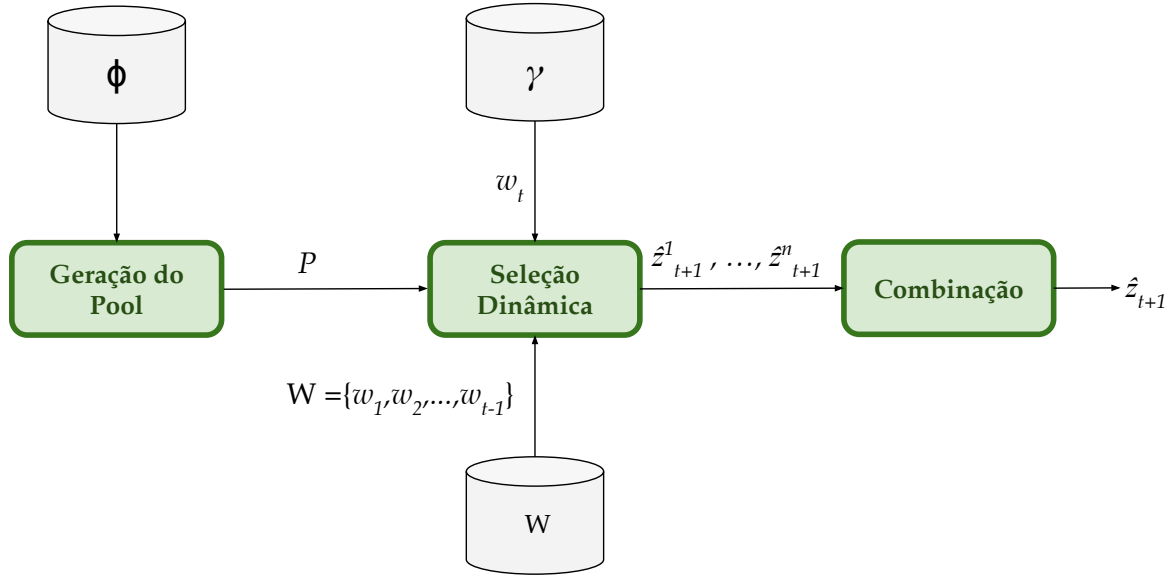
Figura 4 – Percentual médio de janelas em cada RoC que apresenta a mesma distribuição das respectivas janelas do *out-of-sample*. As barras em cinza e preto corresponde as abordagens "Literatura RoC" e "Proposta RoC", respectivamente.



Fonte: o autor (2021)

Figura 4 mostra o resultado obtido, em que o número de janelas em cada abordagem com a mesma distribuição de  $w_t$  é diretamente proporcional ao valor de  $GM$ . Para a maioria das séries temporais avaliadas, as RoC formadas pela suposição proposta apresentaram melhores resultados. Em outras palavras, a RoC formada por janelas que antecedem  $w_t$  foi capaz de selecionar mais janelas com a mesma distribuição das janelas do *out-of-sample* que a suposição da literatura. Além de selecionar janelas mais promissoras que as presentes na literatura, a proposta tem um menor custo computacional, pois é uma busca mais direta que não precisar percorrer todas as janelas do *in-sample* e calcular a similaridade.

Figura 5 – Fluxograma do SMP proposto utilizando seleção dinâmica.  $\phi$  é o conjunto de treinamento;  $\gamma$  é o conjunto de teste;  $P$  é o *Pool* de preditores;  $w_t$  é a nova janela de teste  $\gamma$ ;  $W$  é o histórico de janelas;  $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$  são as previsões dos preditores selecionados pelo DSNaw, e  $\hat{z}_{t+1}$  é a previsão final do SMP.



Fonte: o autor (2021)

## 3.2 ARQUITETURA PROPOSTA

Figura 5 apresenta o Sistema de Múltiplo Preditores proposto, composto por três fases: geração, seleção dinâmica e integração. A primeira fase gera um *Pool* de preditores ( $P$ ) treinados utilizando o conjunto de treinamento ( $\phi$ ). Em seguida, a fase de seleção dinâmica escolhe um subconjunto de modelos ( $P' \subset P$ ) para cada nova janela de teste ( $w_t \in \gamma$ ). Se mais de um modelo for selecionado, então as previsões  $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$  dos  $n$  preditores serão combinadas na última fase; por outro lado, se apenas um preditor for selecionado  $n = 1$ , não será necessária a utilização da combinação. Essas três fases são detalhadas nas sub-seções a seguir.

### 3.2.1 Geração

O objetivo da fase de geração é criar um *Pool* de preditores diversos que sejam capazes de modelar diferentes comportamentos da série temporal. Para tal, o *Pool* pode ser gerado utilizando duas estratégias. A primeira consiste em utilizar o mesmo modelo de aprendizado de máquina e variar o conjunto de treinamento desses modelos, criando assim um *Pool* de preditores homogêneos. Algoritmos como o *Bagging* (MORETTI et al., 2015) e o *Adaboosting* (BAR-



ROW; CRONE, 2016) são aplicados para a construção de um *Pool* homogêneo. A segunda abordagem consiste em utilizar diferentes modelos de aprendizado de máquina, criando assim um *Pool* heterogêneo, por meio de modelos como MLP, SVR, RBF e LSTM, treinados no mesmo conjunto de dados. Por apresentarem diferentes algoritmos de treinamento, é esperado que apresentem diferentes previsões e que sejam especialistas em comportamentos específicos da série temporal, resultando assim em um *Pool* com diversidade.

Neste trabalho utilizamos a alternativa de gerar um *pool* homogêneo, em que todos os preditores utilizam o mesmo algoritmo de aprendizado de máquina, e cada um é treinado em uma amostra diferente do conjunto de treinamento que é re-amostrado por meio do algoritmo *Bagging*, que realiza uma re-amostragem aleatória e com reposição dos padrões de treinamento. Dessa forma, a diversidade é obtida por meio dos modelos treinados em diferentes comportamentos da série temporal. Importante destacar que para que seja aplicado o *Bagging* na série temporal, é preciso que a série seja antes organizada no seguinte padrão  $\{w_t, z_{t+h}\}$ , em que  $w_t$  é a janela temporal (entrada para o modelo) e  $z_{t+h}$  é a observação a ser predita (saída do modelo), dessa forma é possível manter a relação temporal.

### 3.2.2 Seleção Dinâmica

A fase de seleção dinâmica consiste em selecionar um preditor ou um *Ensemble* (conjunto com  $n > 1$  preditores) para cada nova janela de teste ( $w_t$ ). Algoritmo 1 e Figura 6 mostram as etapas da abordagem de seleção dinâmica proposta: *Dynamic Selection from Nearest Antecedent Windows* (DSNAW).

A primeira etapa do DSNAW é a definição da região de competência  $\Delta_t^{PW}$ , composta por  $k$  janelas  $\Delta_t^{PW} = (w_{t-1}, \dots, w_{t-k})$  que antecedem  $w_t$  (Módulo "Definição Região de Competência" na Figura 6 e linha 4 no Algoritmo 1). A estratégia é selecionar as janelas que antecedem a nova janela de teste, seguindo a suposição proposta na Seção 3.1, a qual mostra que por meio dessa estratégia, a RoC tem maior chance de ser composta por janelas com a mesma distribuição de  $w_t$  e, conseqüentemente, de selecionar preditores mais acurados para o novo padrão em comparação à estratégia do DS-LA, que utiliza de medida de similaridade para formar a RoC.

Na segunda etapa (Módulo "Seleção dos Modelos" na Figura 6 e linhas 5 - 12 no Algoritmo 1), todos os modelos são avaliados na RoC. Cada modelo  $p \in P$  é utilizado para prever as observações da região de competência  $\Delta_t^{PW}$ . A acurácia do modelo é calculada por meio da

---

**Algoritmo 1:** Dynamic Selection from Nearest Antecedent Windows
 

---

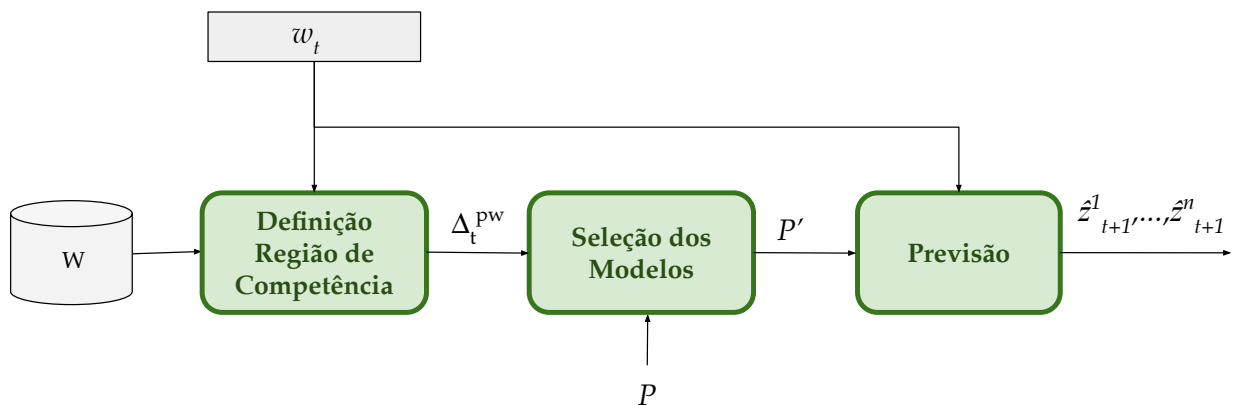
```

1 // Input: Histórico de janelas temporais ( $W$ ), nova janela de teste ( $w_t$ ), Pool de
  preditores ( $P$ ), tamanho da RoC ( $k$ ), quantidade de modelos selecionados ( $n$ )
  Input:  $W, P, k, n$ 
  Output:  $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$ 
2 // Definição da RoC composta por  $k$  janelas que antecedem  $w_t$ 
3  $\Delta_t^{PW} = \text{region\_of\_competence\_definition}(W, k)$ 
4 // Avaliação do desempenho de cada modelo do Pool  $P$ 
5  $errors = []$ 
6 for each predictor  $p \in P$  do
7    $\hat{y} = p(\Delta_t^{PW})$ 
8    $e = \text{error\_measure}(\text{target}, \hat{y})$ 
9    $errors = errors \cup \{e\}$ 
10 // Seleção dos  $n$  preditores com maiores desempenho
11  $ranking = \text{ascending\_order}(errors)$ 
12  $P' = \text{select\_best\_models}(ranking, n)$ 
13 // Obtenção da previsão dos  $n$  preditores selecionados  $w_t$ 
14  $forecast = []$ 
15 for  $i = 1$  to  $n$  do
16    $p = P'(i)$ 
17    $\hat{z}_{t+1}^i = p(w_t)$ 
18    $forecast = forecast \cup \{\hat{z}_{t+1}^i\}$ 
19 return  $forecast$ 

```

---

Figura 6 – Fluxograma da fase de seleção dinâmica utilizando a abordagem proposta DSNAW.  $W$  é o histórico de janelas, que vai sendo atualizado de acordo com as novas observações da série temporal;  $\Delta_t^{PW}$  é a região de competência;  $P$  é o Pool de preditores;  $P'$  é conjunto de modelos selecionados e  $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$  são as previsões obtidas de cada preditor selecionado.



Fonte: o autor (2021)

métrica *Sum of Absolute Errors* (SAE) (HU; LAM; NG, 2001), que foi escolhida por apresentar robustez e confiabilidade (SUN et al., 2008). Um *Ranking* em ordem crescente é criado levando em consideração o valor SAE de cada modelo (Linha 11 no Algoritmo 1), de forma que quanto

menor o valor SAE, maior é a acurácia do modelo. Ao término dessa etapa, são retornados os  $n$  primeiros preditores no *Ranking* (Linha 12 no Algoritmo 1). No caso de  $n = 1$ , temos a abordagem aplicada para seleção dinâmica de preditores, já no caso de  $n > 1$ , temos a abordagem aplicada para a seleção dinâmica de *Ensemble*.

Na última etapa (Módulo "Previsão" na Figura 6 e linhas 14 - 18 no Algoritmo 1), cada modelo selecionado ( $p \in P'$ ) recebe como entrada a nova janela de teste  $w_t$ , resultando em  $n$  previsões  $\hat{z}_{t+1}^1, \dots, \hat{z}_{t+1}^n$ .

### 3.2.3 Integração

Na última fase, a previsão final  $\hat{z}_{t+1}$  é obtida. Quando a fase anterior seleciona um *Ensemble* ( $n > 1$ ), a previsão final será resultado da combinação de todas as saídas do *Ensemble* selecionado. Essa combinação pode ser realizada a partir de diferentes estratégias, como, por exemplo, a média ou mediana, que são abordagens simples de combinação, mas que tende a apresentar os melhores resultados (KOURENTZES; BARROW; CRONE, 2014; ADHIKARI, 2015). Se apenas um preditor é selecionado ( $n = 1$ ), a previsão final será a previsão do modelo escolhido.

## 3.3 CONSIDERAÇÕES DO CAPÍTULO

Este capítulo apresentou o resultado de um experimento que compara duas abordagens de construção de região de competência. A primeira abordagem cria a região de competência com base nos trabalhos da literatura (NADEEM et al., 2019; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; YAO; DAI; SONG, 2018) e consiste em selecionar as  $k$  janelas de treinamento ou validação mais similares ao novo padrão. A segunda abordagem consiste em utilizar as  $k$  janelas que antecedem o novo padrão. O resultado desse experimento mostra que a segunda abordagem tende a apresentar uma maior ocorrência de janelas na região de competência com a mesma distribuição do novo padrão.

Em seguida foi apresentada a arquitetura proposta composta por três fases: geração, seleção dinâmica e integração. Na etapa de geração, é importante que seja criado um conjunto de preditores com previsões acuradas e que apresentem diversidade. Na etapa de seleção, é apresentado o método DSNAW, que consiste em selecionar os  $n$  preditores mais acurados na região de competência formada por  $k$  padrões que antecedem o novo padrão de teste. Por

fim, é realizada a combinação das previsões obtidas na etapa anterior. O código do DSNAW está disponível no github (<https://github.com/EraylsonGaldino/DSNAW>).

## 4 METODOLOGIA DOS EXPERIMENTOS

Neste capítulo são apresentadas as bases de dados, medidas de desempenho, teste estatístico e as abordagens da literatura utilizadas para realizar os experimentos da pesquisa.

### 4.1 BASE DE DADOS

Os experimentos foram realizados em 10 séries temporais do mundo real: Amazon, Apple, Electricity, Goldman Sachs, Microsoft, Pollution, Star, Sunspot, Vehicle e Wine<sup>1</sup>. Essas séries foram selecionadas por serem frequentemente utilizadas na literatura e apresentarem diferentes comportamentos, tais como sazonalidade, tendência, estacionariedade e não estacionariedade (WANG et al., 2018), comportamentos avaliados neste trabalho por meio do correlograma e do teste *Dickey-Fuller* (DOLADO; GONZALO; MAYORAL, 2002).

Em cada série temporal foi aplicado apenas a normalizando, transformando as observações para o intervalo  $[0, +1]$  e particionando em duas amostras: treinamento e teste. O conjunto de treinamento é composto pelos primeiros 75% pontos da série temporal e utilizado para treinamento e validação dos modelos. O conjunto de teste é composto pelos últimos 25% pontos da série temporal e empregado para avaliar as abordagens em dados ainda não apresentados aos preditores. Ambas as amostras são organizadas em janelas temporais com o máximo de 20 *lags* selecionados por meio da função de auto-correlação (ACF) (BOX et al., 2015). Esses parâmetros foram selecionados com base em trabalhos da literatura (SILVA et al., 2018; SILVA et al., 2020). Para as séries temporais que apresentam comportamentos não estacionários ou sazonalidades, não foi aplicado nenhum tratamento, utilizando apenas a série normalizada.

#### 4.1.1 Amazon

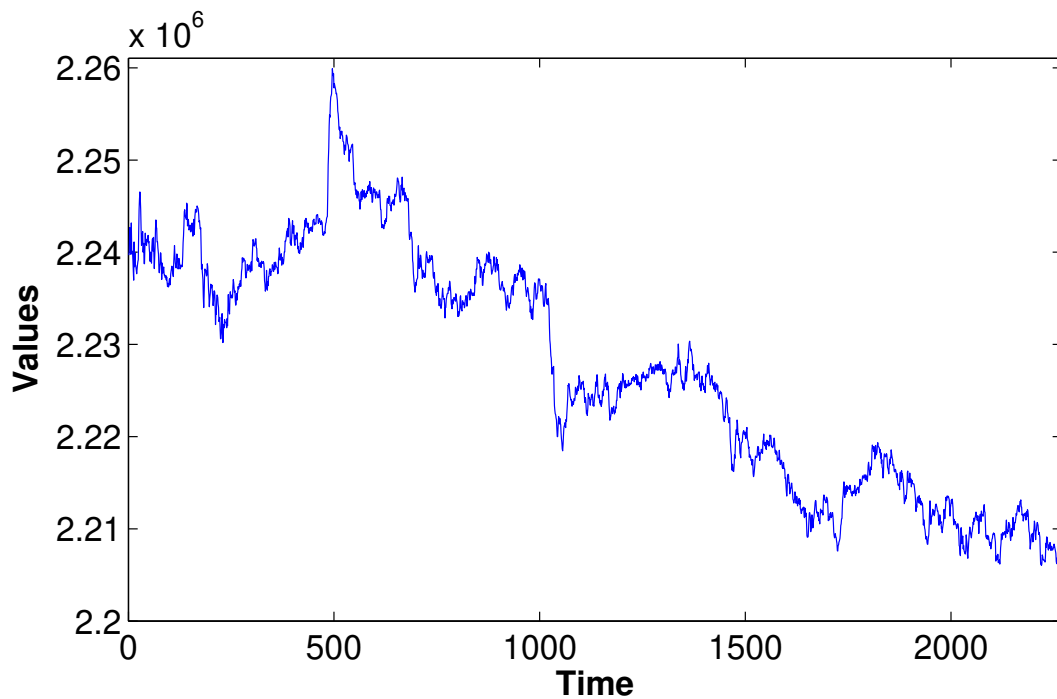
A *Amazon* é uma série do mercado financeiro composta por 2261 observações em nanosegundos. Cada observação é o valor de cotação da ação AMZN<sup>2</sup> na bolsa Nasdaq. A Figura 7 mostra o comportamento de toda série temporal, que apresenta não estacionariedade e tendência de queda.

---

<sup>1</sup> [https://github.com/EraylsonGaldino/dataset\\_time\\_series](https://github.com/EraylsonGaldino/dataset_time_series)

<sup>2</sup> <https://www.nasdaq.com/market-activity/stocks/amzn>

Figura 7 – Série Amazon



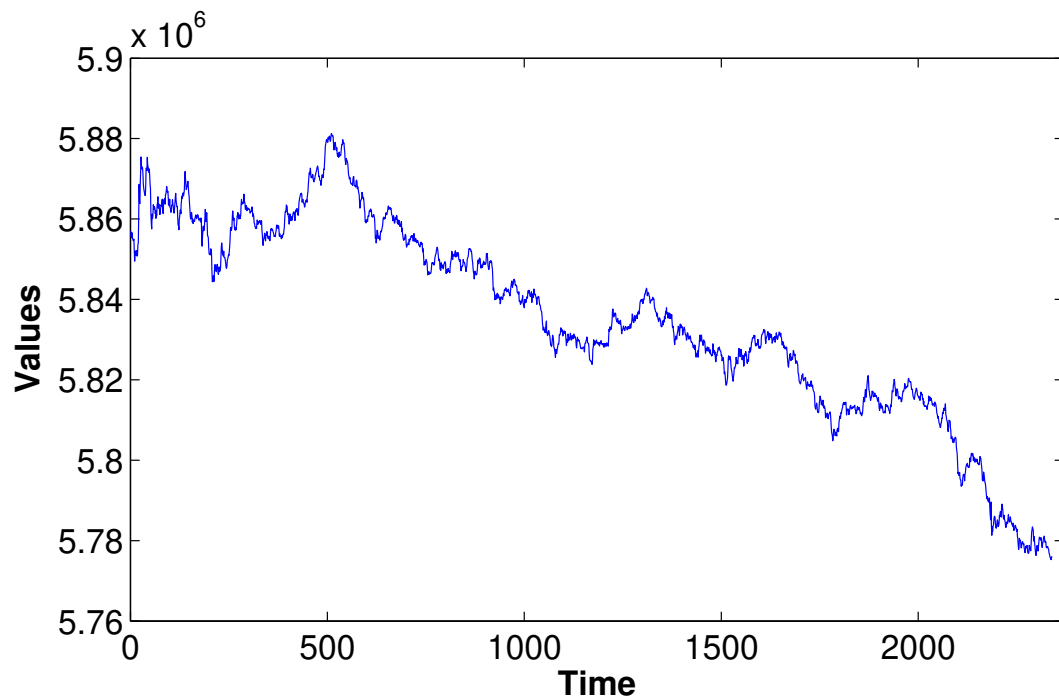
Fonte: o autor (2021)

#### 4.1.2 Apple

A *Apple* é uma série do mercado financeiro composta por 2340 observações em nano-segundos. Cada observação é o valor de cotação da ação AAPL<sup>3</sup> na bolsa Nasdaq. A Figura 8 mostra o comportamento de toda série temporal, que apresenta não estacionariedade e tendência de queda.

<sup>3</sup> <https://www.nasdaq.com/market-activity/stocks/aapl>

Figura 8 – Série Apple

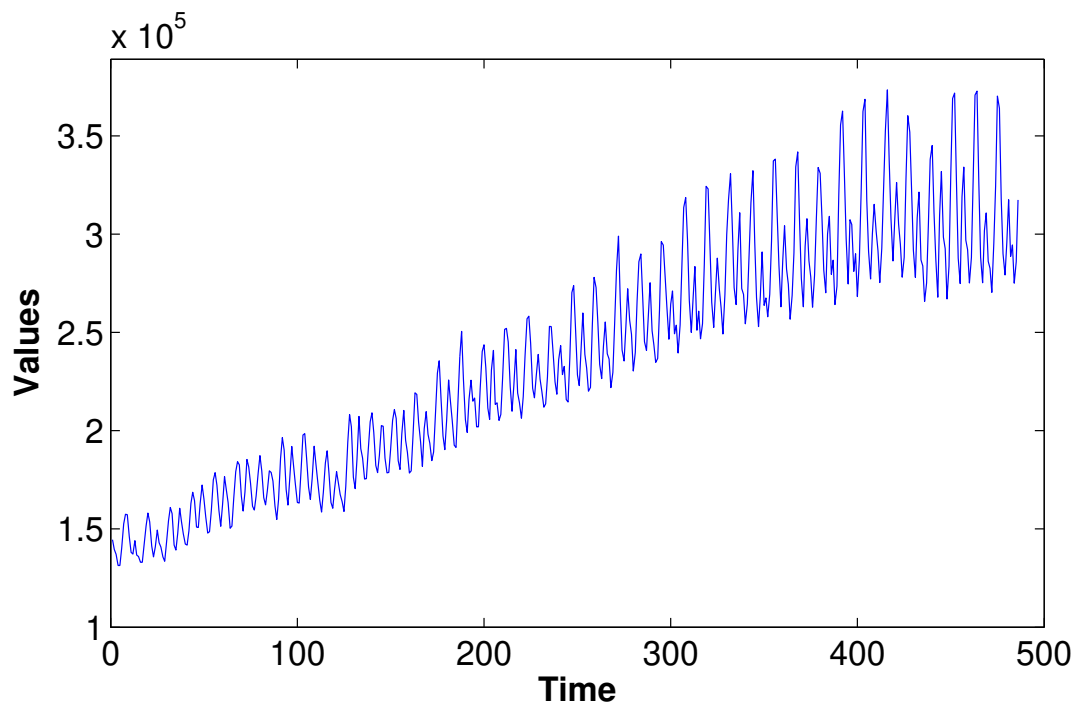


Fonte: o autor (2021)

#### 4.1.3 Electricity

A *Electricity* é uma série de produção de energia na Austrália composta por 486 observações mensais. A Figura 9 mostra o comportamento de toda a série temporal, que apresenta uma tendência de alta e sazonalidade.

Figura 9 – Série Electricity



Fonte: o autor (2021)

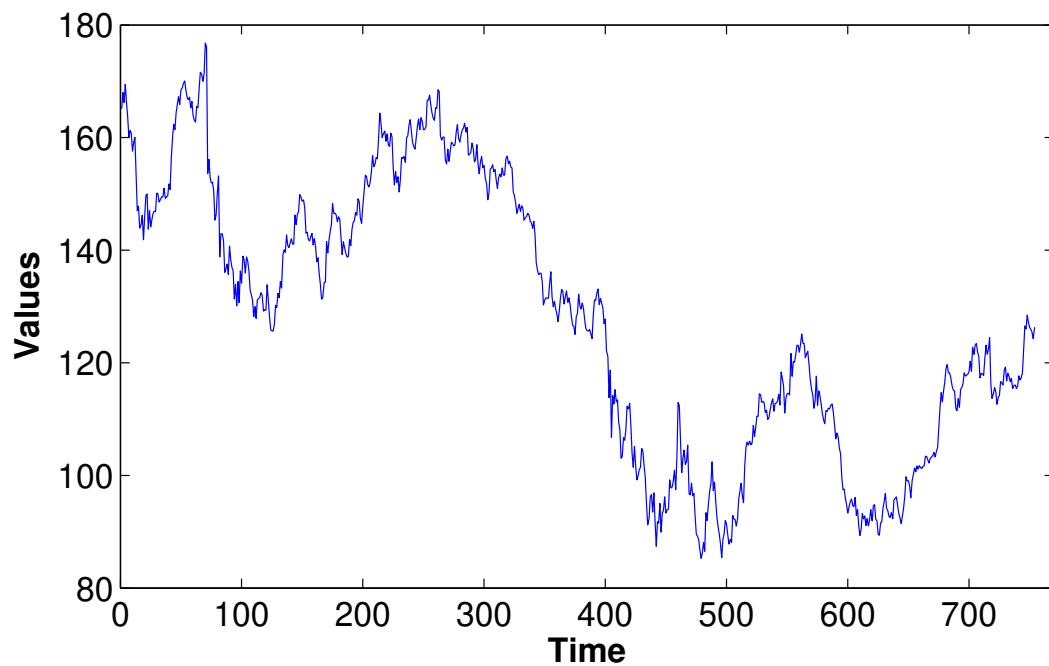
#### 4.1.4 Goldman Sachs

A *Goldman Sachs* é uma série do mercado financeiro composta por 754 observações diárias. Cada observação é o valor de cotação da ação GS<sup>4</sup> na bolsa *New York*. A Figura 10 mostra o comportamento de toda a série temporal, que apresenta não estacionariedade.

<sup>4</sup> <https://www.nyse.com/quote/XNYS:GS>



Figura 10 – Série Goldman Sachs



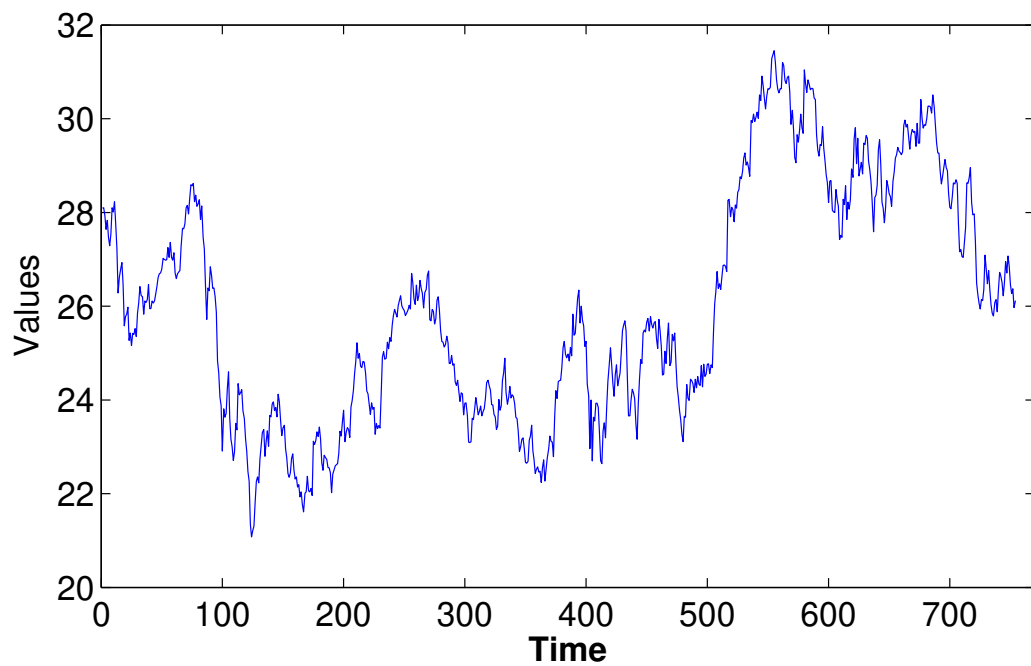
Fonte: o autor (2021)

#### 4.1.5 Microsoft

A *Microsoft* é uma série do mercado financeiro composta por 754 observações diárias. Cada observação é o valor de cotação da ação MSFT<sup>5</sup> na bolsa Nasdaq. A Figura 11 mostra o comportamento de toda a série temporal, que apresenta não estacionariedade.

<sup>5</sup> <https://www.nasdaq.com/market-activity/stocks/msft>

Figura 11 – Série Microsoft

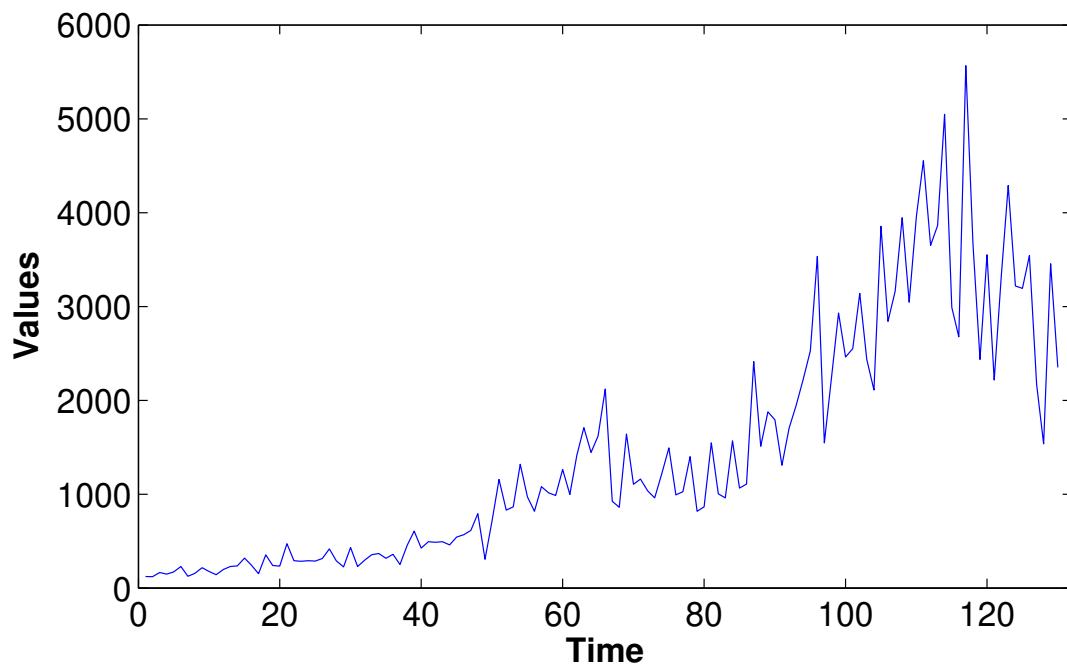


Fonte: o autor (2021)

#### 4.1.6 Pollution

A *Pollution* é uma série da quantidade de equipamentos poluentes na França e é composta por 130 observações mensais. A Figura 12 mostra o comportamento de toda a série temporal, que apresenta não estacionariedade.

Figura 12 – Série Pollution

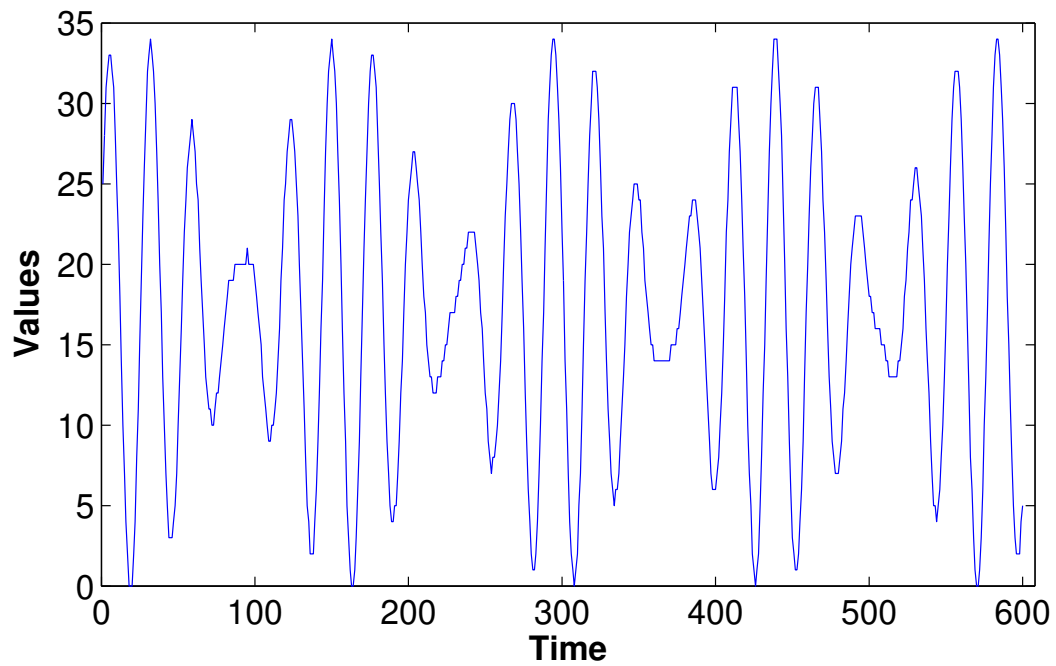


Fonte: o autor (2021)

#### 4.1.7 Star

A *Star* é uma série temporal da intensidade de brilho de uma estrela observada durante a meia-noite por 600 noites consecutivas, resultando em 600 observações diárias. A Figura 13 mostra o comportamento de toda a série temporal, que apresenta estacionariedade e sazonalidade.

Figura 13 – Série Star

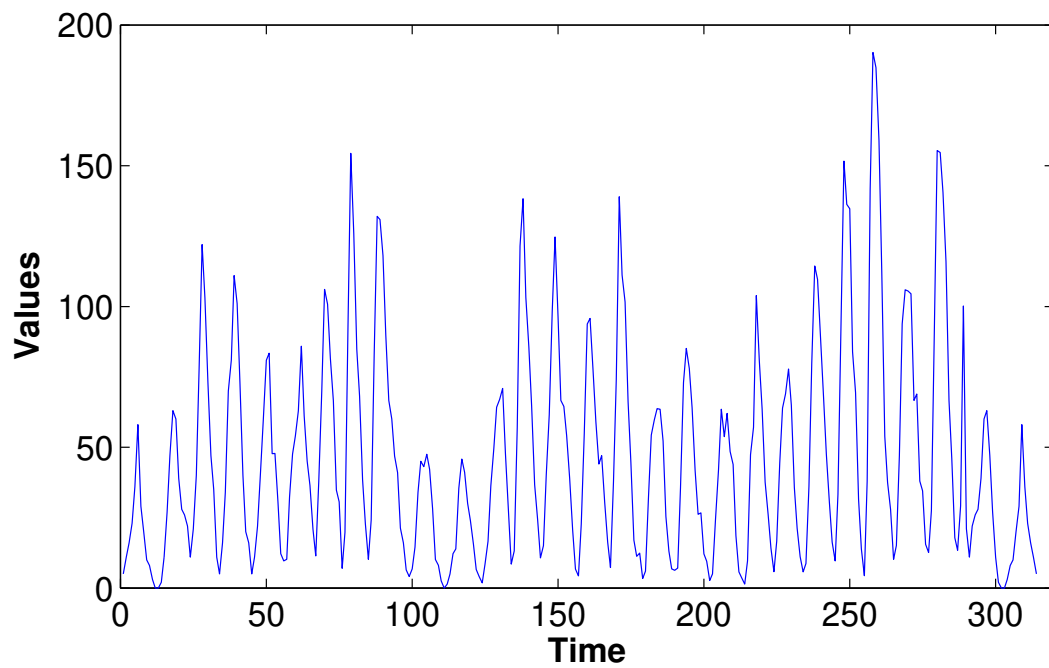


Fonte: o autor (2021)

#### 4.1.8 Sunspot

A *Sunspot* é uma série temporal que apresenta o número anual de manchas solares e é composta por 314 observações. A Figura 14 mostra o comportamento de toda a série temporal, que apresenta estacionariedade e sazonalidade.

Figura 14 – Série Sunspot

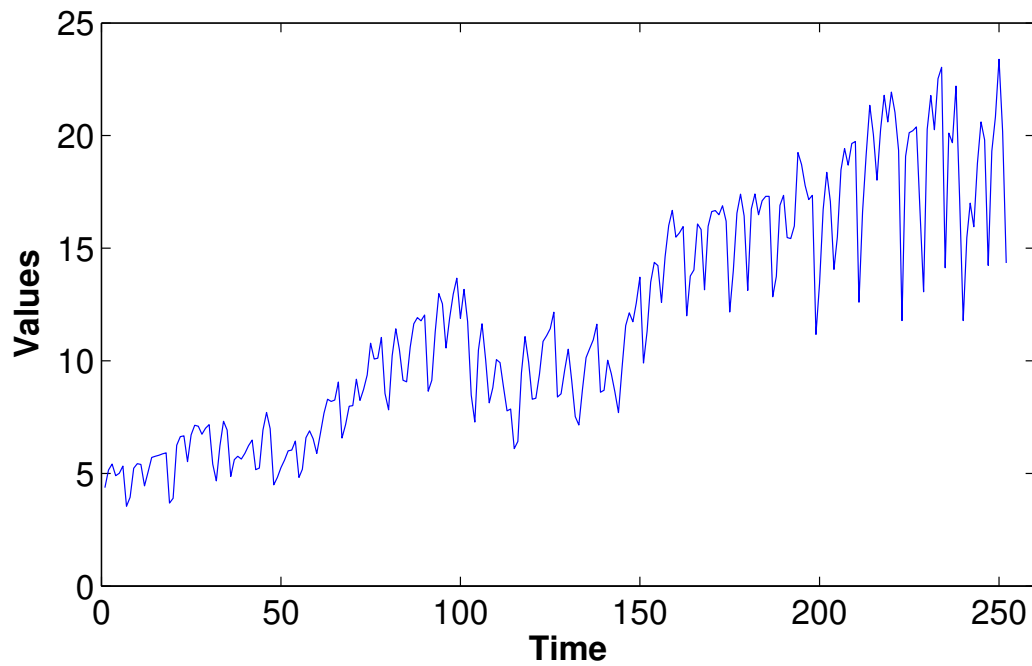


Fonte: o autor (2021)

#### 4.1.9 Vehicle

A *Vehicle* é uma série de vendas mensais de carros nos Estados Unidos e é composta por 252 observações. A Figura 15 mostra o comportamento de toda a série temporal, que apresenta não estacionariedade e sazonalidade.

Figura 15 – Série Vehicle

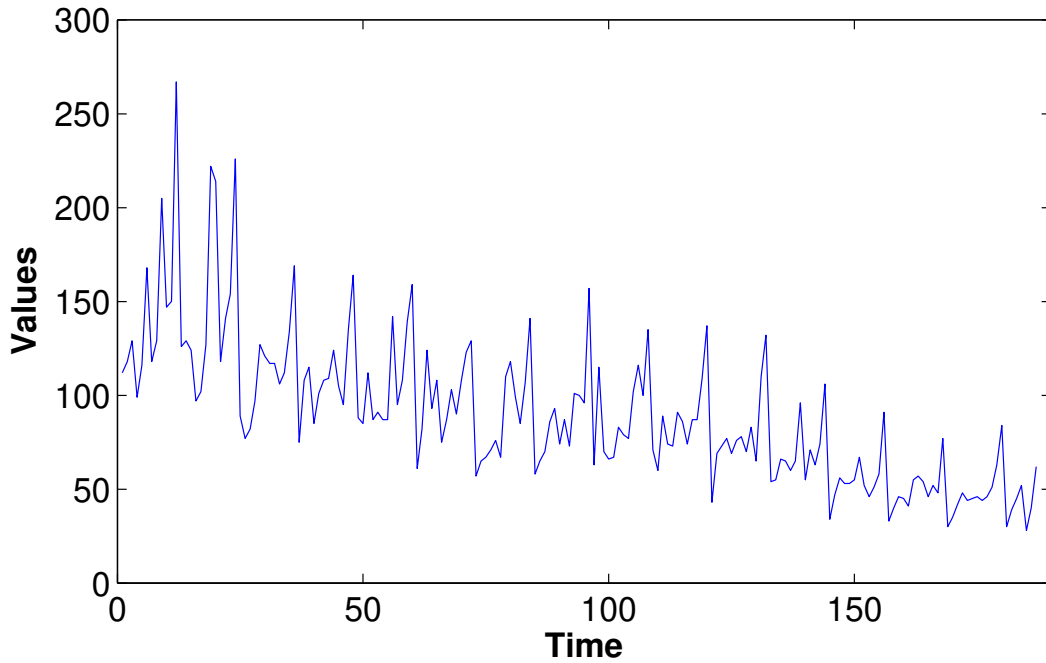


Fonte: o autor (2021)

#### 4.1.10 Wine

A *Wine* é uma série de vendas mensais de vinhos australianos e é composta por 187 observações. A Figura 16 mostra o comportamento de toda a série temporal, que apresenta não estacionariedade e sazonalidade.

Figura 16 – Série Wine



Fonte: o autor (2021)

## 4.2 MÉTRICAS DE AVALIAÇÃO

Quatro métricas foram utilizadas para avaliar as abordagens: *Mean Squared Error* (MSE), *Symmetric Mean Absolute Percentage Error* (SMAPE), *Average Relative Prediction Error Variation* (ARV) e *Mean Absolute Error* (MAE) (OLIVEIRA; LUDERMIR, 2014; MAO; XIAO, 2019). A seguinte nomenclatura é aplicada para definir as métricas:  $N$  é o número de observações no conjunto de teste ( $\gamma$ ),  $\hat{z}_t$  é a previsão e  $z_t$  é o valor real da observação, correspondente ao tempo  $t$  da amostra. O valor médio do conjunto de teste é representado por  $\bar{\gamma}$ .

MSE é uma métrica frequentemente utilizada na literatura para avaliar modelos de previsão (OLIVEIRA; LUDERMIR, 2014). O MSE calcula a distância ao quadrado da previsão ( $\hat{z}_t$ ) em relação ao valor real ( $z_t$ ), portanto, penaliza previsões que são distantes dos valores reais, independentemente do sinal do erro. A métrica MSE é definida por meio da seguinte equação:

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^N (z_t - \hat{z}_t)^2. \quad (4.1)$$

SMAPE, também conhecido como MAPE simétrico, é uma métrica que calcula o erro médio em percentual, independentemente da escala das observações. Diferentemente do MAPE, o SMAPE não aplica uma penalidade maior aos erros negativos, ou seja, quando as previsões

têm valores mais altos que os valores reais da série, tal característica torna a métrica simétrica e permite que os resultados estejam dentro do intervalo de 0% à 200%. O SMAPE é definido por meio da seguinte equação:

$$\text{SMAPE} = \frac{100}{N} \sum_{t=1}^n \frac{|z_t - \hat{z}_t|}{(|z_t| + |\hat{z}_t|)/2}. \quad (4.2)$$

ARV é uma métrica relativa que compara a previsão do modelo com a previsão da série utilizando a média. Se o valor ARV for menor que 1, a previsão do modelo é melhor que o uso da média, caso contrário, o modelo é pior que a previsão pela média. Se o valor do ARV for igual a 1, o modelo de previsão é equivalente à média. O ARV é definido pela seguinte equação:

$$\text{ARV} = \frac{\sum_{t=1}^N (z_t - \hat{z}_t)^2}{\sum_{t=1}^N (\bar{\gamma} - \hat{z}_t)^2}. \quad (4.3)$$

MAE é uma métrica utilizada para medir a acurácia do modelo a partir do erro médio absoluto, seu resultado é obtido por meio da distância média entre as previsões e os respectivos valores reais. O MAE é definido pela seguinte equação:

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |z_t - \hat{z}_t|. \quad (4.4)$$

A Equação 4.5 foi utilizada para calcular a diferença em percentual de desempenho entre a abordagem proposta e as abordagens presentes na literatura.  $\varepsilon_a$  é o desempenho obtido pela abordagem da literatura, enquanto que  $\varepsilon_b$  é o desempenho da abordagem proposta. Quando o valor resultante (*ratio*) for positivo, então a abordagem proposta apresenta um ganho de desempenho em relação à abordagem da literatura.

$$\text{ratio} = \frac{(\varepsilon_a - \varepsilon_b)}{\varepsilon_a} \times 100, \quad (4.5)$$

### 4.3 TESTE ESTATÍSTICO

Para realizar a comparação entre os desempenhos das abordagens, é necessário que se aplique um teste estatístico a fim de avaliar se as previsões apresentam diferenças estatisticamente significativas. Para tal, foi utilizado o teste *Diebold-Mariano Test* (DM) (DIEBOLD; MARIANO, 2002), frequentemente reportado na literatura (TILLY; EBNER; LIVAN, 2021; MUTAVHATSINDI; SIGAUKE; MBUVHA, 2020; ABDOLLAHI, 2020).



A aplicação do teste DM consiste nas seguintes etapas. Inicialmente, dadas as previsões obtidas pelos preditores A e B, calcula-se o custo  $L(e_t)$  para cada observação em (t), em que  $e_t = z_t - \hat{z}_t$  e  $L$  é uma métrica de desempenho, como, por exemplo, MSE. Em seguida, é calculada a diferença ( $d_t$ ) entre os valores de custo de cada abordagem:  $d_t = L_t^A - L_t^B$ . A hipótese nula do teste DM é de que em média não existe diferença entre os valores de custo do preditor A em relação aos valores de custo preditor B, ou seja,  $E[d_t] = 0$ , a hipótese alternativa será  $E[d_t] \neq 0$ .

#### 4.4 CONFIGURAÇÃO DO SISTEMA DE MÚLTIPLOS PREDITORES

Esta seção apresenta os parâmetros do SMP proposto e está organizada de acordo com as seguintes etapas: geração, seleção dinâmica e integração.

Na etapa de geração do SMP, utiliza-se o *Bagging*, que gera amostras com o mesmo tamanho de treinamento. Cada amostra gerada pelo *Bagging* é dividida em sub-amostras, sendo 33% para validação e 67% para treinamento. O *Pool* é formado por 100 SVRs (tamanho pré-selecionado com base em trabalhos da literatura (YAO; DAI; SONG, 2018; SILVA et al., 2020)). A SVR foi escolhida por apresentar estabilidade (DRUCKER et al., 1997), característica importante para este trabalho, pois reduz o custo computacional e a necessidade de várias execuções do modelo, já que o treinamento de uma SVR com os mesmos hiper-parâmetros e conjunto de treinamento retornará sempre o mesmo modelo. A abordagem *Grid Search* (HUANG; MAO; LIU, 2012) é utilizada para selecionar a melhor combinação de hiper-parâmetros para cada modelo. A Tabela 1 apresenta os valores utilizados na busca para cada parâmetro da SVR. O parâmetro *Kernel* define a função kernel que será utilizada no modelo SVR, o parâmetro *Gamma* define o coeficiente da função kernel gaussiana e sigmoid, o *C* é o parâmetro de regularização e *Epsilon* é o parâmetro que define o hipertubo da SVR. Para cada modelo do *Pool*, seleciona-se a combinação de parâmetros que resultou no menor MSE para o conjunto de validação.

A construção de SMP utilizando *Bagging* e SVR se justifica com base na literatura (VALENTINI; DIETTERICH, 2004), que demonstra que o *Bagging* possibilita obter modelos com baixa variância, enquanto que SVR gera modelos com baixo viés, reduzindo assim o risco de modelos com problemas de *Overfitting* ou *Underfitting*.

A segunda etapa se utiliza do método proposto (DSNAW) para selecionar  $n$  preditores. O DSNAW tem dois parâmetros:  $k$  (tamanho da região de competência) e  $n$  (número de preditores selecionados). O valor para cada parâmetro foi selecionado por meio de um conjunto

Tabela 1 – Valores para os parâmetros da SVR

Parâmetro	Valores
Kernel	Radial basis function and Sigmoid
Gamma	0.5, 1, 10, 20, ..., 100, 200, ..., 1000
Cost	0.1, 1, 100, 1000, 10000
Epsilon	1, 1E-01, 1E-02, 1E-03, 1E-04, 1E-05, 1E-06

de validação formado por 25% dos padrões que antecedem o conjunto de teste. Para cada série temporal, selecionou-se a combinação de valores, dentro do intervalo  $[1, 20]$ , para  $k$  e  $n$  que apresentam o menor valor de MSE no conjunto de validação. A utilização desse intervalo tem como motivação aumentar a região de busca utilizada na literatura (BUDI; AJI; WIDODO, 2014; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011). Além disso, é empregado o conjunto de validação para selecionar o valor para o parâmetro  $n$ , o DSNaw poderá ser aplicado para a seleção de preditores (quando  $n = 1$ ) ou seleção de *Ensemble* (quando  $n = 1$ ) com base no comportamento da série temporal.

O parâmetro da terceira etapa (Integração) é a abordagem de combinação utilizada quando  $n > 1$ . Para tal, quando na etapa anterior for selecionado o  $n > 1$ , é também realizada a seleção entre a combinação por média ou por mediana. Essas abordagens de combinação foram pré-selecionadas por serem frequentemente utilizadas e apresentarem simplicidade e previsões acuradas (KOURENTZES; BARROW; CRONE, 2014; ADHIKARI, 2015; LIVIERIS et al., 2020; ALKANDARI; AHMAD, 2020).

## 4.5 COMPARAÇÃO COM A LITERATURA

Foi realizada uma comparação com 11 abordagens da literatura, organizadas em quatro grupos: monolíticos, seleção dinâmica de preditores, *Ensemble*, seleção dinâmica de *Ensemble*. Cada sub-seção a seguir descreve as abordagens de cada grupo.

### 4.5.1 Modelos Monolíticos

As abordagens monolíticas são assim denominadas por utilizarem apenas um modelo treinado e validado em todo o conjunto de treinamento e aplicado para prever todas as observações do conjunto de teste. Por meio da comparação da abordagem proposta em relação aos monolíticos, é possível avaliar se a utilização de SMP com seleção dinâmica é capaz de gerar

previsões mais acuradas que o uso de apenas um preditor para prever todo o conjunto de teste.

Para esse grupo, foram pré-selecionadas as abordagens tradicionais da estatística: Auto Regressivo Integrado de Média Móveis (ARIMA)(BOX et al., 2015) e *Exponential Smoothing* (ETS) (HYNDMAN; ATHANASOPOULOS, 2018), já os monolíticos de aprendizado de máquina, tais como LSTM, Rede RBF e MLP, não foram implementados neste trabalho, pois o objetivo é avaliar a utilização do SMP proposto em comparação ao respectivo monolítico, utilizando, portanto, o modelo SVR.

O modelo ARIMA foi estimado com base na biblioteca *pmdarima* (SMITH et al., 2017), escrita na linguagem Python. O *pmdarima* segue o conceito de auto ARIMA e cria diferentes modelos ARIMA, inclusive com o componente sazonal, e seleciona o modelo com menor *Akaike's Information Criterion* (AIC) no conjunto de treinamento. O modelo ETS foi estimado utilizando a biblioteca *Statsmodel* (SEABOLD; PERKTOLD, 2010), escrita na linguagem Python. Para cada série temporal, foram criados diferentes modelos ETS e selecionado o modelo com menor MSE no conjunto de validação. O modelo SVR foi construído utilizando a abordagem *Grid Search* para selecionar a melhor combinação de hiper-parâmetros empregando os valores apresentados na Tabela 1.

#### 4.5.2 Seleção Dinâmica de Preditores

Este grupo inclui as duas principais abordagens da literatura que realizam a seleção dinâmica de preditores, ou seja, para cada observação de teste é selecionado um preditor. A comparação da abordagem proposta em relação às abordagens de seleção dinâmica de preditores da literatura permite avaliar se a abordagem DSNAW é capaz de selecionar preditores mais acurados que a literatura. Além disso, verifica se selecionar o número de preditores por meio do conjunto de validação pode resultar em um maior desempenho que selecionar apenas um preditor.

A abordagem DS-LA (1) utiliza o mesmo *Pool* de preditores da abordagem proposta e difere quanto à forma como a região de competência é definida, pois seleciona as  $k$  janelas de treinamento que são mais similares à janela de teste. Além disso, essa abordagem foi utilizada com valores pré-definidos para os parâmetros, utilizando  $k = 10$ ,  $n = 1$  e distância euclidiana para calcular a similaridade, seguindo a literatura (BUDI; AJI; WIDODO, 2014; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011). A segunda abordagem é a *Temporal-window*

*Framework* (TWF) (NETO et al., 2020), que consiste em particionar o conjunto de treinamento em  $j$  partições e treinar um preditor para partição. Essa abordagem busca construir um conjunto de preditores que serão capazes de aprender o comportamento local se tornando especialistas nos padrões da partição na qual foi treinado. Ainda, para cada nova janela de teste, é selecionado o preditor respectivo à partição que apresenta a menor distância (calculada por meio da métrica DTW (SAKOE; CHIBA, 1978) ). O TWF tem dois parâmetros, o  $k$  (tamanho da partição) e  $pint$  (percentual de interseção entre as partições), sendo que neste trabalho, os valores para os parâmetros foram selecionados por meio do conjunto de validação.

#### 4.5.3 *Ensemble*

Este grupo reúne as abordagens que aplicam a combinação para todos os modelos do *Pool*. A comparação dessas abordagens em relação à abordagem proposta permite avaliar quando deve-se utilizar a seleção dinâmica e quando deve-se combinar todos os modelos do *Pool*. As duas abordagens utilizadas são o  $Bagg_A$  e  $Bagg_M$ , que combinam todos os preditores do *Pool* do SMP proposto por meio da média e mediana, respectivamente.

#### 4.5.4 Seleção Dinâmica de *Ensemble*

Este grupo apresenta as abordagens que selecionam de forma dinâmica um *Ensemble* e combinam as previsões. A comparação dessas abordagens em relação à abordagem proposta permite avaliar se o DSNW é capaz de selecionar preditores mais acurados.

A abordagem DS-LA (BUDI; AJI; WIDODO, 2014; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011) foi aplicada para o contexto de seleção de *Ensemble*, os valores para os parâmetros  $k$  e  $n$  foram selecionados a partir do conjunto de validação. As abordagens  $DES_A$  e  $DES_M$  aplicam a abordagem *Dynamic Ensemble Selection* (DES) (ROONEY et al., 2004) para selecionar de forma dinâmica 10 preditores e combinar por meio da média e da mediana, respectivamente. A abordagem DES-PALR) (YAO; DAI; SONG, 2018) seleciona um *Ensemble* com os preditores que apresentam menor MSE nos padrões do *cluster* (região de competência) que apresenta a menor distância (calculada pela a distância Euclidiana) do centroide para a nova janela de teste.

Essas abordagens foram empregadas utilizando o mesmo *Pool* de preditores da abordagem proposta. E os respectivos parâmetros foram selecionados no conjunto de validação.

## 4.6 CONSIDERAÇÕES DO CAPÍTULO

A metodologia dos experimentos descreve as bases de dados utilizadas, as configurações, as métricas de avaliação e as abordagens da literatura empregadas para avaliar o método proposto.

O uso de séries temporais com comportamentos diferentes permite avaliar o método proposto em relação a comportamentos como sazonalidade, tendência, estacionariedade e não estacionariedade, assim como em diferentes áreas, tais como mercado financeiro, demanda de energia, demanda de produtos, ambiente, economia e astronomia.

As diferentes métricas de desempenho possibilitam avaliar o modelo em diferentes perspectivas e critérios, enquanto que a comparação com as abordagens da literatura permite avaliar se o método proposto apresenta previsões mais acuradas que: os métodos monolíticos, que utilizam apenas um preditor para toda a série e os métodos de seleção dinâmica de preditores, que seguem a suposição de que usar apenas um preditor gera uma acurácia maior que a combinação; os métodos que combinam as previsões de todos os preditores do *Pool*, seguindo a suposição de que combinar todas as previsões gera uma acurácia maior que aplicar antes da seleção, e os métodos de seleção dinâmica de *Ensemble*, que selecionam mais de um preditor e depois aplicam a combinação, mas com abordagens diferentes para a construção da região de competência.

## 5 ANÁLISE E RESULTADOS

Este capítulo apresenta os resultados das previsões obtidas a partir das diferentes abordagens seguindo a metodologia apresentada no capítulo anterior e se organiza conforme apresenta-se a seguir. A seção 5.1 introduz uma análise dos parâmetros selecionados no conjunto da validação, já a seção 5.2 apresenta os resultados do DSNaw em comparação aos métodos monolíticos. Por sua vez, a seção 5.3 traz os resultados do método proposto em relação aos métodos de seleção dinâmica de preditores, enquanto que a seção 5.4 descreve os resultados do método proposto em comparação às abordagens da literatura que combinam as previsões de todos os preditores do *Pool*. Já a seção 5.5 apresenta os resultados do DSNaw em comparação às abordagens que selecionam *Ensemble* forma dinâmica, enquanto que a seção 5.6 mostra uma análise geral dos resultados. Finalmente, é apresentada uma consideração final do capítulo na seção 5.7.

### 5.1 ANÁLISE DOS PARÂMETROS SELECIONADOS

A Tabela 2 apresenta os parâmetros selecionados por meio do conjunto de validação em comparação aos parâmetros que apresentam o menor MSE no conjunto de teste. Os resultados apresentados na tabela permitem observar que a combinação de parâmetros selecionados na validação difere da combinação de parâmetros que apresentou o melhor desempenho em teste. Em algumas séries temporais, como, por exemplo *Amazon*, *Goldman Sachs*, *Star* e *Vehicle*, apesar da diferença dos parâmetros, os desempenhos foram próximos a diferenças abaixo de 20%. A maior diferença de desempenho foi obtida na série *Apple*.

Tabela 2 – Valores dos parâmetros do DSNaw selecionados por meio do conjunto de validação em comparação aos parâmetros que apresentam o melhor desempenho no conjunto de teste.

Base de Dados	Parâmetros selecionados (Validação)				Melhores Parâmetros (Teste)			
	k	n	Abordagem de Combinação	$MSE_{teste}$	k	n	Abordagem de Combinação	$MSE_{teste}$
Amazon	20	14	Mediana	1.30E-4	20	10	Mediana	1.28E-4
Apple	7	11	Média	1.90E-2	12	1	-	2.16E-4
Electricity	2	16	Mediana	3.74E-3	13	3	Média	2.28E-3
Goldman Sachs	10	11	Mediana	4.06E-4	13	11	Mediana	3.89E-4
Microsoft	12	10	Média	1.99E-3	18	13	Mediana	1.38E-3
Pollution	12	3	Mediana	4.47E-2	4	3	Média	3.18E-2
Star	10	16	Mediana	6.77E-5	2	12	Mediana	5.79E-5
Sunspot	9	5	Mediana	3.31E-2	3	1	-	1.92E-2
Vehicle	18	10	Mediana	1.81E-2	20	3	Média	1.68E-2
Wine	8	11	Mediana	2.94E-3	9	2	Média	1.92E-3

Outra análise se dá em relação à aplicação do DSNAW para seleção de preditor (quando  $n = 1$ ) e seleção de *Ensemble* (quando  $n > 1$ ). Os parâmetros selecionados em validação demonstram que para todas as séries foi aplicado o DSNAW para seleção de *Ensemble*. Porém, nas séries *Apple* e *Sunspot*, o melhor desempenho em teste acontece ao se utilizar a seleção dinâmica de preditores.

Por meio dessa comparação entre os parâmetros selecionados no conjunto de validação e os melhores parâmetros do conjunto de teste é possível inferir que: (I) O desempenho do método DSNAW depende dos valores selecionados para cada parâmetro, (II) apesar de a seleção de parâmetros no conjunto de validação não encontrar os valores que resultaram nos maiores desempenho em teste, os parâmetros selecionados obtiveram desempenho próximo do melhor desempenho possível para a maioria das séries temporais avaliadas.

Para as próximas análises, foram utilizados os resultados da abordagem proposta utilizando os parâmetros selecionados por meio do conjunto de validação.

## 5.2 COMPARAÇÃO COM OS MODELOS MONOLÍTICOS

A Tabela 3 apresenta os resultados da comparação entre o método proposto e as abordagens monolíticas. Na perspectiva da métrica MSE, o método proposto obteve os melhores resultados em 8 das 10 séries temporais avaliadas, perdendo apenas na série *Sunspot*, na qual o SVR apresentou o melhor resultado, e na série *Wine*, na qual o ARIMA obteve o melhor MSE. Por meio da métrica ARV, é possível observar que nas séries *Apple*, *Pollution* e *Vehicle*, todos os preditores apresentaram previsões menos acuradas que a utilização da média da série. Na série *Wine*, apenas a SVR apresentou previsões mais acuradas que a média. Na série *Sunspot* apenas o método proposto e a SVR apresentaram ARV abaixo de 1. Já nas demais séries, *Amazon*, *Electricity*, *Goldman*, *Microsoft* e *Star*, os melhores resultados do ARV foram obtidos por meio do método proposto. Na perspectiva da métrica MAE, o método proposto apresentou as previsões mais acuradas em 9 das 10 séries temporais, perdendo apenas na série *Wine*, na qual a melhor abordagem foi a SVR. Em relação à métrica SMAPE, o método proposto apresentou os melhores resultados em 8 séries temporais, perdendo apenas na série *Pollution* para o método ETS e na série *Star* para o método SVR.

Tabela 3 – Comparação do desempenho do método proposto em relação aos métodos monolíticos. Em negrito estão os valores com o maior desempenho.

Série Temporal	Métrica	ARIMA	ETS	SVR	Proposto
Amazon	MSE	6.08E-3	6.04E-3	3.32E-2	<b>1.30E-4</b>
	ARV	2.52	2.54	1.20	<b>3.93E-2</b>
	MAE	6.04E-2	6.03E-2	1.64E-1	<b>8.94E-3</b>
	SMAPE	63.83	63.59	90.75	<b>13.80</b>
Apple	MSE	2.74E-2	2.73E-2	1.38E-1	<b>1.90E-2</b>
	ARV	3.07	3.09	<b>1.51</b>	2.25
	MAE	1.19E-1	1.19E-1	2.82E-1	<b>8.02E-2</b>
	SMAPE	52.79	52.79	74.04	<b>42.15</b>
Electricity	MSE	2.12E-2	1.51E-2	1.89E-1	<b>3.74E-3</b>
	ARV	3.52	4.65E+2	1.02	<b>2.67E-1</b>
	MAE	1.23E-1	1.02E-1	4.26E-1	<b>5.18E-2</b>
	SMAPE	16.35	13.83	83.13	<b>7.25</b>
Goldman	MSE	2.85E-2	1.88E-2	4.30E-2	<b>4.06E-4</b>
	ARV	9.33	5.74	1.42	<b>2.64E-2</b>
	MAE	1.52E-1	1.16E-1	1.68E-1	<b>1.53E-2</b>
	SMAPE	68.98	51.73	61.50	<b>9.44</b>
Microsoft	MSE	6.01E-2	6.01E-2	2.30E-2	<b>1.99E-3</b>
	ARV	1.32	1.32	1.31	<b>1.28E-1</b>
	MAE	2.13E-1	2.13E-1	1.23E-1	<b>3.41E-2</b>
	SMAPE	27.42	27.41	17.96	<b>5.08</b>
Pollution	MSE	7.69E-2	4.76E-2	1.26E-1	<b>4.47E-2</b>
	ARV	1.36	1.67	<b>1.08</b>	1.78
	MAE	2.29E-1	1.79E-1	3.35E-1	<b>1.74E-1</b>
	SMAPE	37.41	<b>31.52</b>	50.10	37.43
Star	MSE	1.72E-1	8.99E-2	8.57E-5	<b>6.77E-5</b>
	ARV	3.05	3.41	1.35E-3	<b>1.07E-3</b>
	MAE	3.43E-1	2.45E-1	7.04E-3	<b>6.08E-3</b>
	SMAPE	72.77	52.13	<b>3.61</b>	3.77
Sunspot	MSE	6.28E-2	7.75E-2	<b>2.66E-2</b>	3.31E-2
	ARV	1.85E+1	9.54	<b>5.45E-1</b>	9.55E-1
	MAE	1.94E-1	1.97E-1	<b>1.14E-1</b>	<b>1.14E-1</b>
	SMAPE	74.80	75.59	60.31	<b>46.31</b>
Vehicle	MSE	2.94E-2	2.38E-2	2.86E-1	<b>1.81E-2</b>
	ARV	4.44	3.26E+1	<b>1.08</b>	1.45
	MAE	1.23E-1	1.15E-1	5.14E-1	<b>1.09E-1</b>
	SMAPE	16.71	15.82	97.24	<b>15.14</b>
Wine	MSE	<b>2.69E-3</b>	3.15E-3	4.60E-3	2.94E-3
	ARV	8.76	6.71	<b>9.43E-1</b>	1.39
	MAE	<b>3.32E-2</b>	4.48E-2	5.50E-2	3.54E-2
	SMAPE	44.77	54.29	93.91	<b>43.84</b>

Essa comparação do método proposto com os métodos monolíticos permitiu inferir que: (I) a utilização de Sistemas de Múltiplos Preditores (SMP) tende a obter previsões mais acuradas que a utilização de apenas um preditor monolítico e (II) o método proposto foi capaz de obter os melhores resultados em todas as métricas na maioria das séries temporais avaliadas.



### 5.3 COMPARAÇÃO COM OS MÉTODOS DE SELEÇÃO DINÂMICA DE PREDITORES

A Tabela 4 apresenta os resultados da comparação entre o método proposto e as abordagens de seleção dinâmica de preditores. Na perspectiva da métrica MSE, o método proposto obteve os melhores resultados em 8 das 10 séries temporais, perdendo apenas na série *Star*, na qual o melhor resultado foi obtido por meio do método DS-LA (1) e na série *Sunspot*, na qual o melhor resultado foi obtido por meio do método TWF. A métrica ARV demonstra que nas séries *Apple*, *Pollution* e *Wine*, todas as abordagens apresentaram previsões menos acuradas que a utilização da média da série. Os métodos DS-LA (1) e TWF apresentaram os melhores resultados ARV nas séries *Star* e *Vehicle*, respectivamente. Já para as demais 5 séries, os melhores resultados do ARV foram obtidos por meio do método proposto. Na perspectiva das métricas MAE e SMAPE, o método proposto obteve os melhores resultados em 9 de 10 séries temporais avaliadas, perdendo apenas na série *Star* para o método DS-LA (1).

Tabela 4 – Comparação do desempenho do método proposto em relação aos métodos de seleção dinâmica de preditores. Em negrito estão os valores com o maior desempenho.

Série Temporal	Métrica	DS-LA (1)	TWF	Proposto
Amazon	MSE	9.29E-3	1.66E-4	<b>1.30E-4</b>
	ARV	1.38	5.62E-2	<b>3.93E-2</b>
	MAE	5.51E-2	1.01E-2	<b>8.94E-3</b>
	SMAPE	43.18	16.19	<b>13.80</b>
Apple	MSE	1.70E-1	6.89E-2	<b>1.90E-2</b>
	ARV	<b>1.51</b>	2.02	2.25
	MAE	3.02E-1	1.85E-1	<b>8.02E-2</b>
	SMAPE	73.79	62.03	<b>42.15</b>
Electricity	MSE	4.76E-2	6.09E-3	<b>3.74E-3</b>
	ARV	1.15	2.69E-1	<b>2.67E-1</b>
	MAE	1.80E-1	6.11E-2	<b>5.18E-2</b>
	SMAPE	28.01	9.40	<b>7.25</b>
Goldman	MSE	3.65E-3	6.21E-4	<b>4.06E-4</b>
	ARV	1.72E-1	4.32E-2	<b>2.64E-2</b>
	MAE	4.40E-2	1.90E-2	<b>1.53E-2</b>
	SMAPE	31.82	11.56	<b>9.44</b>
Microsoft	MSE	1.56E-2	1.30E-2	<b>1.99E-3</b>
	ARV	5.08E-1	4.77E-1	<b>1.28E-1</b>
	MAE	9.82E-2	8.68E-2	<b>3.41E-2</b>
	SMAPE	15.36	13.70	<b>5.08</b>
Pollution	MSE	9.96E-2	1.16E-1	<b>4.47E-2</b>
	ARV	1.24	<b>1.20</b>	1.78
	MAE	2.85E-1	3.11E-1	<b>1.74E-1</b>
	SMAPE	68.09	76.77	<b>37.43</b>
Star	MSE	<b>6.24E-5</b>	7.08E-5	6.77E-5
	ARV	<b>9.79E-4</b>	1.11E-3	1.07E-3
	MAE	<b>5.38E-3</b>	6.84E-3	6.08E-3
	SMAPE	3.35	<b>3.02</b>	3.77
Sunspot	MSE	4.13E-2	<b>3.16E-2</b>	3.31E-2
	ARV	1.34	1.17	<b>9.55E-1</b>
	MAE	1.42E-1	1.33E-1	<b>1.14E-1</b>
	SMAPE	63.34	56.45	<b>46.31</b>
Vehicle	MSE	8.99E-2	6.34E-2	<b>1.81E-2</b>
	ARV	1.31	<b>9.97E-1</b>	1.45
	MAE	2.63E-1	2.15E-1	<b>1.09E-1</b>
	SMAPE	40.89	34.55	<b>15.14</b>
Wine	MSE	1.75E-2	7.88E-3	<b>2.94E-3</b>
	ARV	<b>1.06</b>	1.41	1.39
	MAE	1.12E-1	8.25E-2	<b>3.54E-2</b>
	SMAPE	100.50	77.54	<b>43.84</b>

Esses resultados demonstram que o método proposto apresenta desempenho superior às técnicas de seleção dinâmica de preditores na maioria das séries temporais avaliadas. Esse desempenho superior está associado a dois pontos: (I) o método proposto seleciona por meio da validação o valor de  $n$ , podendo realizar tanto a seleção dinâmica de preditores quanto a seleção dinâmica de *Ensemble* e (II) o método proposto seleciona os  $n$  preditores a partir da região de competência formada por padrões temporalmente próximos da observação a ser predita.

#### 5.4 COMPARAÇÃO COM OS MÉTODOS *ENSEMBLE*

A Tabela 5 apresenta os resultados da comparação entre o método proposto e as abordagens que combinam todas as previsões do *Pool*. Em relação à métrica MSE, o método proposto obteve os melhores resultados em 9 de 10 séries temporais avaliadas, perdendo apenas para o método *Bagg<sub>M</sub>* na série *Star*. Os resultados da métrica ARV possibilita observar que nas séries *Apple*, *Pollution*, *Vehicle* e *Wine*, todas as abordagens apresentaram desempenho inferior que aquele se dá ao se utilizar a média como abordagem de previsão. Na série *Star*, o melhor valor ARV foi obtido por meio da abordagem *Bagg<sub>M</sub>*, já para as demais 5 séries, o método proposto obteve os melhores resultados. Na perspectiva da métrica MAE, o método proposto obteve as previsões mais acuradas em 9 de 10 séries temporais, perdendo apenas para o método *Bagg<sub>M</sub>* na série *Star*. Em relação à métrica SMAPE, o método proposto obteve os melhores resultados em 8 de 10 séries temporais, perdendo apenas nas séries *Pollution* e *Star*, nas quais os melhores valores foram obtidos por meio das abordagens *Bagg<sub>A</sub>* e *Bagg<sub>M</sub>*, respectivamente.

Tabela 5 – Comparação do desempenho do método proposto em relação aos métodos que combinam todas as previsões do *Pool*. Em negrito estão os valores com o maior desempenho.

Série Temporal	Métrica	$Bagg_A$	$Bagg_M$	Proposto
Amazon	MSE	2.36E-3	1.08E-3	<b>1.30E-4</b>
	ARV	9.26E-1	4.57E-1	<b>3.93E-2</b>
	MAE	2.94E-2	1.97E-2	<b>8.94E-3</b>
	SMAPE	33.13	26.32	<b>13.80</b>
Apple	MSE	1.26E-1	1.49E-1	<b>1.90E-2</b>
	ARV	1.68	<b>1.62</b>	2.25
	MAE	2.46E-1	2.65E-1	<b>8.02E-2</b>
	SMAPE	66.92	67.92	<b>42.15</b>
Electricity	MSE	2.05E-2	1.56E-2	<b>3.74E-3</b>
	ARV	8.91E-1	6.96E-1	<b>2.67E-1</b>
	MAE	1.30E-1	1.15E-1	<b>5.18E-2</b>
	SMAPE	19.01	16.98	<b>7.25</b>
Goldman	MSE	7.11E-4	5.46E-4	<b>4.06E-4</b>
	ARV	4.42E-2	3.47E-2	<b>2.64E-2</b>
	MAE	2.03E-2	1.81E-2	<b>1.53E-2</b>
	SMAPE	12.90	11.46	<b>9.44</b>
Microsoft	MSE	5.25E-3	6.05E-3	<b>1.99E-3</b>
	ARV	2.74E-1	3.14E-1	<b>1.28E-1</b>
	MAE	5.83E-2	5.97E-2	<b>3.41E-2</b>
	SMAPE	8.66	8.90	<b>5.08</b>
Pollution	MSE	4.56E-2	1.09E-1	<b>4.47E-2</b>
	ARV	1.64	<b>1.22</b>	1.78
	MAE	1.79E-1	2.99E-1	<b>1.74E-1</b>
	SMAPE	<b>36.88</b>	72.24	37.43
Star	MSE	5.26E-5	<b>5.25E-5</b>	6.77E-5
	ARV	8.28E-4	<b>8.27E-4</b>	1.07E-3
	MAE	5.07E-3	<b>5.02E-3</b>	6.08E-3
	SMAPE	2.48	<b>2.36</b>	3.77
Sunspot	MSE	3.26E-2	3.61E-2	<b>3.31E-2</b>
	ARV	1.32	1.44	<b>9.55E-1</b>
	MAE	1.22E-1	1.25E-1	<b>1.14E-1</b>
	SMAPE	47.94	46.36	<b>46.31</b>
Vehicle	MSE	5.66E-2	5.69E-2	<b>1.81E-2</b>
	ARV	1.22	<b>1.18</b>	1.45
	MAE	2.15E-1	2.15E-1	<b>1.09E-1</b>
	SMAPE	30.77	30.80	<b>15.14</b>
Wine	MSE	2.46E-2	4.04E-2	<b>2.94E-3</b>
	ARV	1.09	<b>1.07</b>	1.39
	MAE	1.50E-1	1.95E-1	<b>3.54E-2</b>
	SMAPE	102.30	113.68	<b>43.84</b>

Esses resultados permitem inferir que selecionar um subconjunto de preditores de forma dinâmica tende a gerar previsões mais acuradas que combinar as previsões de todos os preditores do *Pool*. Além disso, o método proposto se mostrou capaz de realizar de forma dinâmica a poda do *Pool*, resultando em  $n$  preditores que, ao serem combinados, resultaram em previsões mais acuradas que a combinação de todos os preditores do *Pool*.

## 5.5 COMPARAÇÃO COM OS MÉTODOS DE SELEÇÃO DINÂMICA DE *ENSEMBLE*

A Tabela 6 apresenta os resultados da comparação entre o método proposto e as abordagens da literatura que selecionam de forma dinâmica mais de um preditor. Na perspectiva da métrica MSE, o método proposto apresentou os melhores resultados em 8 de 10 séries temporais avaliadas, perdendo nas séries *Star* e *Sunspot* para os métodos  $DES_M$  e  $DES_A$ , respectivamente. Os resultados da métrica ARV demonstram que nas séries *Apple*, *Pollution*, *Vehicle* e *Wine*, todas as abordagens apresentam previsões menos acuradas que aquelas geradas ao se utilizar a média da série com abordagem de previsão. Na série *Star*, o melhor resultado foi obtido pela abordagem  $DES_M$ , já para as demais 5 séries, os melhores resultados foram obtidos por meio do método proposto. Em relação às métricas MAE e SMAPE, o método proposto apresentou os melhores resultados em 9 de 10 séries temporais, perdendo apenas na série *Star*, na qual o melhor resultado foi obtido por meio da abordagem  $DES_M$ .

Tabela 6 – Comparação do desempenho do método proposto em relação aos métodos de seleção dinâmica de *Ensemble*. Em negrito estão os valores com o maior desempenho.

Série Temporal	Métrica	DS-LA	$DES_A$	$DES_M$	DES-PALR	Proposto
Amazon	MSE	7.20E-3	9.16E-3	7.20E-3	2.18E-2	<b>1.30E-4</b>
	ARV	1.35	1.39	1.35	1.38	<b>3.93E-2</b>
	MAE	4.93E-2	5.48E-2	4.92E-2	9.16E-2	<b>8.94E-3</b>
	SMAPE	41.23	42.99	41.21	57.12	<b>13.80</b>
Apple	MSE	1.62E-1	1.53E-1	1.65E-1	1.71E-1	<b>1.90E-2</b>
	ARV	1.53	1.55	1.52	<b>1.48</b>	2.25
	MAE	2.90E-1	2.84E-1	2.96E-1	3.10E-1	<b>8.02E-2</b>
	SMAPE	72.17	72.09	73.25	75.61	<b>42.15</b>
Electricity	MSE	3.72E-2	3.11E-2	3.61E-2	6.58E-2	<b>3.74E-3</b>
	ARV	1.21	1.21	1.35	1.20	<b>2.67E-1</b>
	MAE	1.58E-1	1.46E-1	1.45E-1	2.29E-1	<b>5.18E-2</b>
	SMAPE	23.65	21.43	21.40	36.18	<b>7.25</b>
Goldman	MSE	2.36E-3	2.20E-3	2.50E-3	3.58E-3	<b>4.06E-4</b>
	ARV	1.24E-1	1.17E-1	1.32E-1	1.79E-1	<b>2.64E-2</b>
	MAE	3.71E-2	3.53E-2	3.66E-2	4.43E-2	<b>1.53E-2</b>
	SMAPE	24.91	23.69	24.65	31.20	<b>9.44</b>
Microsoft	MSE	1.56E-2	1.16E-2	1.36E-2	1.78E-2	<b>1.99E-3</b>
	ARV	5.08E-1	4.51E-1	4.87E-1	5.48E-1	<b>1.28E-1</b>
	MAE	9.82E-2	8.67E-2	9.29E-2	1.06E-1	<b>3.41E-2</b>
	SMAPE	15.36	13.34	14.50	16.92	<b>5.08</b>
Pollution	MSE	1.07E-1	1.03E-1	1.06E-1	1.15E-1	<b>4.47E-2</b>
	ARV	1.23	1.24	1.23	<b>1.20</b>	1.78
	MAE	2.95E-1	2.89E-1	2.93E-1	3.09E-1	<b>1.74E-1</b>
	SMAPE	71.13	69.14	70.43	76.04	<b>37.43</b>
Star	MSE	5.29E-5	5.26E-5	<b>5.25E-5</b>	6.84E-5	6.77E-5
	ARV	8.31E-4	8.28E-4	<b>8.27E-4</b>	1.07E-3	1.07E-3
	MAE	5.48E-3	5.07E-3	<b>5.02E-3</b>	6.06E-3	6.08E-3
	SMAPE	4.32	2.48	<b>2.36</b>	2.74	3.77
Sunspot	MSE	3.52E-2	<b>3.26E-2</b>	3.61E-2	3.97E-2	3.31E-2
	ARV	1.24	1.32	1.44	1.55	<b>9.55E-1</b>
	MAE	1.24E-1	1.22E-1	1.25E-1	1.37E-1	<b>1.14E-1</b>
	SMAPE	50.00	47.94	46.36	54.54	<b>46.31</b>
Vehicle	MSE	5.62E-2	5.03E-2	5.54E-2	3.72E-2	<b>1.81E-2</b>
	ARV	1.21	1.22	<b>1.19</b>	1.42	1.45
	MAE	2.14E-1	2.02E-1	2.12E-1	1.66E-1	<b>1.09E-1</b>
	SMAPE	30.84	28.91	30.57	23.56	<b>15.14</b>
Wine	MSE	3.80E-2	2.58E-2	3.79E-2	3.99E-2	<b>2.94E-3</b>
	ARV	1.07	1.08	1.07	<b>1.06</b>	1.39
	MAE	1.88E-1	1.53E-1	1.88E-1	1.94E-1	<b>3.54E-2</b>
	SMAPE	111.88	103.13	112.01	113.60	<b>43.84</b>

Esses resultados permitem observar que o método proposto foi capaz de selecionar modelos com previsões mais acuradas que as abordagens da literatura na maioria das séries temporais avaliadas. Esses resultados superiores às técnicas de seleção dinâmica de *Ensemble* estão associados principalmente à abordagem utilizada para a construção da região de competência.

## 5.6 ANÁLISE GERAL

A Tabela 7 mostra os resultados do método proposto em relação às abordagens da literatura. Na perspectiva da métrica MSE, o DSNAW apresentou os melhores resultados em 7 de 10 séries temporais. Em relação à métrica ARV, o método proposto apresentou os melhores resultados em 5 de 10 séries temporais. Na perspectiva da métrica MAE, o DSNAW apresentou as previsões mais acuradas em 8 séries temporais, enquanto que para a métrica SMAPE, o método proposto apresentou os melhores resultados em 8 de 10 séries temporais.

Tabela 7 – Comparação do desempenho do método proposto em relação aos métodos da literatura. Em negrito estão os valores com o maior desempenho.

Séries Temporais	Métricas	Monolíticos			Seleção Dinâmica de Preditores		Ensemble		Seleção Dinâmica de Ensemble				Proposto
		ARIMA	ETS	SVR	DS-LA (1)	TWF	<i>Bagg<sub>A</sub></i>	<i>Bagg<sub>M</sub></i>	DS-LA	<i>DES<sub>A</sub></i>	<i>DES<sub>M</sub></i>	DES-PALR	
Amazon	MSE	6.08E-3	6.04E-3	3.32E-2	9.29E-3	1.66E-4	2.36E-3	1.08E-3	7.20E-3	9.16E-3	7.20E-3	2.18E-2	<b>1.30E-4</b>
	ARV	2.52	2.54	1.20	1.38	5.62E-2	9.26E-1	4.57E-1	1.35	1.39	1.35	1.38	<b>3.93E-2</b>
	MAE	6.04E-2	6.03E-2	1.64E-1	5.51E-2	1.01E-2	2.94E-2	1.97E-2	4.93E-2	5.48E-2	4.92E-2	9.16E-2	<b>8.94E-3</b>
	SMAPE	63.83	63.59	90.75	43.18	16.19	33.13	26.32	41.23	42.99	41.21	57.12	<b>13.80</b>
Apple	MSE	2.74E-2	2.73E-2	1.38E-1	1.70E-1	6.89E-2	1.26E-1	1.49E-1	1.62E-1	1.53E-1	1.65E-1	1.71E-1	<b>1.90E-2</b>
	ARV	3.07	3.09	1.51	1.51	2.02	1.68	1.62	1.53	1.55	1.52	<b>1.48</b>	2.25
	MAE	1.19E-1	1.19E-1	2.82E-1	3.02E-1	1.85E-1	2.46E-1	2.65E-1	2.90E-1	2.84E-1	2.96E-1	3.10E-1	<b>8.02E-2</b>
	SMAPE	52.79	52.79	74.04	73.79	62.03	66.92	67.92	72.17	72.09	73.25	75.61	<b>42.15</b>
Electricity	MSE	2.12E-2	1.51E-2	1.89E-1	4.76E-2	6.09E-3	2.05E-2	1.56E-2	3.72E-2	3.11E-2	3.61E-2	6.58E-2	<b>3.74E-3</b>
	ARV	3.52	4.65	1.02	1.15	2.69E-1	8.91E-01	6.96E-01	1.21	1.21	1.35	1.20	<b>2.67E-01</b>
	MAE	1.23E-1	1.02E-1	4.26E-1	1.80E-1	6.11E-2	1.30E-1	1.15E-1	1.58E-1	1.46E-1	1.45E-1	2.29E-1	<b>5.18E-2</b>
	SMAPE	16.35	13.83	83.13	28.01	9.40	19.01	16.98	23.65	21.43	21.40	36.18	<b>7.25</b>
Goldman	MSE	2.85E-2	1.88E-2	4.30E-2	3.65E-3	6.21E-4	7.11E-4	5.46E-4	2.36E-3	2.20E-3	2.50E-3	3.58E-3	<b>4.06E-4</b>
	ARV	9.33	5.74	1.42	1.72E-1	4.32E-2	4.42E-2	3.47E-2	1.24E-1	1.17E-1	1.32E-1	1.79E-1	<b>2.64E-2</b>
	MAE	1.52E-1	1.16E-1	1.68E-1	4.40E-2	1.90E-2	2.03E-2	1.81E-2	3.71E-2	3.53E-2	3.66E-2	4.43E-2	<b>1.53E-2</b>
	SMAPE	68.98	51.73	61.50	31.82	11.56	12.90	11.46	24.91	23.69	24.65	31.20	<b>9.44</b>
Microsoft	MSE	6.01E-2	6.01E-2	2.30E-2	1.56E-2	1.30E-2	5.25E-3	6.05E-3	1.56E-2	1.16E-2	1.36E-2	1.78E-2	<b>1.99E-3</b>
	ARV	1.32	1.32	1.31	5.08E-1	4.77E-1	2.74E-1	3.14E-1	5.08E-1	4.51E-1	4.87E-1	5.48E-1	<b>1.28E-1</b>
	MAE	2.13E-1	2.13E-1	1.23E-1	9.82E-2	8.68E-2	5.83E-2	5.97E-2	9.82E-2	8.67E-2	9.29E-2	1.06E-1	<b>3.41E-2</b>
	SMAPE	27.42	27.41	17.96	15.36	13.70	8.66	8.90	15.36	13.34	14.50	16.92	<b>5.08</b>
Pollution	MSE	7.69E-2	4.76E-2	1.26E-1	9.96E-2	1.16E-1	4.56E-2	1.09E-1	1.07E-1	1.03E-1	1.06E-1	1.15E-1	<b>4.47E-2</b>
	ARV	1.36	1.67	<b>1.08</b>	1.24	1.20	1.64	1.22	1.23	1.24	1.23	1.20	1.78
	MAE	2.29E-1	1.79E-1	3.35E-1	2.85E-1	3.11E-1	1.79E-1	2.99E-1	2.95E-1	2.89E-1	2.93E-1	3.09E-1	<b>1.74E-1</b>
	SMAPE	37.41	<b>31.52</b>	50.10	68.09	76.77	36.88	72.24	71.13	69.14	70.43	76.04	37.43
Star	MSE	1.72E-1	8.99E-2	8.57E-5	6.24E-5	7.08E-5	5.26E-5	<b>5.25E-5</b>	5.29E-5	5.26E-5	<b>5.25E-5</b>	6.84E-5	6.77E-5
	ARV	3.05	3.41	1.35E-3	9.79E-4	1.11E-3	8.28E-4	<b>8.27E-4</b>	8.31E-4	8.28E-4	<b>8.27E-4</b>	1.07E-3	1.07E-3
	MAE	3.43E-1	2.45E-1	7.04E-3	5.38E-3	6.84E-3	5.07E-3	<b>5.02E-3</b>	5.48E-3	5.07E-3	<b>5.02E-3</b>	6.06E-3	6.08E-3
	SMAPE	72.77	52.13	3.61	3.35	3.02	2.48	<b>2.36</b>	4.32	2.48	<b>2.36</b>	2.74	3.77
Sunspot	MSE	6.28E-2	7.75E-2	<b>2.66E-2</b>	4.13E-2	3.16E-2	3.26E-2	3.61E-2	3.52E-2	3.26E-2	3.61E-2	3.97E-2	3.31E-2
	ARV	1.85E+1	9.54	<b>5.45E-1</b>	1.34	1.17	1.32	1.44	1.24	1.32	1.44	1.55	9.55E-1
	MAE	1.94E-1	1.97E-1	<b>1.14E-1</b>	1.42E-1	1.33E-1	1.22E-1	1.25E-1	1.24E-1	1.22E-1	1.25E-1	1.37E-1	<b>1.14E-1</b>
	SMAPE	74.80	75.59	60.31	63.34	56.45	47.94	46.36	50.00	47.94	46.36	54.54	<b>46.31</b>
Vehicle	MSE	2.94E-2	2.38E-2	2.86E-1	8.99E-2	6.34E-2	5.66E-2	5.69E-2	5.62E-2	5.03E-2	5.54E-2	3.72E-2	<b>1.81E-2</b>
	ARV	4.44	3.26E+1	1.08	1.31	<b>9.97E-1</b>	1.22	1.18	1.21	1.22	1.19	1.42	1.45
	MAE	1.23E-1	1.15E-1	5.14E-1	2.63E-1	2.15E-1	2.15E-1	2.15E-1	2.14E-1	2.02E-1	2.12E-1	1.66E-1	<b>1.09E-1</b>
	SMAPE	16.71	15.82	97.24	40.89	34.55	30.77	30.80	30.84	28.91	30.57	23.56	<b>15.14</b>
Wine	MSE	<b>2.69E-3</b>	3.15E-3	4.60E-3	1.75E-2	7.88E-3	2.46E-2	4.04E-2	3.80E-2	2.58E-2	3.79E-2	3.99E-2	2.94E-3
	ARV	8.76	6.71	9.43E-1	<b>1.06</b>	1.41	1.09	1.07	1.07	1.08	1.07	<b>1.06</b>	1.39
	MAE	<b>3.32E-2</b>	4.48E-2	5.50E-2	1.12E-1	8.25E-2	1.50E-1	1.95E-1	1.88E-1	1.53E-1	1.88E-1	1.94E-1	3.54E-2
	SMAPE	44.77	54.29	93.91	100.50	77.54	102.30	113.68	111.88	103.13	112.01	113.60	<b>43.84</b>

Tabela 8 – Diferença Percentual (%) entre o DSNAW e as abordagens da literatura calculado por meio da equação 4.5 em relação aos resultados do MSE

Séries Temporais	ARIMA	ETS	SVR	DS-LA (1)	TWF	$Bagg_A$	$Bagg_M$	DS-LA	$DES_A$	$DES_M$	DES-PALR
Amazon	97.87	97.85	99.61	98.60	21.97	94.50	88.01	98.20	98.59	98.20	99.40
Apple	30.58	30.41	86.19	88.85	72.42	84.92	87.25	88.29	87.60	88.46	88.89
Electricity	82.40	75.19	98.03	92.14	38.68	81.75	76.07	89.96	87.97	89.65	94.32
Goldman	98.58	97.84	99.06	88.87	34.61	42.89	25.68	82.83	81.59	83.74	88.67
Microsoft	96.69	96.69	91.35	87.26	84.75	62.13	67.14	87.26	82.88	85.43	88.86
Pollution	41.91	6.26	64.56	55.15	61.54	2.06	58.85	58.28	56.73	57.71	61.06
Star	99.96	99.92	21.05	-8.47	4.44	-28.55	-28.81	-27.96	-28.55	-28.81	1.04
Sunspot	47.22	57.20	-24.69	19.66	-5.03	-1.65	8.30	5.79	-1.65	8.30	16.49
Vehicle	38.45	24.21	93.67	79.91	71.48	68.06	68.23	67.83	64.07	67.40	51.39
Wine	-9.52	6.71	36.02	83.18	62.65	88.03	92.72	92.25	88.60	92.23	92.62

A Tabela 8 mostra a diferença percentual (calculada por meio da equação 4.5) de desempenho entre o DSNAW e as abordagens da literatura considerando os valores do MSE. Em relação às abordagens monolíticas, ARIMA, ETS e SVR, o DSNAW apresenta um ganho de desempenho na maioria das séries temporais avaliadas, perdendo apenas para o ARIMA na série *Wine* e para o SVR na série *Sunspot*. Em relação às abordagens de seleção dinâmica de preditores, DS-LA (1) e TWF, o método proposto apresenta um ganho de desempenho na maioria das séries e tem um desempenho inferior aos métodos DS-LA (1) e TWF nas séries *Star* e *Sunspot*, respectivamente. Em relação às abordagens que combinam todas as previsões do *Pool*:  $Bagg_A$  e  $Bagg_M$ , o método proposto apresentou um ganho de desempenho em 8 séries em relação ao método  $Bagg_A$  e um ganho em 9 séries em relação ao método  $Bagg_M$ . Em relação aos métodos de seleção dinâmica de *Ensemble*, o método proposto obteve um ganho de desempenho em 9 séries em comparação aos métodos DS-LA e  $DES_M$ , obteve um ganho em 8 séries em comparação ao  $DES_A$  e um ganho em todas as séries avaliadas em relação ao DES-PALR. Ao calcularmos a mediana dos ganhos percentuais em relação às abordagens da literatura para cada série temporal, temos os seguintes valores: 98% na série *Amazon*, 88% na série *Wine*, 87% nas séries *Apple*, *Electricity* e *Microsoft*, 83% na *Goldman*, 67% na série *Vehicle*, 57% na série *Pollution*, 8% na série *Sunspot* e -8% na série *Star*. Com base nesses resultados, podemos observar que o DSNAW obteve em mediana um ganho percentual positivo em 9 de 10 séries temporais.

A Tabela 9 mostra os resultados do teste estatístico DM (valor crítico  $\alpha = 0.05$ ) que compara o desempenho do DSNAW em relação aos métodos da literatura na perspectiva da métrica MSE. Os símbolos “+”, “-” e “ $\approx$ ” correspondem a desempenhos superior, pior e igual, respectivamente. As últimas três linhas da tabela apresentam um resumo da comparação, no qual Vitórias corresponde à quantidade de séries em que o método proposto obteve desempenho

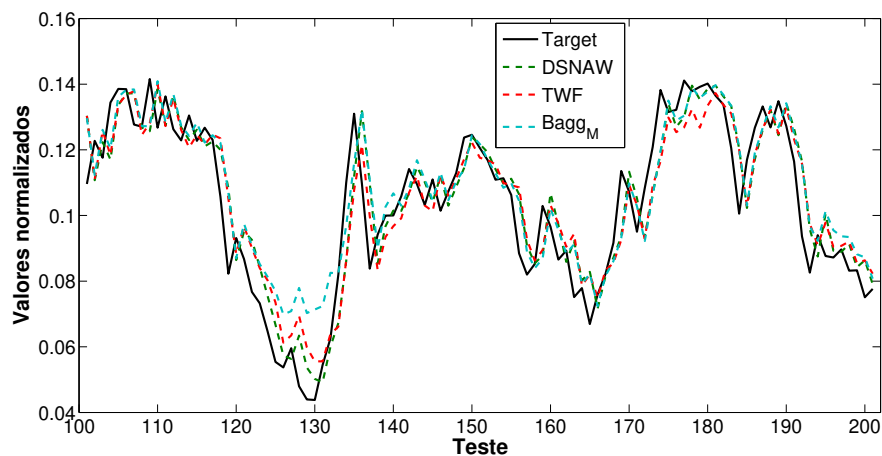


Tabela 9 – Resultado do teste estatístico DM (valor crítico  $\alpha = 0.05$ ) que compara o desempenho do DSNAW em relação aos métodos da literatura na perspectiva da métrica MSE. Os símbolos “+”, “-” e “≈” corresponde ao desempenho superior, pior e igual, respectivamente.

Time Series	ARIMA	ETS	SVR	DS-LA (1)	TWF	$Bagg_A$	$Bagg_M$	DS-LA	$DES_A$	$DES_M$	DES-PALR
Amazon	+	+	+	+	+	+	+	+	+	+	+
Apple	+	+	+	+	+	+	+	+	+	+	+
Electricity	+	+	+	+	+	+	+	+	+	+	+
Goldman	+	+	+	+	+	+	+	+	+	+	+
Microsoft	+	+	+	+	+	+	+	+	+	+	+
Pollution	≈	≈	+	+	+	≈	+	+	+	+	+
Star	+	+	≈	≈	≈	-	-	≈	-	-	≈
Sunspot	+	+	≈	≈	≈	≈	≈	≈	≈	≈	≈
Vehicle	≈	≈	+	+	+	+	+	+	+	+	+
Wine	≈	≈	≈	+	+	+	+	+	+	+	+
Vitórias	7	7	7	8	8	7	8	8	8	8	8
Empates	3	3	3	2	2	2	1	2	1	1	2
Derrotas	0	0	0	0	0	1	1	0	1	1	0

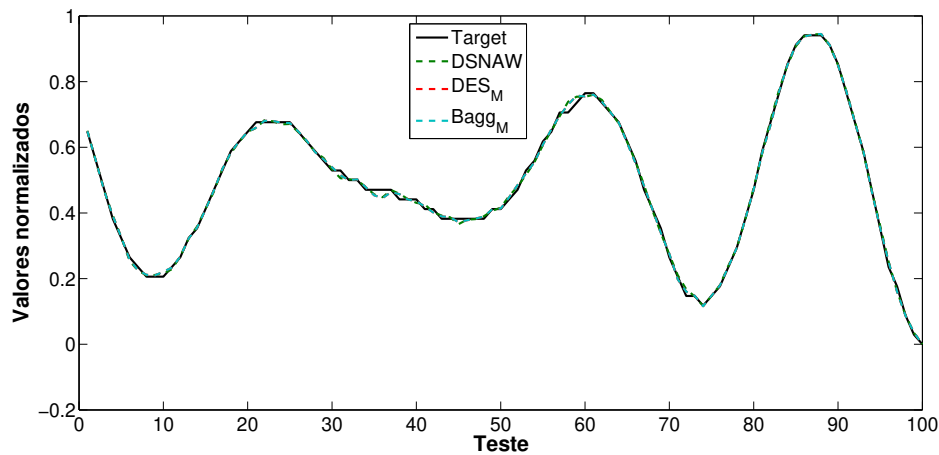
estatisticamente superior ao respectivo método da literatura; Empates é a quantidade de séries temporais em que o método proposto e o respectivo método da literatura não apresentam desempenhos estatisticamente diferentes, enquanto que Derrotas é a quantidade de séries em que o DSNAW apresentou desempenho estatisticamente inferior ao respectivo método da literatura. O método proposto apresentou um desempenho estatisticamente igual ou superior às abordagens da literatura em 9 de 10 séries temporais. Na série *Star*, o método proposto obteve desempenho inferior em comparação às abordagens  $Bagg_A$ ,  $Bagg_M$ ,  $DES_A$  e  $DES_M$ . Considerando as 110 (10 Séries Temporais x 11 abordagens) comparações diretas, o método proposto foi melhor em 84 casos, empatou em 22 e foi pior em 4.

Figura 17 – Previsões para o intervalo de 100 até 200 observações do conjunto de teste da série *Amazon*, utilizando o DSNAW e as abordagens da literatura que apresentaram os melhores resultados.



Fonte: o autor (2021)

Figura 18 – Previsões para as 100 primeiras observações do conjunto de teste da série *Star*, utilizando o DSNAW e as abordagens da literatura que apresentaram os melhores resultados.



Fonte: o autor (2021)

As Figuras 17 e 18 mostram as previsões do DSNAW e das duas abordagens que apresentam os melhores desempenho na série em que o método proposto obteve o melhor desempenho (*Amazon*) e na série em que o DSNAW obteve o pior desempenho (*Star*) com base nos resultados apresentados. Para a série *Star*, apesar de o método proposto ter apresentado desempenho inferior, as previsões realizadas são próximas do valor desejado e também das previsões das melhores abordagens. Na série *Amazon* é possível observar que as previsões do DSNAW estão mais próximas do *Target* que as previsões obtidas por meio das abordagens TWF e *Bagg<sub>M</sub>*.

Os resultados apresentados na Tabela 7 demonstram que as abordagens que utilizam Sistemas de Múltiplos Preditores (SMP) tendem a obter desempenho superior às abordagens monolíticas, corroborando com os resultados encontrados na literatura (YAO; DAI; SONG, 2018; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; BUDI; AJI; WIDODO, 2014; OLIVEIRA; SILVA; NETO, 2021).

A partir de uma análise em relação às abordagens de SMP, é possível verificar que as abordagens de seleção dinâmica de preditores e de *Ensemble* da literatura tendem a ter uma maior variabilidade de desempenho em comparação às abordagens que combinam todas as previsões do *Pool*, sendo que podemos inferir que tal variabilidade se relaciona à maneira com a qual os modelos são selecionados, utilizando uma Região de Competência (RoC) formada por padrões de treinamento ou validação semelhante ao novo padrão de teste (YAO; DAI; SONG, 2018; YANKOV; DECOSTE; KEOGH, 2006; WIDODO; BUDI, 2011; BUDI; AJI; WIDODO,

2014; OLIVEIRA; SILVA; NETO, 2021). Porém, não existe garantia de que a RoC será composta por padrões de treinamento que tenham um comportamento similar ao novo padrão de teste, seja por conta de ruído nos dados ou por alterações no processo gerador da série temporal, que pode mudar no decorrer do tempo.

Porém, o método proposto cria uma RoC utilizando os padrões que são temporalmente mais próximos do padrão de teste. O objetivo é aumentar a chance de selecionar preditores em um RoC formada por padrões que tenham comportamentos mais próximos do novo padrão de teste, como apresentado na seção 3.1 e, consequentemente, selecionar preditores com previsões mais acuradas para a observação desconhecida.

Os resultados apresentados nas Tabelas 7, 8 e 9 mostram que a estratégia apresentada no método proposto tende a ter desempenho superior aos métodos da literatura na maioria das séries temporais avaliadas e em diferentes métricas. A Tabela 8, além de mostrar o ganho de desempenho ao utilizar o DSNaw, revela que o método proposto obteve os melhores resultados nas séries *Amazon* e *Wine* com um ganho mediano (em relação às abordagens da literatura) de 98% e 92%, respectivamente, já os piores resultados foram obtidos nas séries *Star* e *Sunspot* com um ganho mediano de -8.4% e 8.3%, respectivamente. Em relação ao comportamento das séries, é possível observar que os melhores casos do DSNaw ocorreram em séries não estacionárias, enquanto que os piores casos se deram em série estacionária.

Ao compararmos esse resultado com os resultados apresentados na Figura 4, podemos observar que para a série *Amazon*, enquanto a RoC formada pelo método proposto apresenta um percentual próximo de 100% de janelas com a mesma distribuição do *out-of-sample*, a RoC formada pela literatura apresenta apenas um percentual próximo de 20%, enquanto que na *Star* ambas abordagens de construção de RoC apresentam um percentual próximo de 100%. Essa possível associação entre o ganho percentual e o percentual de janelas na RoC com a mesma distribuição demonstra que podemos aplicar o experimento realizado na seção 3.1 utilizando o teste estatístico KS para selecionar entre uma RoC formada seguindo a literatura e uma RoC formada com base no DSNaw. Para tal, quando o resultado do experimento aplicado seguindo os passos descritos na seção 3.1 mostrar que ambas as RoC apresentam um mesmo percentual de janelas com a mesma distribuição, então a RoC formada pela literatura pode ter um desempenho superior (é o caso da série *Star*), caso contrário, o DSNaw tende a ter um desempenho superior (é o caso da série *Amazon*).

## 5.7 CONSIDERAÇÕES FINAIS

Este capítulo apresentou os resultados obtidos nos experimentos realizados. Primeiramente, foi realizada uma comparação entre os parâmetros do DSNAW selecionados por meio do conjunto de validação em comparação aos parâmetros que apresentaram os melhores resultados em teste. A partir dessa comparação, foi possível observar que apesar de não ser possível selecionar as melhores combinações de parâmetros no conjunto de validação, os resultados obtidos foram próximos na maioria das séries temporais.

Em seguida, foi apresentada uma comparação entre o método proposto e as abordagens da literatura organizadas de acordo com os seguintes grupos: monolíticos (seção 5.2), seleção dinâmica de preditores (seção 5.3), *Ensemble* (seção 5.4) e seleção dinâmica de *Ensemble* (seção 5.5). Em todos os casos, o método proposto apresentou desempenho superior às abordagens da literatura na maioria das séries temporais e por meio de diferentes métricas.

Na última análise apresentada (seção 5.6), foram apresentados os ganhos percentuais do método proposto em relação às abordagens da literatura. Na maioria das séries temporais, o DSNAW apresentou um ganho percentual positivo, obtendo o melhor resultado na série e o pior resultado na série *Star*. Ainda nessa última análise, foi aplicado o teste estatístico Diebold Mariano (DM) para verificar se o DSNAW apresenta previsões estatisticamente diferentes das abordagens da literatura. O teste foi aplicado em 100 comparação em pares, utilizando as previsões do DSNAW e um respectivo método da literatura. Dentre estas, o método proposto foi o melhor em 84 casos, empatou em 22 e foi pior em 4. Por fim, foi realizada uma análise associando os ganhos percentuais aos resultados obtidos no experimento apresentado na 3.1. Essa análise permite inferir que o desempenho do método de seleção dinâmica se relaciona à ocorrência de padrões na região de competência com o mesmo comportamento do novo padrão.

## 6 CONCLUSÃO

Este capítulo apresenta as respostas para as questões de pesquisa que guiaram o desenvolvimento desta tese, bem como o resumo do trabalho, as principais contribuições, limitações, trabalhos futuros e trabalhos aprovados.

### 6.1 RESUMO DO TRABALHO

O desenvolvimento de sistemas de previsão acurados para modelagem de séries temporais do mundo real é uma tarefa desafiadora. Devido à presença de padrões temporais que mudam ao longo do tempo, a adoção de um modelo único pode gerar previsões de baixa acurácia. Nesse cenário, Sistemas de Múltiplo Preditores (SMP) surgem como uma alternativa promissora. A seleção dinâmica de preditores é um caso especial de SMP, no qual cada modelo é um especialista em padrões específicos da série temporal. Na seleção dinâmica, os modelos mais competentes são selecionados para cada novo padrão de teste. Frequentemente emprega-se o critério de avaliar o desempenho dos modelos na região de competência, formada pelos padrões (presentes nos conjuntos de treinamento ou validação) mais similares ao padrão de teste. A qualidade da região de competência é um fator-chave na precisão do SMP, entretanto, definir adequadamente a medida de similaridade e o tamanho da região de competência representa um desafio. Além disso, não há garantia de que existam padrões similares no conjunto de treinamento ou validação. Esta tese propõe uma abordagem de seleção dinâmica intitulada *Dynamic Selection based on the Nearest Windows* (DSNAW), que escolhe um ou mais modelos de acordo com seu desempenho na região de competência composta pelas janelas que antecedem à nova janela da observação de teste. Essa estratégia assume que as janelas temporais mais próximas do novo padrão de teste têm um comportamento mais similar à observação a ser predita que os dados de treinamento. O estudo experimental utilizando dez séries temporais mostrou que a abordagem proposta supera as abordagens da literatura, consolidando-a como uma nova abordagem para seleção dinâmica de preditores.

## 6.2 QUESTÕES DE PESQUISA

Nesta seção são apresentadas as respostas para as questões de pesquisa com base nos resultados introduzidos no capítulo anterior, levando em consideração o conjunto de métricas e base de dados utilizadas neste trabalho.

### 6.2.1 Uma região de competência composta por janelas provenientes do conjunto de treinamento possui a mesma distribuição de um dado novo padrão de teste?

Com base nos resultados obtidos, é possível observar que na maioria dos casos uma região de competência formada por janelas de treinamento ou validação tende a ter uma baixa ocorrência de janelas com a mesma distribuição do novo padrão. Apenas em séries temporais como *Star* e *Sunspot* esse tipo de abordagem tende a apresentar bons resultados. Em comum, essas duas séries têm o comportamento estacionário.

Esta pesquisa permite aplicar o experimento descrito na Seção 3.1 para avaliar em quais casos a RoC formada por janelas de treinamento será composta por padrões com a mesma distribuição do novo padrão.

### 6.2.2 Uma região de competência composta por padrões mais recentes possui uma maior quantidade de janelas com a mesma distribuição do padrão de teste que regiões de competência usadas na literatura?

A partir da análise experimental, é possível observar que a RoC formada por janelas que antecedem o ponto a ser predito, além de apresentar uma maior ocorrência de padrões com a mesma distribuição, também seleciona modelos com previsões mais acuradas na maioria das séries temporais avaliadas, principalmente nas séries não estacionárias.

## 6.3 PRINCIPAIS CONTRIBUIÇÕES DA TESE

Os resultados obtidos por meio do método proposto tornam válida a proposição de que uma região de competência formada por janelas que antecedem o padrão a ser predito tende a apresentar o mesmo comportamento e consequentemente aumenta a chance de selecionar

preditores com previsões acuradas. Contudo, podemos destacar as seguintes contribuições da tese:

- Abordagem para criação da região de competência composta por padrões temporalmente próximos da observação a ser predita.
- Define de forma automática o tamanho da região de competência e o número de preditores a serem selecionados, podendo assim ser aplicado para seleção dinâmica de preditores ou seleção dinâmica de *Ensemble*.
- Quando for selecionado mais de um preditor, a abordagem proposta é capaz de escolher de forma automática a melhor estratégia de combinação.

Em resumo, a abordagem emprega uma nova maneira de definir a região de competência e de escolher a abordagem mais adequada entre Seleção Dinâmica de Preditor e Seleção Dinâmica de *Ensemble*, para cada série temporal.

#### 6.4 LIMITAÇÕES DA PESQUISA

Entre as limitações deste estudo está o uso de um SMP com 100 SVRs treinadas durante os experimentos utilizando a abordagem de *Bagging*, por essa razão, os resultados obtidos podem estar limitados a um SMP com as mesmas características. Consequentemente, um SMP composto por um conjunto de preditores de algoritmos diferentes, como, por exemplo, MLP, LSTM, RBF, SVR, pode apresentar respostas diferentes para as questões levantadas nesta pesquisa.

Outra limitação consiste na quantidade e na variedade de séries temporais. O conjunto de 10 séries temporais utilizadas na pesquisa é apenas uma amostra da população de séries temporais que existem no mundo real. Consequentemente, os resultados obtidos nesta pesquisa são limitados às séries temporais estudadas ou àquelas com características similares.

#### 6.5 TRABALHOS FUTUROS

Uma continuação imediata deste trabalho será uma análise do método DSNAW em um conjunto de preditores heterogêneos, composto por diferentes modelos treinados na mesma amostra de dados. Esse tipo de análise pode vir a mostrar se o desempenho do DSNAW se

associa à forma de geração do *Pool* ou se trata-se uma abordagem independente do conjunto de preditores.

Outro ponto a ser avaliado é o desempenho do DSNaw em séries temporais com comportamentos estacionários. Os piores resultados do DSNaw apresentados neste trabalho ocorrem em séries estacionárias. Portanto, é válida a realização de experimentos para compreender se esse tipo de comportamento seria uma limitação do DSNaw.

Em relação ao SMP apresentado no método proposto, é possível a utilização de novas abordagens de combinação e de outros tipos de modelos-base para a construção do *Pool*, tais como LSTM, RBF e MLP. Além disso, os parâmetros deste trabalho foram selecionados a partir de um intervalo limitado entre 1 e 20, que pode ser alterado, sendo que os valores para os parâmetros podem ser então selecionados por meio de algoritmos de otimização.

## 6.6 ARTIGOS PUBLICADOS

Os trabalhos listados a seguir são resultados da pesquisa realizada durante o doutorado:

- "On the evaluation of dynamic selection parameters for time series forecasting." (SILVA et al., 2020).
- "A hybrid system based on dynamic selection for time series forecasting"(OLIVEIRA; SILVA; NETO, 2021).
- "A dynamic predictor selection method based on recent temporal windows for time series forecasting"(SILVA; NETO; CAVALCANTI, 2021).



## REFERÊNCIAS

- ABDOLLAHI, H. A novel hybrid model for forecasting crude oil price based on time series decomposition. *Applied Energy*, Elsevier, v. 267, p. 115035, 2020.
- ADHIKARI, R. A neural network based linear ensemble framework for time series forecasting. *Neurocomputing*, Elsevier, v. 157, p. 231–242, 2015.
- ADHIKARI, R.; AGRAWAL, R. Combining multiple time series models through a robust weighted mechanism. In: IEEE. *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*. [S.l.], 2012. p. 455–460.
- ADHIKARI, R.; VERMA, G.; KHANDELWAL, I. A model ranking based selective ensemble approach for time series forecasting. *Procedia Computer Science*, Elsevier, v. 48, p. 14–21, 2015.
- AHMED, N. K.; ATIYA, A. F.; GAYAR, N. E.; EL-SHISHINY, H. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, Taylor & Francis, v. 29, n. 5-6, p. 594–621, 2010.
- AKÇA, E.; YOZGATLIGIL, C. Mutual information model selection algorithm for time series. *Journal of Applied Statistics*, Taylor & Francis, p. 1–16, 2020.
- ALKANDARI, M.; AHMAD, I. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Applied Computing and Informatics*, Emerald Publishing Limited, 2020.
- BARROW, D. K.; CRONE, S. F. A comparison of adaboost algorithms for time series forecast combination. *International Journal of Forecasting*, Elsevier, v. 32, n. 4, p. 1103–1119, 2016.
- BONTEMPI, G.; TAIEB, S. B.; BORGNE, Y.-A. L. Machine learning strategies for time series forecasting. In: SPRINGER. *European business intelligence summer school*. [S.l.], 2012. p. 62–77.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. [S.l.]: springer, 2016.
- BROCKWELL, P. J.; DAVIS, R. A.; FIENBERG, S. E. *Time series: theory and methods: theory and methods*. [S.l.]: Springer Science & Business Media, 1991.
- BUDI, I.; AJI, R. F.; WIDODO, A. Prediction of research topics using ensemble of best predictors from similar dataset. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, v. 8, n. 1, 2014.
- CANG, S.; YU, H. A combination selection algorithm on forecasting. *European Journal of Operational Research*, Elsevier, v. 234, n. 1, p. 127–139, 2014.
- CAO, L.; GU, Q. Dynamic support vector machines for non-stationary time series forecasting. *Intelligent Data Analysis*, IOS Press, v. 6, n. 1, p. 67–83, 2002.

- CHATFIELD, C. *The analysis of time series: an introduction*. [S.l.]: Chapman and Hall/CRC, 2003.
- CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Prototype selection for dynamic classifier and ensemble selection. *Neural Computing and Applications*, Springer, v. 29, n. 2, p. 447–457, 2018.
- CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D.; REN, T. I. Meta-des: A dynamic ensemble selection framework using meta-learning. *Pattern recognition*, Elsevier, v. 48, n. 5, p. 1925–1935, 2015.
- DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business & economic statistics*, Taylor & Francis, v. 20, n. 1, p. 134–144, 2002.
- DOLADO, J. J.; GONZALO, J.; MAYORAL, L. A fractional dickey–fuller test for unit roots. *Econometrica*, Wiley Online Library, v. 70, n. 5, p. 1963–2006, 2002.
- DRUCKER, H.; BURGESS, C. J.; KAUFMAN, L.; SMOLA, A. J.; VAPNIK, V. Support vector regression machines. In: *Advances in neural information processing systems*. [S.l.: s.n.], 1997. p. 155–161.
- EHSAN, R. M.; SIMON, S. P.; VENKATESWARAN, P. Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. *Neural Computing and Applications*, Springer, v. 28, n. 12, p. 3981–3992, 2017.
- ELWELL, R.; POLIKAR, R. Incremental learning of concept drift in nonstationary environments. *IEEE Transactions on Neural Networks*, IEEE, v. 22, n. 10, p. 1517–1531, 2011.
- GONZÁLEZ, J. P.; ROQUE, A. M. S.; PÉREZ, E. A. Forecasting functional time series with a new hilbertian ARMAX model: Application to electricity price forecasting. *IEEE Transactions on Power Systems*, IEEE, v. 33, n. 1, p. 545–556, 2018.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural computation*, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.
- HONG-ZHI, A.; BING, C. A kolmogorov-smirnov type statistic with application to test for nonlinearity in time series. *International Statistical Review/Revue Internationale de Statistique*, JSTOR, p. 287–307, 1991.
- HU, T.; LAM, K.; NG, S. River flow time series prediction with a range-dependent neural network. *Hydrological Sciences Journal*, Taylor & Francis, v. 46, n. 5, p. 729–745, 2001.
- HUANG, Q.; MAO, J.; LIU, Y. An improved grid search algorithm of svr parameters optimization. In: IEEE. *2012 IEEE 14th International Conference on Communication Technology*. [S.l.], 2012. p. 1022–1026.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018.
- JUNIOR, S. L. D.; CECATTO, J. R.; FERNANDES, M. M.; RIBEIRO, M. X. An optimized data mining method to support solar flare forecast. *Information Technology-New Generations*, Springer, p. 467–474, 2018.

- KARASU, S.; ALTAN, A.; BEKIROU, S.; AHMAD, W. A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy*, Elsevier, v. 212, p. 118750, 2020.
- KHAN, M. Y.; MITTNIK, S. Nonlinear time series modeling and forecasting the seismic data of the hindu kush region. *Journal of Seismology*, Springer, p. 1–24, 2017.
- KOURENTZES, N.; BARROW, D. K.; CRONE, S. F. Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, Elsevier, v. 41, n. 9, p. 4235–4244, 2014.
- KOWALSKI, A. M.; MARTIN, M. T.; PLASTINO, A.; JUDGE, G. On extracting probability distribution information from time series. *Entropy*, Molecular Diversity Preservation International, v. 14, n. 10, p. 1829–1841, 2012.
- KRIKUNOV, A. V.; KOVALCHUK, S. V. Dynamic selection of ensemble members in multi-model hydrometeorological ensemble forecasting. *Procedia Computer Science*, Elsevier, v. 66, p. 220–227, 2015.
- LAHMIRI, S. Intraday stock price forecasting based on variational mode decomposition. *Journal of Computational Science*, Elsevier, v. 12, p. 23–27, 2016.
- LEMKE, C.; GABRYS, B. Meta-learning for time series forecasting and forecast combination. *Neurocomputing*, Elsevier, v. 73, n. 10-12, p. 2006–2016, 2010.
- LIANG, R.; JIA, P. Coal mine gas concentration prediction model based on selective ensemble learning. In: IEEE. *2018 Chinese Control And Decision Conference (CCDC)*. [S.l.], 2018. p. 3789–3793.
- LILLIEFORS, H. W. On the kolmogorov-smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 64, n. 325, p. 387–389, 1969.
- LIVIERIS, I. E.; PINTELAS, E.; STAVROYIANNIS, S.; PINTELAS, P. Ensemble deep learning models for forecasting cryptocurrency time-series. *Algorithms*, Multidisciplinary Digital Publishing Institute, v. 13, n. 5, p. 121, 2020.
- MA, X.; TAO, Z.; WANG, Y.; YU, H.; WANG, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, Elsevier, v. 54, p. 187–197, 2015.
- MA, Z.; DAI, Q. Selected an stacking elms for time series prediction. *Neural Processing Letters*, Springer, v. 44, n. 3, p. 831–856, 2016.
- MA, Z.; DAI, Q.; LIU, N. Several novel evaluation measures for rank-based ensemble pruning with applications to time series prediction. *Expert Systems with Applications*, Elsevier, v. 42, n. 1, p. 280–292, 2015.
- MAO, S.; XIAO, F. A novel method for forecasting construction cost index based on complex network. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 527, p. 121306, 2019.
- MEADE, N. Evidence for the selection of forecasting methods. *Journal of forecasting*, Wiley Online Library, v. 19, n. 6, p. 515–535, 2000.

- MENDES-MOREIRA, J.; SOARES, C.; JORGE, A. M.; SOUSA, J. F. D. Ensemble approaches for regression: A survey. *ACM Computing Surveys (CSUR)*, ACM, v. 45, n. 1, p. 10, 2012.
- METCALFE, A. V.; COWPERTWAIT, P. S. *Introductory time series with R*. [S.l.]: Springer, 2009.
- MITCHELL, T. M. et al. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, v. 45, n. 37, 1997.
- MORA-LÓPEZ, L.; MORA, J. An adaptive algorithm for clustering cumulative probability distribution functions using the kolmogorov-smirnov two-sample test. *Expert Systems with Applications*, Elsevier, v. 42, n. 8, p. 4016–4021, 2015.
- MORETTI, F.; PIZZUTI, S.; PANZIERI, S.; ANNUNZIATO, M. Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. *Neurocomputing*, Elsevier, v. 167, p. 3–7, 2015.
- MUTAVHATSINDI, T.; SIGAUKE, C.; MBUVHA, R. Forecasting hourly global horizontal solar irradiance in south africa using machine learning models. *IEEE Access*, IEEE, v. 8, p. 198872–198885, 2020.
- NADEEM, F.; ALGHAZZAWI, D.; MASHAT, A.; FAQEEH, K.; ALMALAISE, A. Using machine learning ensemble methods to predict execution time of e-science workflows in heterogeneous distributed systems. *IEEE Access*, IEEE, v. 7, p. 25138–25149, 2019.
- NETO, P. S. M.; CAVALCANTI, G. D.; FIRMINO, P. R.; SILVA, E. G.; FILHO, S. R. V. N. A temporal-window framework for modelling and forecasting time series. *Knowledge-Based Systems*, Elsevier, v. 193, p. 105476, 2020.
- OLIVEIRA, J. F. de; LUDERMIR, T. B. A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*, Elsevier, v. 180, p. 27–34, 2016.
- OLIVEIRA, J. F. de; SILVA, E. G.; NETO, P. S. de M. A hybrid system based on dynamic selection for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, 2021.
- OLIVEIRA, J. F. L. de; LUDERMIR, T. B. A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting. In: IEEE. *2014 Brazilian Conference on Intelligent Systems*. [S.l.], 2014. p. 73–78.
- OLIVEIRA, R. T. de; ASSIS, T. F. O. de; FIRMINO, P. R. A.; FERREIRA, T. A. Copulas-based time series combined forecasters. *Information Sciences*, Elsevier, v. 376, p. 110–124, 2017.
- PARZEN, E. Some recent advances in time series modeling. *IEEE Transactions on Automatic Control*, IEEE, v. 19, n. 6, p. 723–730, 1974.
- PRUDÊNCIO, R. B.; LUDERMIR, T. B. Meta-learning approaches to selecting time series models. *Neurocomputing*, Elsevier, v. 61, p. 121–137, 2004.
- QIU, X.; REN, Y.; SUGANTHAN, P. N.; AMARATUNGA, G. A. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, Elsevier, v. 54, p. 246–255, 2017.

- RAFTERY, A. E.; GNEITING, T.; BALABDAOUI, F.; POLAKOWSKI, M. Using bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, v. 133, n. 5, p. 1155–1174, 2005.
- RAHMAN, M. M.; ISLAM, M. M.; MURASE, K.; YAO, X. Layered ensemble architecture for time series forecasting. *IEEE transactions on cybernetics*, IEEE, v. 46, n. 1, p. 270–283, 2015.
- ROONEY, N.; PATTERSON, D.; ANAND, S.; TSYMBAL, A. Dynamic integration of regression models. In: SPRINGER. *International Workshop on Multiple Classifier Systems*. [S.I.], 2004. p. 164–173.
- RUBIO, J. de J.; ELIAS, I.; CRUZ, D. R.; PACHECO, J. Uniform stable radial basis function neural network for the prediction in two mechatronic processes. *Neurocomputing*, Elsevier, v. 227, p. 122–130, 2017.
- SACCHI, R.; OZTURK, M. C.; PRINCIPE, J. C.; CARNEIRO, A. A.; SILVA, I. N. D. Water inflow forecasting using the echo state network: A brazilian case study. *International Joint Conference on Neural Networks (IJCNN)*, IEEE, p. 2403–2408, 2007.
- SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 26, n. 1, p. 43–49, fev. 1978. ISSN 0096-3518.
- SAMARA, S.; NATSHEH, E. Modeling the output power of heterogeneous photovoltaic panels based on artificial neural networks using low cost microcontrollers. *Heliyon*, Elsevier, v. 4, n. 11, p. e00972, 2018.
- SEABOLD, S.; PERKTOLD, J. Statsmodels: Econometric and statistical modeling with python. In: AUSTIN, TX. *Proceedings of the 9th Python in Science Conference*. [S.I.], 2010. v. 57, p. 61.
- SILVA, E. G.; CAVALCANTI, G. D.; OLIVEIRA, J. F. L. de; NETO, P. S. de M. On the evaluation of dynamic selection parameters for time series forecasting. In: IEEE. *2020 International Joint Conference on Neural Networks (IJCNN)*. [S.I.], 2020. p. 1–7.
- SILVA, E. G.; DOMINGOS, S. d. O.; CAVALCANTI, G. D.; NETO, P. S. de M. Improving the accuracy of intelligent forecasting models using the perturbation theory. In: IEEE. *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.I.], 2018. p. 1–7.
- SILVA, E. G.; NETO, P. S. D. M.; CAVALCANTI, G. D. A dynamic predictor selection method based on recent temporal windows for time series forecasting. *IEEE Access*, IEEE, 2021.
- SIN, E.; WANG, L. Bitcoin price prediction using ensembles of neural networks. In: IEEE. *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. [S.I.], 2017. p. 666–671.
- SMITH, T. G. et al. pmdarima: Arima estimators for python. *Retrieved from*, v. 309, 2017.
- SUDHEER, C.; MAHESWARAN, R.; PANIGRAHI, B. K.; MATHUR, S. A hybrid SVM-PSO model for forecasting monthly streamflow. *Neural Computing and Applications*, Springer, v. 24, n. 6, p. 1381–1389, 2014.

- SUN, G.; JIANG, C.; CHENG, P.; LIU, Y.; WANG, X.; FU, Y.; HE, Y. Short-term wind power forecasts by a synthetical similar time series data mining method. *Renewable Energy*, Elsevier, v. 115, p. 575–584, 2018.
- SUN, J.; SHEN, Z.; LI, H.; SHEN, Y. Clustering via local regression. In: SPRINGER. *Joint European conference on machine learning and knowledge discovery in databases*. [S.l.], 2008. p. 456–471.
- TILLY, S.; EBNER, M.; LIVAN, G. Macroeconomic forecasting through news, emotions and narrative. *Expert Systems with Applications*, Elsevier, v. 175, p. 114760, 2021.
- TSOUMAKAS, G.; PARTALAS, I.; VLAHAVAS, I. A taxonomy and short review of ensemble selection. In: *Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications*. [S.l.: s.n.], 2008. p. 1–6.
- VALENTINI, G.; DIETTERICH, T. G. Bias-variance analysis of support vector machines for the development of svm-based ensemble methods. *Journal of Machine Learning Research*, v. 5, n. Jul, p. 725–775, 2004.
- WANG, J.; ZHANG, N.; LU, H. A novel system based on neural networks with linear combination framework for wind speed forecasting. *Energy Conversion and Management*, Elsevier, v. 181, p. 425–442, 2019.
- WANG, L.; WANG, Z.; QU, H.; LIU, S. Optimal forecast combination based on neural networks for time series forecasting. *Applied Soft Computing*, Elsevier, v. 66, p. 1–17, 2018.
- WIDODO, A.; BUDI, I. Model selection for time series forecasting using similarity measure. In: IEEE. *2011 International Conference on Advanced Computer Science and Information Systems*. [S.l.], 2011. p. 221–226.
- WOLPERT, D. H. The lack of a priori distinctions between learning algorithms. *Neural computation*, MIT Press, v. 8, n. 7, p. 1341–1390, 1996.
- WOODS, K.; KEGELMEYER, W. P.; BOWYER, K. Combination of multiple classifiers using local accuracy estimates. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 19, n. 4, p. 405–410, 1997.
- WU, J.; CHENG, E. A novel hybrid particle swarm optimization for feature selection and kernel optimization in support vector regression. In: IEEE. *2010 International Conference on Computational Intelligence and Security*. [S.l.], 2010. p. 189–194.
- WU, J.; XIE, Y. Hybrid support vector regression with parallel co-evolution algorithm based on ga and pso for forecasting monthly rainfall. *Journal of Software Engineering and Applications*, Scientific Research Publishing, v. 12, n. 12, p. 524–539, 2019.
- XIN, N.; GU, X.; WU, H.; HU, Y.; YANG, Z. Application of genetic algorithm-support vector regression (ga-svr) for quantitative analysis of herbal medicines. *Journal of Chemometrics*, Wiley Online Library, v. 26, n. 7, p. 353–360, 2012.
- YANKOV, D.; DECOSTE, D.; KEOGH, E. Ensembles of nearest neighbor forecasts. In: SPRINGER. *European Conference on Machine Learning*. [S.l.], 2006. p. 545–556.
- YAO, C.; DAI, Q.; SONG, G. Several novel dynamic ensemble selection algorithms for time series prediction. *Neural Processing Letters*, Springer, p. 1–41, 2018.

- 
- YOSHIDA, S.; CHAKRABORTY, B. A comparative study of similarity measures for time series classification. In: SPRINGER. *JSAI International Symposium on Artificial Intelligence*. [S.l.], 2015. p. 397–408.
- ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, Elsevier, 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*, Elsevier, v. 160, n. 2, p. 501–514, 2005.
- ZHANG, S.; CHEN, Y.; ZHANG, W.; FENG, R. A novel ensemble deep learning model with dynamic error correction and multi-objective ensemble pruning for time series forecasting. *Information Sciences*, Elsevier, v. 544, p. 427–445, 2021.
- ZHU, X.; WU, X.; YANG, Y. Dynamic classifier selection for effective mining from noisy data streams. In: IEEE. *Fourth IEEE International Conference on Data Mining (ICDM'04)*. [S.l.], 2004. p. 305–312.
- ZOU, H.; YANG, Y. Combining time series models for forecasting. *International journal of Forecasting*, Elsevier, v. 20, n. 1, p. 69–84, 2004.