Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Estudo Comparativo de Modelos Supervisionados com Janela Deslizante para Predição do Mercado de Ações com Base em Indicadores Técnicos

Rômulo de Vasconcelos Feijão Filho

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Orientador
Prof. Dr. Jan Mendonça Correa

Brasília
2021

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Estudo Comparativo de Modelos Supervisionados com Janela Deslizante para Predição do Mercado de Ações com Base em Indicadores Técnicos

Rômulo de Vasconcelos Feijão Filho

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Prof. Dr. Jan Mendonça Correa (Orientador)
CIC/UnB

Prof.a Dr.a Maristela Terto De Holanda
Universidade de Brasília

Prof.a Dr.a Roberta Barbosa Oliveira
Universidade de Brasília

Prof. Dr. Marcelo Grandi Mandelli
Coordenador do Bacharelado em Ciência da Computação

Brasília, 21 de maio de 2021

Dedicatória

Eu dedico esse trabalho à minha família e amigos.

Agradecimentos

Agradeço primeiramente à minha família por sempre confiarem em mim por me ensinar que nada é impossível, basta acreditar e correr atrás. Aos meus amigos, que sempre estiveram ao meu lado, desde os momentos bons aos mais turbulentos. Ao meu orientador, que mesmo em meio à pandemia e orientações por vídeo chamada, me auxiliou durante todo o processo de pesquisa e escrita da monografia. À UnB, que me deu toda a base necessária ao longo dos anos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

O mercado de ações apresenta grandes desafios ao tentar prever suas tendências, por isso, existem muitos artigos que buscam fazer essa previsão. Contudo, em muitos desses artigos, utiliza-se dados do próprio dia para fazer o cálculo do valor estimado do dia, o que pode ser considerado um cenário não realista, já que na realidade, não se sabe quais serão os valores futuros das ações. Esse trabalho tem como objetivo fazer um estudo comparativo de modelos para predição dos valores de ações de empresas de diferentes setores da bolsa de valores brasileira utilizando suas cotações e diferentes indicadores técnicos conhecidos. Para isso serão comparados quatro modelos baseados nos algoritmos SVR, ν-SVR, KRR e regressão linear. Esses modelos terão seus hiper-parâmetros ajustados e utilizarão apenas dados passados para suas predições por meio da técnica de janela deslizante, e diferentes janelas serão utilizadas, para observar como cada algoritmo se comporta em diferentes cenários. Com isso, será possível observar, para quase todos os cenários de teste e métricas adotadas, que o algoritmo KRR desempenha melhor que os outros algoritmos comparados para prever os preços e as tendências do mercado.

Palavras-chave: ciência de dados, mercado de ações, svm, krr, support vector machines, kernel ridge regression, regressão linear, predição

Abstract

The stock market present great challenges when trying to predict its tendencies, and for that, there are many articles that aim to make this prediction. Yet, in its greater part, these articles use data from the same day which the value prediction is made, which is not a realistic scenario, because in reality, this data is still in the future. This work has the objective of making a comparative study of prediction models for stock values of companies from different sectors of Brazilian's stock market, making use of its prices and known technical indicators. For that, a comparison of four different models based on the algorithm SVR, ν-SVR, KRR and linear regression will be made. These models will have its hyperparameters adjusted and will only make use of past data due to the sliding window technique, with different values for these windows, so it's possible to observe how each algorithm behaves in different scenarios. With that, it will be possible to observe that, for almost every scenario and metric adopted, the KRR model performs best in predicting these market prices and tendencies than the other algorithms.

Keywords: data science, stock market, svm, krr, support vector machines, kernel ridge regression, linear regression, prediction

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

CVM Comissão de Valores Mobiliários.

EBITDA Earning Before Interests, Taxes, Depreciation and Amortization.

EMAs Médias Móveis Exponenciais.

IFR Índice de Força Relativa.

KRR Kernel Ridge Regression.

MACD Média Móvel de Convergência/Divergência.

MAE Mean Absolute Error.

MSE Mean Squared Error.

P/L Preço/Lucro.

RBF Radial Basis Function.

RMSE Root Mean Squared Error.

RNA Redes Neurais Artificiais.

SVC Support Vector Classification.

SVM Support Vector Machine.

SVR Support Vector Regression.

Capítulo 1

Introdução

No começo de 2021, mais de 400 empresas podiam ter suas ações negociadas na bolsa. Existem diversos riscos associados ao mercado de ações, tais como: riscos de mercado, riscos de liquidez e riscos da empresa [1]. Os riscos de mercado estão relacionados a eventos externos, não necessariamente relacionados ao desempenho da empresa. Os riscos de liquidez tratam da capacidade de venda de uma ação, ou seja, em alguns casos investidores podem ter dificuldades em conseguir um comprador do papel. Por fim, temos o risco da empresa, que tem a ver com resultados negativos ou abaixo do mercado, ou até mesmo por práticas ilegais.

Existem diversos índices técnicos tais como Índice de Força Relativa [2], Bandas de Bollinger [3], Oscilador Estocástico [4], entre outros, que são utilizados no âmbito de investimentos de renda variável com o intuito de auxiliar o investidor. Esses índices têm como base valores quantitativos passados da ação, e podem indicar se, naquele momento, esta se encontra hipervalorizada, hipovalorizada, se sua volatilidade está alta ou baixa, e outras características importantes.

Devido aos desafios que apresenta, há uma oferta muito grande de artigos que tratam sobre predição no mercado de ações [5] [6] [7] [8]. Um sistema de predições eficiente pode ser interessante para o melhor entendimento do comportamento do mercado e de como ele deve variar nos dias seguintes, podendo identificar possíveis *"crashes"* ou até uma hipovalorização do mercado.

1.1 Problema

Devido à sua volatilidade e complexidade, existem muitas ferramentas de avaliação das ações, porém é complicado manter registro de múltiplas ao mesmo tempo. Portanto é interessante um modelo que automatize esse trabalho e dê estimativas de ações escolhidas. Existem muitos artigos na área apresentando bons resultados, porém muitos deles falham

em detalhar seu processo de desenvolvimento ou, usam dados do próprio dia para calcular seus valores, o que não é um cenário realista, já que no dia atual, não temos ainda conhecimento sobre qual seria o valor dos dias seguintes. [5] [7]

1.2 Objetivos

O objetivo principal deste trabalho é fazer um estudo comparativo de algoritmos de aprendizado de máquina no âmbito de predição do mercado de renda variável, utilizando apenas dados passados para prever valores futuros:

- Coleta e cálculo das *features* que serão utilizadas
- Implementação de modelos de regressão.
- Treinamento e teste de diferentes modelos de regressão.
- Experimentação dos modelos de predição utilizando empresas de múltiplos setores.
- Análise dos resultados por meio de métricas comumente utilizadas na área de aprendizagem de máquina.

1.3 Estruturação do Documento

Este trabalho foi dividido da seguinte maneira:

- Capítulo 2: Trata da fundamentação teórica do presente trabalho. Nele alguns conceitos serão introduzidos, esses sendo: o mercado de ações e como este funciona, indicadores técnicos conhecidos utilizados para o treinamento e teste dos modelos, e por fim, os algoritmos de aprendizagem de máquina utilizados.
- Capítulo 3: Uma breve revisão da literatura, aonde serão analisados artigos que inspiraram este trabalho, apresentando também possíveis pontos de aprofundamento nos estudos encontrados.
- Capítulo 4: Metodologia empregada na pesquisa, desde a coleta de dados, até a implementação e uso dos modelos e metodologia de avaliação dos resultados.
- Capítulo 5: Os resultados obtidos serão apresentados, separados por tipo de métrica avaliativa e cenário de teste.
- Capítulo 6: Será apresentada uma conclusão para a pesquisa, incluindo considerações finais e possíveis trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Para o melhor entendimento desta monografia e de seus resultados, esse capítulo visa inicialmente dar ao leitor alguns conceitos necessários do mercado de ações e dos indicadores utilizados. Em seguida, terá uma introdução ao aprendizado de máquina, uma apresentação conceitual de cada modelo e de métricas avaliativas incorporadas na pesquisa.

2.1 Mercado de Ações

Ao adquirir uma ação, o proprietário se torna um sócio minoritário daquela empresa específica, caso esse investidor adquira uma quantidade expressiva de ações da empresa, ele poderá virar um sócio majoritário e poderá desempenhar papéis de tomada de decisão de dentro da empresa.

Existem algumas vantagens que o investidor usufrui ao comprar uma ação. Se uma empresa aberta tem lucro, uma parcela desse lucro é dividido a todos os acionistas, de acordo com a proporção de ações que este tiver, esse valor distribuído é chamado de dividendo. Além disso, tem o valor da própria ação, que pode subir ou descer de acordo com perspectivas do setor em que esta atua, e do próprio desempenho da empresa.

Em alguns casos, essas ações podem ser negociadas publicamente na bolsa de valores, no caso do Brasil, a empresa deve estar registrada na Comissão de Valores Mobiliários (CVM) [9]. A CVM é uma entidade vinculada ao Ministério da Economia, com personalidade jurídica e patrimônio próprios, dotada de autoridade administrativa independente. Essa comissão cuida de toda a parte de regulamentação e desenvolvimento do mercado de valores mobiliários no Brasil.

Existem diferentes perfis de investidores, alguns buscando lucro em um curto espaço de tempo e realizando mais operações, outros já visando mais ao longo prazo. Independente do perfil do investidor, estes, quando operando de forma correta, podem realizar uma análise fundamentalista ou uma análise técnica [10]. A análise fundamentalista trata do

estado atual de uma empresa, como por exemplo, perpectivas de crescimento baseado em acontecimentos recentes. Já a análise técnica, faz o uso de métricas mensuráveis, para que seja possível observar padrões de comportamento dos valores de uma ação baseados no sentimento de seus acionistas, futuros ou atuais.

2.2 Ferramentas Utilizadas

Para a pesquisa, foi utilizada a linguagem Python, em sua versão 3.9.1. Sua escolha é justificada devido à grande quantidade de bibliotecas disponíveis voltada para a área de *Data Science*.

A captura dos dados foi feita por meio de uma biblioteca da Yahoo Finanças [1]. Com ela foi possível obter todos os dados necessários para os cálculos dos indicadores técnicos, em um período diário.

Para facilitar a análise e manipulação dos dados, foram utilizadas duas bibliotecas, Pandas [2] e NumPy [3]. A biblioteca Pandas consegue transformar os dados em *dataframes*, que são estruturas de dados que contêm linhas e colunas nomeadas. Já a NumPy, transforma os dados em NumPy *arrays*, esses, que podem ser facilmente apresentados em gráficos utilizando outras bibliotecas.

Outra biblioteca usada na pesquisa foi a scikit-learn [4]. Essa biblioteca possui uma implementação dos algoritmos de predição de dados, as métricas de avaliação utilizadas para pesquisa e também uma implementação do *grid search*.
Por fim, para criação dos gráficos, foram utilizadas as bibliotecas Seaborn e matplotlib. Essas bibliotecas permitiram a criação dos gráficos utilizando os NumPy *arrays* criados.

2.3 Indicadores Técnicos

Como *features* para treino e predição dos modelos, foram usados diferente indicadores técnicos, conhecidos no mercado de ações e em artigos relacionados. Esses indicadores fazem uso de dados quantitativos, como por exemplo, o valor de fechamento do dia da ação, o valor mais alto e o mais baixo do dia, entre outros. Esses indicadores foram o Índice de Força Relativa (IFR), Médias Móveis Exponenciais (EMAs), Média Móvel de Convergência/Divergência (MACD) com Linha de Sinal, Bandas de Bollinger e Oscilador Estocástico.

[1] `https://finance.yahoo.com/`
[2] `https://pandas.pydata.org/docs/reference/index.html`
[3] `https://numpy.org/`
[4] `https://scikit-learn.org/stable/`

2.3.1 IFR - Índice de Força Relativa

O IFR foi primeiramente apresentado por J. Welles Wilder, em 1978. É um tipo oscilador de *momentum* que mede a velocidade e as mudanças de valores das ações. O seu cálculo é feito da seguinte maneira:

$$IFR = 100 - \frac{100}{1 + FR}$$

aonde

$$FR = \frac{Média\ dos\ dias\ com\ aumento\ de\ preço\ dos\ últimos\ 14\ dias.}{Média\ dos\ dias\ com\ redução\ de\ preço\ dos\ últimos\ 14\ dias.}$$

Quanto a leitura desse índice, com a proposta dada por Wilder, valores acima de 70 indicam que a ação se encontra hipervalorizada e que existe uma tendência futura de queda em seu preço, valores abaixo de 30 indicam uma hipovalorização [2].

2.3.2 *EMAs* - Médias Móveis Exponenciais

Uma média móvel trata-se da média, com pesos iguais, de n dados anteriores. Já as EMAs, consistem de uma média móvel aonde se dá um peso maior para valores mais recentes. O artigo de Klinker [11], apresenta o cálculo desse indicador com janelas de 12 e 26 dias, e estes valores são utilizados para calcular outro indicador presente nesse estudo, o MACD, que será abordado em seguida. O cálculo das médias móveis exponenciais é feito da seguinte forma:

$$EMA_{hoje} = Close_{hoje} * (\frac{Smoothing}{1 + Dias}) + EMA_{ontem} * (1 - (\frac{Smoothing}{1 + Dias})).$$

Na fórmula, $Close_{hoje}$, representa o valor de fechamento do dia do cálculo da média móvel exponencial (EMA_{hoje}), e Dias representa a janela de dias do cálculo. Para começar o cálculo de, por exemplo, o *EMA* com janela de Dias = 12, no 13º dia deverá ser calculada a média móvel padrão e usar esse valor como a *EMA* do dia anterior, representada na fórmula por EMA_{ontem}. Quanto maior o valor de *smoothing*, maior a influência dos valores

mais recentes. O valor mais comumente utilizado para o *smoothing* na literatura é 2 [11], portanto iremos usá-lo neste estudo.

2.3.3 *MACD e Signal Line* - Média Móvel Convergente e Divergente com Linha de Sinal

Citado na subseção anterior, o MACD é um indicador de *momentum* da tendência de uma ação. Foi proposto por Gerald Appel na década de 70 e publicado no seu livro [12]. Para calculá-lo, deve-se subtrair a média móvel exponencial de janela de 26 dias pela de 12 dias.

$$MACD = EMA_{26dias} - EMA_{12dias.}$$

Em conjunto ao MACD, existe a Linha de Sinal. A Linha de Sinal nada mais é do que a média móvel exponencial do MACD em uma janela de 9 dias. Ambos indicadores, quando combinados, dão "sinais" de venda ou compra de uma ação. Um acionista recebe o sinal de compra quando a linha MACD cruza acima da Linha de Sinal, e o sinal de venda quando o MACD atravessa abaixo dessa linha.

2.3.4 Bandas de Bollinger

As bandas de Bollinger foram concebidas originalmente por John Bollinger na década de 80, e publicadas em seu livro *Bollinger on Bollinger Bands* [3]. Esse índice consiste de dois valores, um inferior e um superior, e é um indicador que trata da volatilidade e valor de mercado de uma ação. Quanto maior a diferença do parâmetro inferior para o parâmetro superior, maior a volatilidade da ação naquele momento. Além disso, existe também o conhecimento de quanto mais próximo o valor de uma ação chega ao seu parâmetro superior, ela tende a cair, e quanto mais próximo do parâmetro inferior, maior a chance do valor da ação subir. O cálculo desses parâmetros é feito da seguinte maneira:

$$Bollinger_{superior} = MédiaMóvel_{nDias}(PreçoTípico) + m * \sigma_{nDias}(PreçoTípico).$$

$$Bollinger_{inferior} = MédiaMóvel_{nDias}(PreçoTípico) - m * \sigma_{nDias}(PreçoTípico).$$

aonde

$$PreçoTípico = \frac{Close + High + Low}{3} \; e \; m = 2.$$

Na fórmula do preço típico, *Close* representa o valor de fechamento do dia, *High* o valor mais alto do dia e *Low* o valor mais baixo do dia. Para o estudo será utilizada a média móvel de 20 dias, que é mais comumente utilizada na área [3]. Além dos valores dos parâmetros inferiores e superiores, iremos utilizar também como *features* a diferença desses parâmetros com o valor da ação.

2.3.5 Oscilador Estocástico

O oscilador estocástico trata-se de um indicador de *momentum*, este mostra a posição do valor da ação com relação aos limites máximos e mínimos obtidos em um intervalo de tempo. Este indicador foi desenvolvido por George C. Lane na década de 1950 e publicado apenas na década de 1980 em seu livro [13] e de acordo com ele o oscilador estocástico não segue o preço, volume ou algo do tipo, mas sim a velocidade, ou *momentum*, do preço, já que de acordo com o autor, o *momentum* muda de direção antes do preço [14]. Seu cálculo é feito da seguinte maneira:

$$\%K = (\frac{C - Ln}{Hn - Ln}) * 100.$$

%K representa o valor do oscilador estocástico, C representa o valor de fechamento do dia mais recente, L_n é o menor valor para o fechamento em n dias, e H_n é o maior valor de fechamento em n dias. Para essa pesquisa, foi utilizado n igual a 14, pois este é mais comumente usado na área [14] [15]. Esse indicador varia de 0 a 100, e de acordo com a literatura [13], valores acima de 80 indicam uma ação hipervalorizada e valores abaixo de 20, hipovalorizada.

2.4 Aprendizagem de Máquina

O conceito inicial de aprendizagem de máquina foi apresentado por Warren S. McCulloch e Walter Pitts em 1943, os autores fazem um estudo comparativo, apresentando similaridades entre eventos neurais e aos conceitos da lógica proposicional, como por exemplo o

caráter binário das atividades do sistema neurológico [16]. A aprendizagem de máquina é o ramo da inteligência artificial que tem como premissa principal de que um sistema é capaz de aprender a identificar padrões por meio de dados. Com a aprendizagem de máquina, é possível por exemplo, identificar padrões, fazer estimações, fazer classificações, entre outras tarefas. Existem diferentes categorias de aprendizagem de máquina, como por exemplo aprendizagem supervisionada e não-supervisionada, e nessa pesquisa, serão utilizados apenas modelos supervisionados, por termos os valores esperados das ações que serão avaliadas.

2.4.1 Aprendizagem Supervisionada

Na aprendizagem supervisionada, os dados de treino que são alimentados ao algoritmo contêm os resultados desejados, estes denominados de *labels* [17]. Dessa forma, o modelo busca uma generalização sobre um conjunto de dados já com os resultados esperados, para que este possa futuramente fazer uma generalização sobre outros conjuntos de dados. Existem vários algoritmos de aprendizado supervisionado, alguns destes para classificação, outros para regressão. Neste estudo, foram utilizados apenas algoritmos de regressão, devido à natureza temporal dos dados das ações e ao uso dos algoritmos KRR e regressão linear, estes que não possuem uma versão de classificação.

2.4.2 *Support Vector Machine (SVM)*

Tendo como base na teoria de aprendizagem estatística, desenvolvida por Vapnik e Chervonenkis desde a década de 70 [18], foi desenvolvido o algoritmo SVM. As aplicações do modelo foram melhores detalhadas no livro *Statistical Learning Theory* [19]. O algoritmo SVM possui aplicação tanto para problemas de classificação, quanto para problemas de regressão, já que o problema da regressão é similar ao problema de classificação, porém o modelo de regressão é usado para dados de formato linear e o modelo de classificação para dados não contínuos [20]. Ambos os algoritmos de regressão *Support Vector Regression (SVR)* e classificação *Support Vector Classification (SVC)* do SVM buscam gerar um hiperplano ótimo para o problema, através dos vetores de suporte [21]. Existe, porém, uma grande diferença, uma margem de tolerância, representada por ϵ, na Figura 2.1. Essa margem de tolerância gera dois limitadores, um em cada lado do hiperplano penalizando previsões que estiverem a uma distância absoluta maior do que ϵ do resultado desejado, portanto, quanto menor for ϵ, menor será a tolerância a erros [20]. Enquanto na regressão buscamos inserir os valores dentro dessa margem de tolerância de forma que estes acompanhem o hiperplano, na classificação o hiperplano serve como um delimitador para classificações diferentes. Esse algoritmo faz o uso também de uma técnica mais conhecida

na literatura como *Kernel trick* [22]. Como em alguns casos pode ser complicado obter uma função que se adeque para as dimensões apresentadas do conjunto de dados, portanto nesses casos, deve-se ajustar a dimensão dos vetores de atributos. O *Kernel trick* realiza esse ajuste de forma otimizada, mapeando os pontos e suas distâncias para uma outra representação espacial.

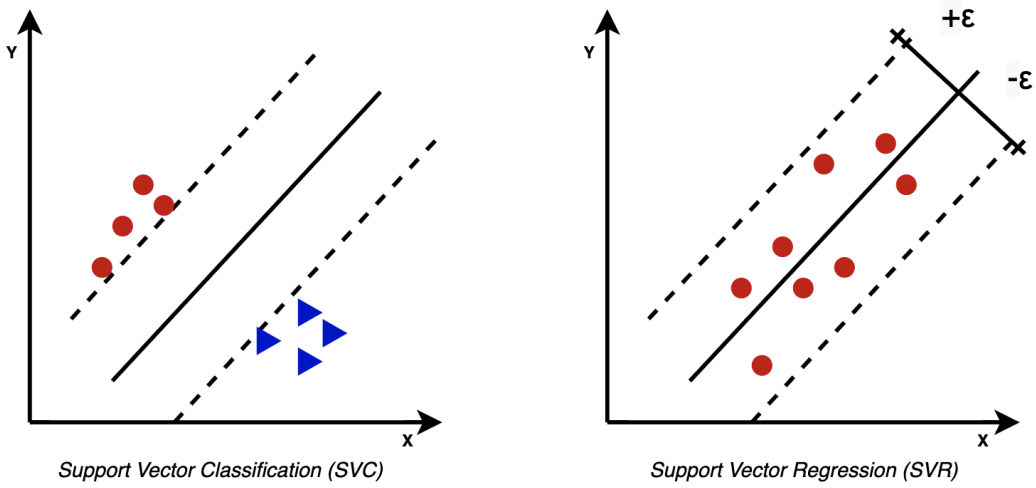

Figura 2.1: SVC x SVR

2.4.3 *Nu Support Vector Machine (ν-SVM)*

O algoritmo ν-SVM foi proposto por Schölkopf na década de 90 [23] e este é uma variação do algoritmo padrão de regressão SVR. De acordo com Scholkopf, B., Peter Bartlett e A. Smola et al. [23], este algoritmo busca minimizar de forma automática a margem de tolerância ϵ, buscando o seu valor ótimo para a regressão. Uma das principais motivações para o uso do ν-SVM, como relatado por Chang, Chih Chung e Chih Jen Lin et al. [24], é o fato da complexidade de definir o valor ideal para ϵ, portanto o algoritmo faz uso de um novo parâmetro ν, permitindo controlar o número de vetores de suporte e o número de erros de treino. Entretanto, essa busca pelo valor ótimo da margem de erro torna esse algoritmo lento para uma quantidade muito extensa de dados [24].

2.4.4 *Kernel Ridge Regression (KRR)*

SVR compartilha uma grande semelhança do algoritmo *Least Squares Support Vector Machine* (LS-SVM), usado no artigo [5] para o mesmo propósito de predição de ações.

Assim como o LS-SVM, o KRR também faz o uso dos conceitos principais de uma SVM, incluindo a técnica de *Kernel trick* apresentada na explicação do *SVM* presente Seção 2.4.2, só que empregando uma função de custo quadrático, para que seja encontrado um conjunto de funções lineares que resolva o problema de forma eficiente [25] [26], ao invés da função ϵ do SVM tradicional. Isso quer dizer que este algoritmo busca minimizar a soma das diferenças dos quadrados dos valores estimados e reais. Devido a essas função quadrática, este algoritmo está mais sujeito a ter uma sensitividade maior para *outliers*, podendo em alguns casos ajustar o modelo para o benefício destes, custando assim na predição de outras amostras.

2.4.5 *Regressão Linear*

A regressão linear busca estabelecer uma relação entre uma variável dependente, por uma ou mais variáveis independentes [27]. A fórmula de um modelo univariável de regressão linear é a seguinte:

$$Y = a + bx.$$

Aonde a representa o ponto de intercepção da reta no eixo vertical e b é o coeficiente angular, ou seja, sua inclinação, com relação à variável independente.
Essa fórmula univariável, pode ser convertida para uma multivariável, representada da seguinte maneira:

$$Y = a + b_1x_1 + b_2x_2 + ... + b_nx_n.$$

Desta forma, cada variável independente x_i, possui seu próprio coeficiente de regressão b_i.

2.4.6 Métricas de Avaliação dos Modelos

Por terem sido usados modelos de regressão, e ainda assim, ser possível avaliar ações de forma classificativa binária, aonde valores positivos representam o aumento de preço com relação ao dia anterior e valores negativos representam queda do preço com relação ao dia anterior [6], métricas para ambos os tipos de problemas puderam ser utilizadas.

Root Mean Squared Error (RMSE)

O *Mean Squared Error (MSE)* trata-se de uma métrica de regressão que representa a média do quadrado da diferença entre o valor original e o valor estimado, medindo assim a variância dos resultados. O *RMSE* é o raiz do MSE, portanto mede o desvio padrão. Ambas as métricas podem variar de 0 a ∞, quanto menor o valor, melhor a predição. Suas fórmulas podem ser representadas por:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_t - y_p)^2 \tag{2.1}$$

$$RMSE = \sqrt{MSE} \tag{2.2}$$

Na Equação 2.1, N representa o número de amostras, y_t o valor esperado e y_p o valor previsto. Para essa métrica, quanto menor o valor, melhor a predição do modelo.

Mean Absolute Error (MAE)

O *MAE* consiste da média dos erros absolutos, isso é, a média da diferença dos valores esperados e os valores obtidos na predição. Sua fórmula é a seguinte:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|v_r - v_e| \tag{2.3}$$

Na Equação 2.3, N representa o número de amostras, v_r representa algum valor esperado e v_e o seu valor estimado. Para essa métrica, maior o valor, pior a predição.

Matriz de Confusão

Usada em modelos de classificação, a matriz de confusão é representada por uma matriz de contingência de tamanho $n \times n$, aonde n é o número de classificações diferentes [28]. Por exemplo, para uma matriz com 2 classificações, a Figura 2.2 representa as seguintes relações entre valor estimado e valor real: verdadeiro positivo (v_p), verdadeiro negativo

(v_n), falso positivo (f_p) e falso negativo (f_n). A partir dessa matriz, podemos calcular outras métricas classificatórias para o modelos, tais como o índice de precisão, de acurácia e o *F1 score*. Para essa pesquisa, valores positivos representam dias em que o valor da ação subiu com relação ao dia anterior, e valores negativos, dias em que o valor da ação caiu com relação ao dia anterior.

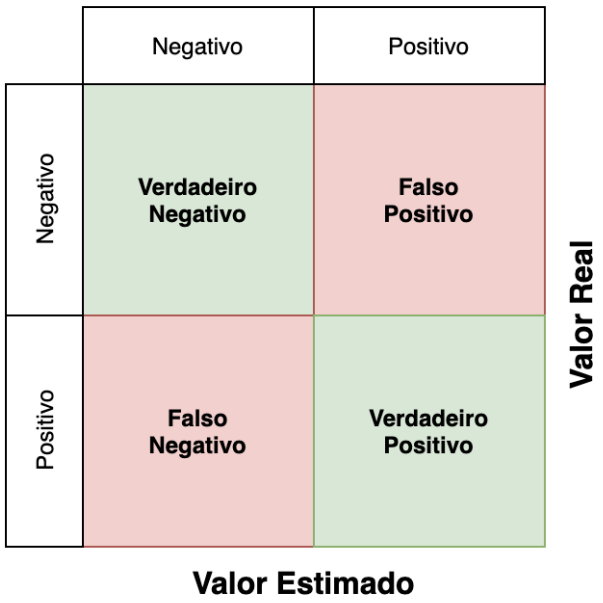

Figura 2.2: Matriz de Confusão

Precisão

A precisão indica a corretude do modelo para avaliações positivas, ou seja, a capacidade de não estimar um valor negativo para um valor positivo. Seu cálculo consiste da razão entre a quantidade de valores positivos previstos de maneira correta e a quantidade de valores positivos do espaço amostral [29]. Sua fórmula, para classificação binária, pode ser melhor descrita da seguinte forma:

$$Precisão = \frac{v_p}{v_p + f_p} \tag{2.4}$$

F1 Score

O *F1 Score* é uma métrica de avaliação de modelos de classificação. Para calculá-lo,

é necessário inicialmente calcular o valor de precisão, representado pela Equação 2.4, e o valor de revocação, representado pela Equação 2.5, pois essa métrica é uma média harmônica desses valores. O valor de revocação, permite avaliar quantas vezes que o valor esperado era positivo, e este de fato foi classificado como positivo [29]. Com isso, para modelos de classificação binários, as fórmulas para essa métrica são:

$$Revocação = \frac{v_p}{v_p + f_n} \tag{2.5}$$

$$F1\ Score = 2 * \frac{Precisão * Revocação}{Precisão + Revocação} \tag{2.6}$$

Para essa métrica, seu valor varia de 0 a 1 e quanto maior o valor, melhor o modelo.

Acurácia

A acurácia do modelo indica a capacidade do modelo de prever os valores corretamente, e é definida simplesmente como a razão da quantidade dos valores estimados corretamente e todos os valores do espaço amostral. Para classificações binárias, pode ser formulada da seguinte maneira:

$$Acurácia = \frac{v_p + v_n}{v_p + f_p + v_n + f_n} \tag{2.7}$$

Capítulo 3

Trabalhos Relacionados

Este capítulo apresenta alguns trabalhos relacionados ao tema de predição no mercado de ações que foram grandes inpiradores para a pesquisa, portanto, dividiremos em duas seções. Na primeira seção será feito um breve resumo sobre cada trabalho e sobre quais técnicas foram empregadas nesta pesquisa, na segunda seção serão apresentadas algumas considerações finais, indicando possíveis pontos de melhoria em relação aos estudos observados.

3.1 Revisão da Literatura

Existem muitos artigos na área de predição no mercado de ações. Apenas nos últimos 5 anos, usando a plataforma IEEExplorer e a *string* de busca *"stock market AND (prediction OR forecasting)"*, no momento da escrita desta monografia, foram encontrados 2449 trabalhos relacionados. Devido a esse número, os artigos principais utilizados para a pesquisa foram obtidos filtrando os resultados. Foi observada uma grande quantidade de artigos obtendo bons resultados utilizando o algoritmo SVM e suas variações, portanto esse algoritmo será o mais utilizado aqui. Foi utilizado também o algoritmo de regressão linear para um efeito comparativo, devido a sua simplicidade. Assim, poderemos observar a diferença deste com um algoritmo um pouco mais robusto, no caso, o SVM.

O artigo [5] foi a base inicial para a pesquisa, devido aos resultados obtidos, as técnicas nele empregadas. Neste artigo o autor fez um estudo comparativo de variações do LS-SVM com o algoritmo de Redes Neurais Artificiais (RNA). Além de ter utilizado o algoritmo de LS-SVM, este artigo fez o uso de indicadores técnicos para prever os valores das ações. Uma das motivações do artigo foi o fato da RNA apresentar, em muitos casos, uma incapacidade de generalizar em cima dos dados, um problema de *over-fitting* [5] [30].

Para os testes, os autores fizeram o uso de dados dos 3 anos antecedentes à data de publicação do artigo, cobrindo ações de todos os setores do SP 500, com a proporção dos

dados de treino e de teste de 70% e 30%, respectivamente. Como análise dos resultados, utilizando a métrica MSE para cada ação e algoritmo diferente empregado na pesquisa, os resultados para o algoritmo RNA variaram de 0,4055 até 3,5049 com um desvio padrão de 0,8723, com o LS-SVM puro variaram de 0,4673 até 2,1034 com um desvio padrão de 0,4411, e por fim, o PSO-LS-SVM, teve uma variação de 0,1735 até 1,5881, com desvio padrão de 0,3469.

Outro artigo utilizado para esse estudo foi o de Gururaj, Vaishnavi, V. R. Shriya e Dr. Ashwini K et al. [7]. Nele os autores fizeram um estudo comparativo dos algoritmos de regressão linear e o algoritmo padrão de SVM. Nesse artigo foi utilizada a técnica de predição por janela deslizante. Essa técnica consiste de usar dados passados como *input* para prever dados futuros [7], onde no caso do artigo avaliado, foi de um dia anterior. Os dados utilizados vieram de uma biblioteca externa chamada Quandl, aonde os dados das ações da Coca-Cola de 01 de Janeiro de 2017 até 01 de Janeiro de 2018 foram capturados e transformados em um arquivo .csv para serem utilizados. Os resultados foram analisados a partir de quatro métricas diferentes, estas sendo o RMSE, MAE, MSE e coeficiente de determinação (R^2). Para o modelo de regressão linear, os resultados das métricas avaliativas foram de 3,22 para o RMSE, 2,53 para o MAE, 10,37 para o MSE e 0,73 para o R^2. Já para o modelo de SVM, os resultados foram 1,58 para o RMSE, 1,33 para o MAE, 2,51 para o MSE, e por fim, 0,93 para o R^2. Sendo assim, os resultados obtidos pelo SVM foram melhores.

O estudo de Stuke, Annika, Patrick Rinke e Milica Todorović et al. [31] apresenta algumas técnicas de otimização de hiperparâmetros, tais como *random search*, *grid search*, otimização bayesiana e suas aplicações para o algoritmo KRR. Neste trabalho, o KRR é empregado para prever a energia molecular orbital do dataset QM9 contendo 134 mil moléculas orgânicas pequenas. Como citado no artigo, o KRR possui três hiperparâmetros que podem ser otimizados, estes sendo γ, α e o *kernel* [31]. Os autores afirmam que, apesar de a otimização bayesiana apresentar os hiperparâmetros ótimos de forma eficiente, a técnica de *grid search* também consegue descobrir os melhores hiperparâmetros, dado um *grid* suficientemente bom.

Por fim, outro artigo da literatura também usado para esse estudo foi o de Ravikumar, S. e P. Saraf et al. [8]. Este detalha muito bem todo o processo de estudo, desde a coleta de dados, até as avaliações dos modelos. Para a predição foram usados alguns indicadores de *momentum* e volatilidade, apresentados pelo autor. Foram usados cinco algoritmos diferentes de regressão por essa pesquisa, esses sendo regressão linear, regressão polinomial, SVR, regressão de árvores de decisão e regressão *random forest*, e seis algoritmos de classificação, SVC, *K-nearest neighbors (KNN)*, regressão logística, *naive bayes*, classificador de árvores de decisão e classificador *random forest*. Além disso, múltiplos *kernels* foram

utilizados para os modelos de SVM, esses sendo *kernel* linear, polinomial, Radial Basis Function (RBF) e sigmoidal.

Para os resultados de regressão, usando a métrica de acurácia, o que obteve o melhor resultado foi o regressor *random forest* com 99.57% e o pior o de regressão linear, com 81.52%. O modelo de regressão SVR obteve um resultado de 87.41% na sua métrica de acurácia. Já para os modelos de classificação, o que obteve maior acurácia foi a regressão logística com 68.27%, e o pior foi o classificador de árvore de decisão, com 57.99% de acurácia. O classificador SVC, obteve acurácia de 68.41% para o kernel linear e 67.86% para o kernel RBF.

3.2 Considerações Finais

Com a análise dos estudos, vimos como as técnicas foram aplicadas para o problema de predição do valor das ações e os resultados obtidos por cada estudo. Apesar de apresentarem bons resultados, alguns deles apresentam possíveis pontos que podem ser modificados ou acrescentados, como por exemplo Hegazy, Osman, Omar S. Soliman e Mustafa Abdul Salam et al. [5], apesar de haver uma descrição detalhada das *features* utilizadas no treinamento e teste dos modelos, para a avaliação destes foi usada apenas uma métrica, já em Gururaj, Vaishnavi, V. R. Shriya e Dr. Ashwini K et al. [7], não ficou claro quais *features* foram utilizadas para o treinamento e teste dos modelos, além disso, foi usada apenas uma única ação para testes. Além disso, apenas um artigo foi encontrado explicitando que os dados utilizados para treinamento e teste dos modelos foram anteriores ao dia que seria estimado. A ideia deste trabalho é utilizar as melhores ideias dos estudos analisados durante a pesquisa, como por exemplo o uso de janela deslizante para o uso de dados passados na previsão de valores futuros [7], o uso de indicador de volatilidade (bandas de Bollinger) nas *features* de treinamento e teste [32], realizar avaliações e ajustes de hiperparâmetros [31], entre outros.

Capítulo 4

Metodologia

Este capítulo tem como objetivo apresentar todas as etapas da metodologia empregada. Na Seção 4.1, será apresentada a forma que foi realizada a coleta dos dados, quais dados de quais empresas foram utilizados, qual período escolhido para avaliação e por que, qual a quantidade de dados obtida e algumas informações sobre estes. Na Seção 4.2, será explicado como foi a etapa de pré-processamento dos dados, de forma a obter cada índice técnico utilizado para treinamento e teste dos modelos. Por fim, na Seção 4.3, terá uma breve explicação dos modelos utilizados e dos parâmetros avaliados para estes, será melhor explicada a técnica de janela deslizante que foi utilizada nos dados de treino e teste, e alguns cenários de teste abordados durante a pesquisa.

4.1 Coleta dos dados

Para a parte de coleta dos dados, foi utilizada uma biblioteca da Yahoo Finanças [1], com ela foram coletados dados das ações de empresas de 9 setores diferentes na bolsa de valores, essas sendo Embraer (bens industriais), Oi (comunicações), Magazine Luiza (consumo cíclico), Ambev (consumo não cíclico), Banco do Brasil (financeiro), Vale (materiais básicos), Petrobras (petróleo, gás e biocombustíveis), Grupo Fleury (saúde), e por fim, Comgás (utilidade pública). Todos os dados coletados vieram no espaço de um ano, iniciando em 1 de Janeiro de 2018 até 31 de Dezembro de 2018. Foi preferido utilizar os dados de 2018 devido a 2019 ter sido um ano de mudança na presidência, e 2020, por conta da pandemia do coronavírus, fatores que poderiam impactar negativamente o treinamento ou teste dos modelos propostos. Os dados coletados de cada empresa continham 245 linhas [2] e 6 colunas e estes foram importados diretamente para um *dataframe*, utilizando a

[1] `https://finance.yahoo.com`

[2] Esses números são pelo fato de não ocorrer abertura na bolsa nos finais de semana e feriados

biblioteca pandas [3]. A data era representada pelo índice (*Date*), e as colunas continham dados sobre o valor mais alto atingido pela ação no dia (*High*), o valor mais baixo do dia (*Low*), o valor no início do dia (*Open*), o valor ao fim do dia (*Close*), a quantidade de cotas negociadas no dia (*Volume*) e o valor ao fim do dia com os ajustes de distribuição de divisões e dividendos (*Adj. Close*). A Figura 4.1 representa as primeiras cinco linhas do *dataframe* da Embraer.

Date	High	Low	Open	Close	Volume	Adj Close
2018-01-02	20.990000	20.230000	20.629999	20.520000	3813300.0	20.389370
2018-01-03	21.600000	20.860001	21.049999	21.299999	9403700.0	21.164400
2018-01-04	22.200001	21.500000	21.500000	21.799999	6597700.0	21.661217
2018-01-05	22.129999	20.650000	21.900000	20.650000	8980400.0	20.518539
2018-01-08	21.170000	20.500000	21.059999	20.700001	4798500.0	20.568220

Figura 4.1: Cinco primeiras linhas do *dataframe* da Embraer

4.2 Pré-processamento

Após a importação dos dados, foi feita uma verificação se não haviam valores faltando, o que poderia causar inconsistências no cálculo dos indicadores e, consequentemente, no desempenho dos modelos. Devido ao fato dos dados terem sido importados de forma estruturada e com os dados básicos totalmente preenchidos, não houve a necessidade do uso de técnicas de preenchimento ou remoção de inconsistências. Após isso, foram calculados todos os indicadores técnicos de cada ação. Os indicadores técnicos para treinamento e teste dos modelos foram escolhidos após uma análise preliminar, aonde foram escolhidos os indicadores que apresentaram uma maior correlação linha com o preço de fechamento. Inicialmente, havia-se pensado de fazer uso de outro indicador técnico, *Money Flow Index* [33], porém após analisar uma matriz de correlação gerada para análise das *features*, foi escolhido não usá-lo já que, em muitos casos, como podemos ver na Figura 4.2, este não possuia uma correlação boa o suficiente para o treinamento e teste dos modelos.

[3] `https://pandas.pydata.org`

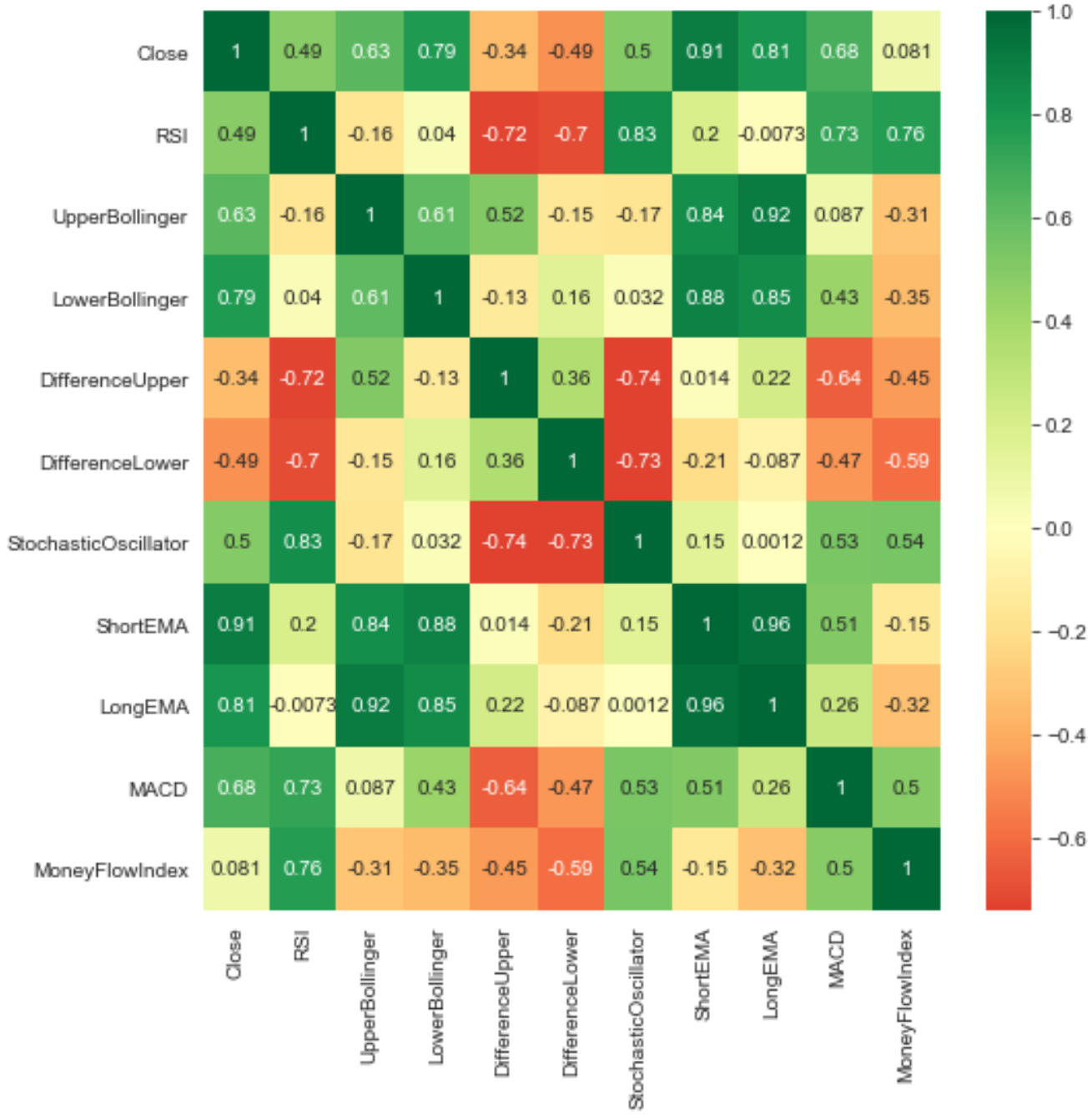

Figura 4.2: Matriz de correlação da Embraer

4.3 Implementação dos Modelos

Como foi mencionado anteriormente, os algoritmos utilizados para a implementação dos modelos de predição dos valores das ações foram *SVR* [4], *ν-SVR*[5], *KRR*[6] e regressão linear[7], explicados na Subseção 2.3.2. Além desses algoritmos, foi empregada também a técnica de GridSearch [8], para buscar os melhores hiperparâmetros dos modelos *SVR*, *ν-SVR* e *KRR*, com exceção do *kernel*, que foi definido anteriormente como RBF, por este ser o mais comum na literatura encontrada [8] [5] para aplicações de modelos baseados em *SVM*. Para o modelos *SVR* e *ν-SVR*, foram escolhidos para avaliação alguns valores definidos por Murphy [34], esse sendo:

- C: 1, 10 e 100;
- γ: 1, 0.1, 0.01, 0.001 e 0.0001.

C é um parâmetro utilizado apenas no SVR e serve para indicar a margem de erro aceitável para cada ponto, enquanto γ indica a curvatura para os limites da margem de decisão e foi modificado tanto para *SVR*, quanto *ν-SVR*. Os valores que apresentaram os melhores resultados para C e γ foram 100 e 0.0001, respectivamente, portanto esses foram utilizados para os modelos finais. Para o modelo *ν-SVR* ainda foi utilizado o parâmetro ν com valor de 0.5, pois este foi definido como padrão da biblioteca utilizada.

Por fim, para o modelo KRR, não foi encontrado nenhum artigo que definisse os melhores parâmetros possíveis para serem avaliados, então foram avaliados valores para os hiperparâmetros:

- α: 1, 10, 100;
- γ: 1, 0.1, 0.01, 0.001 e 0.0001.

Pode-se perceber que os valores avaliados para o parâmetro α são os mesmos do parâmetro C, avaliados na *SVM*, já que este parâmetro representa a função de penalização para esse algoritmo. Para os parâmetros avaliados, o que apresentou o melhor resultado foram 1 e 1, portanto foram usados para o modelo final.

[4] `https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html`

[5] `https://scikit-learn.org/stable/modules/generated/sklearn.svm.NuSVR.html#sklearn.svm.NuSVR`

[6] `https://scikit-learn.org/stable/modules/generated/sklearn.kernel_ridge.KernelRidge.html`

[7] `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html`

[8] `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html`

4.3.1 Janela Deslizante

Além de ajustar os hiperparâmetros, foi aplicada aos dados de entrada, representados por X_i, e saída do modelo, Y_i, a técnica de janela deslizante, também empregada no trabalho de Gururaj, Vaishnavi, V. R. Shriya e Dr. Ashwini K et al. [7], de forma que pudesse usar dados passados para prever dados futuros. Como mostra a Figura 4.3, para uma lista com um conjunto de *features* e uma lista com os resultados do valor de fechamento das ações, foram retirados os últimos 5 registros da lista com o conjunto das *features* e os 5 primeiros da lista com os valores de fechamento, dessa forma, ao treinar o modelo, o treinamento de algum dia será feito com os dados de 5 dias anteriores a esse dia.

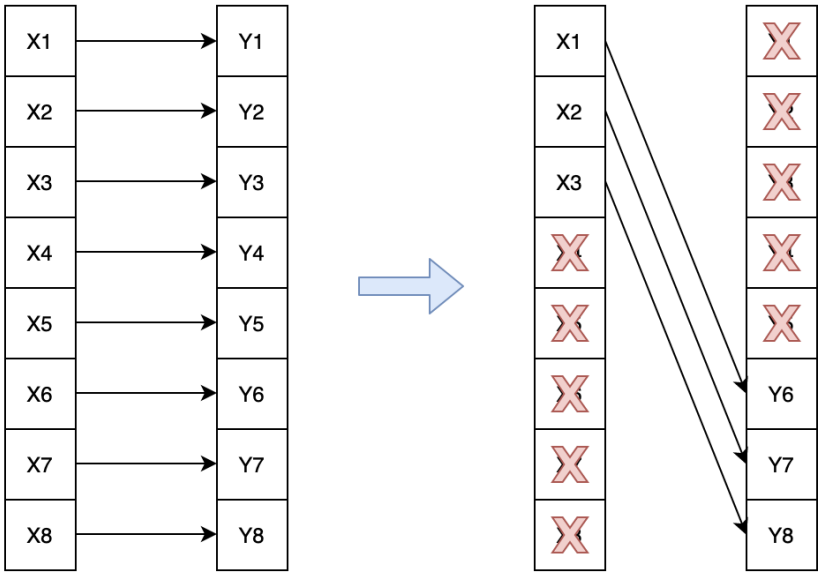

Figura 4.3: Janela deslizante de 5 dias

4.3.2 Cenários de Teste

Foram elaborados diferentes cenários de teste para cada modelo avaliado. Para todos os cenários de teste, os dados de treinamento e de teste foram divididos com uma proporção de 75%/25%, respectivamente. Para que pudesse ser avaliado de forma mais detalhada, foram realizados testes com diferentes janelas deslizantes, essas podendo ser de 1, 3 ou 5 dias. Todos os testes foram avaliados utilizando as mesmas métricas de avaliação anteriormente mencionadas na Subseção 2.4.3.

Capítulo 5

Resultados

Para validar os modelos e a metodologia proposta ao longo do trabalho, esse capítulo apresentará o ambiente ao qual os experimentos foram realizados e os resultados experimentais obtidos para cada caso de teste. No Anexo I, são apresentados gráficos contendo os resultados de cada métrica avaliativa para cada cenário de teste, e no Anexo II, são apresentados gráficos das predições das ações da Ambev de cada modelo, e para cada cenário de teste.

5.1 Ambiente de Experimento

Para todos os cenários de teste, o experimento foi realizado em sistema MacOS Mojave, versão 10.14.6. A máquina possui 8GB de memória primária e uma *CPU Intel Core i5* de 8^a geração com quatro núcleos e *clock* de 1.4 GHz.

5.2 Resultados das Métricas de Classificação

Nas Subseções 5.2.1, 5.2.2 e 5.2.3, serão apresentados os resultados das métricas avaliativas de classificação para cada modelo e empresa testados durante a pesquisa, divididos por janela de teste.

5.2.1 Janela Deslizante de 5 Dias

Avaliando as métricas classificatórias com janela de 5 dias, apresentadas nas tabelas de 5.1 a 5.3, o modelo KRR obteve os melhores resultados em todas as métricas avaliativas, com uma média de 0.8291 no seu *F1-score*, 0.8350 na precisão e 0.8440 na acurácia. Enquanto isso, o modelo de regressão linear obteve os piores resultados em todos os quesitos, com

o *F1-score* de 0.7228, precisão de 0.7435 e acurácia de 0.7423. Entre os modelos SVR e ν-SVR, o modelo SVR apresentou um resultado melhor.

Tabela 5.1: F1-score - Janela de 5 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.7368	0.6842	0.9231	0.8108
Banco do Brasil	0.8421	0.8000	0.8276	0.6538
Comgás	0.7727	0.7826	0.8750	0.7826
Embraer	0.7407	0.7308	0.9020	0.7170
Grupo Fleury	0.6875	0.6667	0.6875	0.5556
Magazine Luiza	0.8235	0.7778	0.8214	0.6667
Oi	0.7879	0.7879	0.7500	0.6250
Petrobras	0.8571	0.8372	0.8182	0.8372
Vale	0.8182	0.9167	0.8571	0.8571
Média	**0.7852**	**0.7760**	**0.8291**	**0.7228**

Tabela 5.2: Precisão - Janela de 5 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.7778	0.7222	0.9474	0.8824
Banco do Brasil	0.8000	0.7857	0.7742	0.6800
Comgás	0.8947	0.8571	0.9130	0.8571
Embraer	0.6897	0.7037	0.8846	0.6786
Grupo Fleury	0.6471	0.6111	0.6471	0.4762
Magazine Luiza	0.8077	0.7241	0.7419	0.6207
Oi	0.8667	0.8667	0.8571	0.7143
Petrobras	0.8182	0.7826	0.7500	0.7826
Vale	0.9000	0.9167	1.0000	1.0000
Média	**0.8002**	**0.7744**	**0.8350**	**0.7435**

Tabela 5.3: Acurácia - Janela de 5 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.7872	0.7447	0.9362	0.8511
Banco do Brasil	0.8085	0.7660	0.7872	0.6170
Comgás	0.7872	0.7872	0.8723	0.7872
Embraer	0.7021	0.7021	0.8936	0.6809
Grupo Fleury	0.7872	0.7660	0.7872	0.6596
Magazine Luiza	0.8085	0.7447	0.7872	0.6170
Oi	0.8511	0.8511	0.8298	0.7447
Petrobras	0.8723	0.8511	0.8298	0.8511
Vale	0.8298	0.9149	0.8723	0.8723
Média	**0.8038**	**0.7920**	**0.8440**	**0.7423**

5.2.2 Janela Deslizante de 3 Dias

Para as métricas classificatórias com janela deslizante de 3 dias, o modelo KRR ainda obteve o melhor resultado em todas as métricas, porém foi possível perceber algumas diferenças no desempenho dos modelos com relação à janela de 5 dias.

Primeiramente, ao observar as Tabelas 5.4, 5.5 e 5.6, todos os modelos obtiveram resultados melhores com relação aos resultados obtidos para a janela de 5 dias. Outra diferença observada foi que o modelo de regressão linear obteve uma precisão de 0.8460, melhor que o resultado obtido pelo modelo ν-SVR, com 0.8184. O modelo de regressão linear também superou o modelo de ν-SVR na métrica de acurácia, o primeiro obtendo um resultado de 0.8299 e o segundo, 0.8209. Apesar disso, quanto ao *F1-score*, o modelo de regressão linear desempenhou pior quando comparado ao ν-SVR, indicando que o modelo de regressão linear é tendencioso quanto a previsão de resultados positivos, ou seja, para dias aonde o valor da ação subiu com relação ao dia anterior e tem uma dificuldade maior para prever valores negativos.

Tabela 5.4: F1-score - Janela de 3 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.8333	0.8500	0.9231	0.8889
Banco do Brasil	0.9153	0.8387	0.9492	0.7843
Comgás	0.8261	0.8000	0.9020	0.8980
Embraer	0.9434	0.7273	1.0000	0.9804
Grupo Fleury	0.6471	0.7568	0.6471	0.5000
Magazine Luiza	0.8364	0.7857	0.8525	0.7458
Oi	0.7879	0.8235	0.7879	0.6875
Petrobras	0.8372	0.8095	0.8636	0.8182
Vale	0.9200	0.9231	0.9796	0.9388
Média	**0.8385**	**0.8127**	**0.8783**	**0.8047**

Tabela 5.5: Precisão - Janela de 3 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.9375	0.8500	0.9474	1.0000
Banco do Brasil	0.9000	0.7879	0.9333	0.9091
Comgás	0.9500	0.9474	0.9200	0.9565
Embraer	0.9259	0.6897	1.0000	1.0000
Grupo Fleury	0.6471	0.7000	0.6471	0.4737
Magazine Luiza	0.8214	0.7586	0.7647	0.6875
Oi	0.9286	0.9333	0.9286	0.8462
Petrobras	0.8182	0.8095	0.8261	0.7826
Vale	0.9200	0.8889	1.0000	0.9583
Média	**0.8721**	**0.8184**	**0.8852**	**0.8460**

Tabela 5.6: Acurácia - Janela de 3 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.8776	0.8776	0.9388	0.9184
Banco do Brasil	0.8980	0.7959	0.9388	0.7755
Comgás	0.8367	0.8163	0.8980	0.8980
Embraer	0.9388	0.6939	1.0000	0.9796
Grupo Fleury	0.7551	0.8163	0.7551	0.6327
Magazine Luiza	0.8163	0.7551	0.8163	0.6939
Oi	0.8571	0.8776	0.8571	0.7959
Petrobras	0.8571	0.8367	0.8776	0.8367
Vale	0.9184	0.9184	0.9796	0.9388
Média	**0.8617**	**0.8209**	**0.8957**	**0.8299**

5.2.3 Janela Deslizante de 1 Dia

Para os testes com janela deslizante de 1 dia foram observadas algumas diferenças consideráveis quando comparando os resultados obtidos, apresentados nas Tabelas 5.7, 5.8 e 5.9, com os resultados para as janelas de 3 e 5 dias, Tabelas 5.1 a 5.6.
Primeiro, o modelo KRR foi superado pelo modelo de regressão linear na métrica de precisão, com o primeiro obtendo 0.9519, e o segundo 0.9533.
Outra diferença observada, foi que o modelo ν-SVR obteve os piores resultados para todas as métricas de classificação avaliadas, quando comparados com os outros modelos.

Tabela 5.7: F1-score - Janela de 1 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.8889	0.8947	0.9474	0.9474
Banco do Brasil	0.9180	0.9333	0.9841	0.9153
Comgás	0.8980	0.8000	0.9434	0.9231
Embraer	0.9259	0.7931	0.8846	0.8462
Grupo Fleury	0.8947	0.8947	0.8649	0.8649
Magazine Luiza	0.8525	0.7667	0.9333	0.9655
Oi	0.7895	0.5455	0.8333	0.8649
Petrobras	0.8444	0.8696	0.9778	0.9778
Vale	0.9412	0.9412	0.9600	0.9600
Média	**0.8837**	**0.8265**	**0.9254**	**0.9183**

Tabela 5.8: Precisão - Janela de 1 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	1.0000	0.9444	1.0000	1.0000
Banco do Brasil	0.9333	0.9655	0.9688	0.9643
Comgás	1.0000	0.8696	0.9615	0.9600
Embraer	0.9259	0.7419	0.9200	0.8800
Grupo Fleury	0.8500	0.8500	0.8421	0.8421
Magazine Luiza	0.7879	0.7188	0.8750	0.9333
Oi	0.8824	0.7500	1.0000	1.0000
Petrobras	0.8636	0.8696	1.0000	1.0000
Vale	0.9600	0.9600	1.0000	1.0000
Média	**0.9115**	**0.8522**	**0.9519**	**0.9533**

Tabela 5.9: Acurácia - Janela de 1 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.9231	0.9231	0.9615	0.9615
Banco do Brasil	0.9038	0.9231	0.9808	0.9038
Comgás	0.9038	0.8077	0.9423	0.9231
Embraer	0.9231	0.7692	0.8846	0.8462
Grupo Fleury	0.9231	0.9231	0.9038	0.9038
Magazine Luiza	0.8269	0.7308	0.9231	0.9615
Oi	0.8462	0.7115	0.8846	0.9038
Petrobras	0.8654	0.8846	0.9808	0.9808
Vale	0.9423	0.9423	0.9615	0.9615
Média	**0.8953**	**0.8462**	**0.9359**	**0.9273**

5.3 Resultados das Métricas de Regressão

Nas Subseções 5.3.1, 5.3.2 e 5.3.3, serão apresentados os resultados das métricas avaliativas para cada modelo e empresa testados durante a pesquisa, divididos por janela de teste.

5.3.1 Janela Deslizante de 5 Dias

Para os resultados obtidos nos testes com janela de 5 dias, apresentados pelas Tabelas 5.10 e 5.11, o modelo KRR apresentou os melhores resultados em ambas as métricas, com um MAE de 0.8345 e um RMSE de 1.0672. Enquanto isso, o modelo de regressão linear apresentou os piores resultados em ambas as métricas, obtendo um resultado médio de 1.2379 para a métrica MAE e 1.4662 para a métrica RMSE. O modelo SVR superou o modelo ν-SVR, comportamento também observado nas as métricas classificatórias com janela de 5 dias.

Tabela 5.10: MAE - Janela de 5 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.2839	0.4522	0.2065	0.2782
Banco do Brasil	0.5478	0.5939	0.9192	1.3035
Comgás	1.3346	1.4789	2.0576	3.7720
Embraer	0.3708	0.3156	0.4817	0.4584
Grupo Fleury	1.2301	1.1530	0.2980	1.6334
Magazine Luiza	0.2873	0.2442	0.1991	0.3446
Oi	0.2741	0.2222	0.0960	0.0845
Petrobras	3.2607	3.7347	2.2669	2.3500
Vale	1.6033	1.4913	0.9854	0.9163
Média	**1.0214**	**1.0762**	**0.8345**	**1.2379**

Tabela 5.11: RMSE - Janela de 5 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.4090	0.5388	0.2437	0.3125
Banco do Brasil	0.6845	0.7232	1.1266	1.4749
Comgás	1.7116	1.8272	2.4266	4.1195
Embraer	0.4187	0.4147	0.6475	0.6022
Grupo Fleury	1.4357	1.4777	0.3615	1.6949
Magazine Luiza	0.3128	0.3031	0.2694	0.4106
Oi	0.2966	0.3320	0.1106	0.1189
Petrobras	3.3638	3.8544	3.2260	3.3361
Vale	1.7764	1.5993	1.1927	1.1266
Média	**1.1566**	**1.2300**	**1.0672**	**1.4662**

5.3.2 Janela Deslizante de 3 Dias

O modelo KRR continuou desempenhando melhor que os outros modelos para ambas as métricas de regressão, porém algumas diferenças foram observadas quando comparados os resultados obtidos com janela de 5 dias.

A primeira diferença foi o modelo ν-SVR ter apresentado um resultado melhor que o modelo SVR em ambas as métricas avaliadas, além disso, o modelo SVR obteve o pior resultado para a métrica MAE, indicando que este teve o maior erro absoluto nos testes realizados.

Outra diferença foi a redução dos valores dos erros obtidos, o que pode ser justificado pela maior proximidade dos dados utilizados para predição com relação aos resultados

esperados.

Tabela 5.12: MAE - Janela de 3 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.2351	0.3491	0.1252	0.1714
Banco do Brasil	0.4194	0.4783	0.6519	1.2643
Comgás	1.1848	1.2055	1.6697	1.6999
Embraer	0.2949	0.3231	0.3475	0.3949
Grupo Fleury	0.7289	0.4805	0.3417	1.2062
Magazine Luiza	0.2404	0.2026	0.1391	0.2455
Oi	0.2267	0.1524	0.0666	0.0680
Petrobras	2.1018	2.3055	1.2514	1.2878
Vale	1.4392	1.0630	0.6099	0.5096
Média	**0.7635**	**0.7289**	**0.5781**	**0.7608**

Tabela 5.13: RMSE - Janela de 3 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.3253	0.4271	0.1482	0.1974
Banco do Brasil	0.5029	0.6375	0.8024	1.6285
Comgás	1.5027	1.5933	2.0554	2.0091
Embraer	0.3322	0.4644	0.4312	0.4987
Grupo Fleury	0.9477	0.7040	0.4047	1.2827
Magazine Luiza	0.2636	0.2484	0.1956	0.2939
Oi	0.2442	0.2152	0.0791	0.0932
Petrobras	2.1895	2.3893	1.7337	1.8431
Vale	1.5999	1.1734	0.6843	0.5746
Média	**0.8787**	**0.8725**	**0.7261**	**0.9357**

5.3.3 Janela Deslizante de 1 Dias

Para as métricas de regressão com janela de 1 dia, o KRR continuou obtendo os melhores resultados e o valor dos erros continuou diminuindo com relação aos outros cenários de teste avaliados. Dessa vez, o modelo de regressão linear superou os modelos SVR e ν-SVR em ambas as métricas, resultando em um MAE médio de 0.2699 e um RMSE de 0.3251. O modelo de ν-SVR apresentou uma melhora menos considerável quando comparado ao modelo SVR e os resultados obtidos com janela de 3 dias, voltando a ser superado pelo SVR quando usada uma janela de 1 dia.

Tabela 5.14: MAE - Janela de 1 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.2439	0.2664	0.0724	0.0671
Banco do Brasil	0.3017	0.3850	0.2733	0.4425
Comgás	0.7526	1.0054	0.7401	0.6354
Embraer	0.1054	0.1684	0.1486	0.1399
Grupo Fleury	0.2739	0.5334	0.1489	0.3808
Magazine Luiza	0.1832	0.1596	0.0890	0.0918
Oi	0.2002	0.1125	0.0326	0.0306
Petrobras	1.1279	1.1276	0.4198	0.4527
Vale	0.2983	0.5043	0.1981	0.1881
Média	**0.3875**	**0.4736**	**0.2359**	**0.2699**

Tabela 5.15: RMSE - Janela de 1 dias (Empresa/Modelo)

	SVR	$\nu - SVR$	KRR	Regressão Linear
Ambev	0.3202	0.3502	0.0967	0.0838
Banco do Brasil	0.3797	0.3850	0.3462	0.5670
Comgás	1.0645	1.4926	0.9684	0.7461
Embraer	0.1468	0.2518	0.2110	0.2000
Grupo Fleury	0.4410	0.5883	0.1764	0.4153
Magazine Luiza	0.2033	0.1958	0.1154	0.1134
Oi	0.2064	0.1330	0.0416	0.0391
Petrobras	1.2518	1.3461	0.5394	0.5551
Vale	0.3858	0.6145	0.2236	0.2065
Média	**0.4888**	**0.5953**	**0.3021**	**0.3251**

Capítulo 6

Conclusão

Este capítulo trata-se da conclusão da monografia, apontando na Seção 6.1 algumas considerações finais sobre a pesquisa e na Seção 6.2, alguns possíveis trabalhos futuros.

6.1 Considerações Finais

Este trabalho apresentou um estudo comparativo de modelos de aprendizagem supervisionados aplicados no contexto de predições de valores e direção no mercado de ações, utilizando a técnica de janela deslizante. Com isso foram utilizadas empresas presentes na bolsa de valores de diferentes setores.

Primeiramente, foram apresentados os desafios presentes na área de predição do mercado de ações, como seus diferentes riscos e volatilidade. Após isso, foram apresentados alguns conceitos sobre o mercado de ações, indicadores técnicos conhecidos da área que seriam utilizados nessa pesquisa, os quatro modelos diferentes que seriam comparados durante a pesquisa e as cinco métricas avaliativas, sendo dessas três métricas de classificação e duas métricas de regressão.

Em seguida, foram apresentados resultados de consulta em bases de pesquisa sobre artigos relacionados, e alguns artigos que serviram como inspiração inicial e consulta para essa monografia foram melhor detalhados. Além das metodologias aplicadas e dos resultados obtidos, foram apresentados também possíveis pontos que poderiam ser explorados ou melhor detalhados nessas pesquisas, de forma que esses pontos pudessem ser incoporados nessa monografia.

Consequentemente, foi apresentada a metodologia seguida para a pesquisa, desde as etapas de coleta de dados, pré-processamento, implementação dos modelos e os cenários de teste criados. Durante essa discussão, foram apresentadas as empresas que foram avaliadas na pesquisa, o formato dos dados obtidos e uma análise correlacional das *features* e os valores das ações. Além disso também foi melhor detalhado o processo de seleção dos

hiperparâmetros dos modelos e quais foram utilizados para avaliação final em busca dos melhores resultados, e também, como funciona a técnica de janela deslizante, empregada na pesquisa para fazer uso de dados passados para prever dados futuros, algo que foi pouco utilizado ou apresentado em artigos relacionados observados.

Por fim, foram apresentados os resultados obtidos separados por tipo de métrica avaliativa e por cenário de teste. Pelos resultados observados, o modelo KRR se apresentou superior em quase todas as métricas e cenários avaliados, com exceção apenas da métrica de precisão para a janela de 1 dia, o qual foi superado pelo modelo de regressão linear. Uma observação a ser feita, foi que todos os modelos apresentaram uma tendência maior de acertar um valor como positivo do que negativo, evidenciado pelos resultados médios do *F1-score* terem sido inferiores aos resultados obtidos para as métricas de precisão e acurácia, em todos os modelos e em todos os cenários de teste avaliados. O modelo de regressão linear se mostrou mais efetivo que os modelos SVR e ν-SVR para a janela mais curta, de 1 dia, enquanto para janelas de 3 a 5 dias, o modelo de regressão linear foi o modelo que obteve os piores resultados. Surpreendentemente, o modelo SVR se provou mais eficaz que o modelo ν-SVR, isso pode ter sido causado pelo uso do valor padrão da biblioteca para o parâmetro ν, presente apenas no modelo ν-SVR, já que os outros parâmetros utilizados foram os mesmos do modelo SVR.

Ainda sobre os resultados, foi possível perceber um melhor desempenho dos modelos à medida que a janela deslizante foi diminuida, isso faz sentido pois os dados utilizados são mais recentes, e portanto estão, em muitos casos, mais próximos do cenário atual dos valores que seriam previstos.

Por fim, podemos afirmar para todos os cenários testados, que o modelo KRR foi o que melhor desempenhou comparado aos outros modelos. O modelo de regressão linear se torna mais viável quando utilizados dados mais recentes, e à medida que os dados para predição se afastam do valor dia do valor previsto, é preferível o uso dos outros modelos. Além desses pontos, essa pesquisa se mostrou favorável ao uso do SVR com relação ao ν-SVR, porém estudos futuros deverão ainda ser feitos, para estudar o comportamento desse modelo com o ajuste do parâmetro ν.

6.2 Trabalhos Futuros

Apesar de algumas técnicas terem sido abordadas, ainda existem muitas oportunidades de pesquisa na área de predição de ações no mercado de ações, tais quais:

- Utilizar dados além de 1 ano:
  Com isso seria possível analizar qual o impacto de um conjunto maior de dados para treinamento e teste dos modelos e como esse se ajustaria a grandes crises;

- Utilizar dados *intraday*:
  Isso nos permitiria verificar a capacidade dos algoritmos utilizados de prever valores ao longo do dia, podendo ser avaliados até casos de *high frequency trading*, os quais se passam em frações de segundo;

- Utilizar dados ajustados à inflação:
  Devido a tendência dos modelos apresentados de apresentarem valores positivos, e de, ao longo do tempo, a tendência de uma ação ser de alta por conta da inflação, seria interessante avaliar se os modelos se adequariam para cenários menos tendenciosos;

- Incorporar algoritmos de aprendizagem profunda:
  Alguns algoritmos de aprendizagem profunda como por exemplo o LSTM e o ANN também são utilizados para pesquisas na área de predição de ações e seus resultados poderiam ser comparados aos resultados obtidos pelos modelos de aprendizagem supervisionada;

- Acrescentar outros indicadores técnicos:
  Apesar de nesse artigo terem sido utilizados alguns indicadores técnicos conhecidos no mercado de renda variável, existem outros indicadores que poderiam ser acrescentados em pesquisas futuras;

- Fazer uso de notícias com técnicas de análise de sentimentos:
  Pelo fato do mercado de ações depender não só dos indicadores técnicos, mas também de como a empresa está sendo vista na mídia, seria interessante incorporar técnicas de análise de sentimento para os modelos, de forma que possamos relacionar acontecimentos a comportamentos observados;

- Fazer uso de indicadores fundamentalistas:
  O desempenho da empresa também pode ser útil para a predição dos seus valores, portanto utilizar indicadores fundamentalistas como por exemplo o indicador Preço/Lucro (P/L) e a margem *Earning Before Interests, Taxes, Depreciation and Amortization (EBITDA)* podem ser utilizados para tentar obter melhores resultados;

- Criar um *bot* para realizar operações de forma automática:
  Testar o modelo em um cenário real agregaria mais para a validação da pesquisa,

portanto seria de muito valor criar um *bot* que realize as operações de acordo com a predição e, no fim, analisar os resultados obtidos;

- Fazer cálculos recursivos dos valores das ações:
  Para esssa pesquisa, foram utilizados dados passados porém, já existentes. Fazer um cálculo recursivo dos valores das ações, seria interessante para verificar a capacidade do modelo de prever os valores para n dias futuros, utilizando apenas valores estimados pelo próprio modelo;

Referências

[1] Investimentos, Toro: *Mercado de ações - o que é e como funciona*. https://artigos.toroinvestimentos.com.br/mercado-de-acoes-como-funciona-curso. 1

[2] Wilder, J.W.: *New Concepts in Technical Trading Systems*. Trend Research, 1978, ISBN 9780894590276. https://books.google.com.br/books?id=WesJAQAAMAAJ. 1, 5

[3] Bollinger, John A.: *Bollinger on Bollinger Bands*. McGraw-Hill, 1ª edição, 2002, ISBN 9780071373685,0071373683. http://gen.lib.rus.ec/book/index.php?md5=D5972376AF9518FF5E665EB4203FA705. 1, 6, 7

[4] Lane, G.C. e Investment Educators (Firm): *Using Stochastics, Cycles & R.S.I.: –to the Moment of Decision –*. G.C. Lane, published under the auspices of Investment Educators, 1986. https://books.google.com.br/books?id=Qx7iPwAACAAJ. 1

[5] Hegazy, Osman, Omar S. Soliman e Mustafa Abdul Salam: *A machine learning model for stock market prediction*. International Journal of Computer Science and Telecommunications, 4:17–23, dezembro 2013. 1, 2, 9, 14, 16, 20

[6] Anbalagan, Thirunavukarasu e S. Uma Maheswari: *Classification and prediction of stock market index based on fuzzy metagraph*. Procedia Computer Science, 47:214–221, 2015, ISSN 1877-0509. https://www.sciencedirect.com/science/article/pii/S1877050915004688, Graph Algorithms, High Performance Implementations and Its Applications (ICGHIA 2014). 1, 10

[7] Gururaj, Vaishnavi, V. R. Shriya e Dr. Ashwini K: *Stock market prediction using linear regression and support vector machines*. 14(8):1931–1934, 2019. 1, 2, 15, 16, 21

[8] Ravikumar, S. e P. Saraf: *Prediction of stock prices using machine learning (regression, classification) algorithms*. Em *2020 International Conference for Emerging Technology (INCET)*, páginas 1–5, 2020. 1, 15, 20

[9] CVM: *Sobre a cvm*. http://www.cvm.gov.br/menu/acesso_informacao/institucional/sobre/cvm.html. 3

[10] Graham, Benjamin: *The intelligent investor rev ed*, 2003, ISBN 0060555661. 3

[11] Klinker, F.: *Exponential moving average versus moving exponential average*. Math Semesterber 58, páginas 97–107, 2011. https://doi.org/10.1007/s00591-010-0080-8. 5, 6

[12] Appel, G.: *Technical analysis: Power tools for active investors*. 2005. https://books.google.com.br/books?id=RFYIAAAACAAJ. 6

[13] Lane, G.C. e Investment Educators (Firm): *Using Stochastics, Cycles & R.S.I.: –to the Moment of Decision –*. G.C. Lane, published under the auspices of Investment Educators, 1986. https://books.google.com.br/books?id=Qx7iPwAACAAJ. 7

[14] StockCharts: *Stochastic oscillator*. https://school.stockcharts.com/doku.php?id=technical_indicators:stochastic_oscillator_fast_slow_and_full. 7

[15] Hayes, A. e Potters, C.: *Stochastic oscillator*. https://www.investopedia.com/terms/s/stochasticoscillator.asp. 7

[16] Mcculloch, Warren e Walter Pitts: *A logical calculus of ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, 5:127–147, 1943. 8

[17] Geron, Aurelien: *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 978-1-492-03264-9. O'reilly, 2nd edição, 2019, ISBN 978-1-492-03264-9. http://gen.lib.rus.ec/book/index.php?md5=40CA3F6E08498377145117D8B48BFD1B. 8

[18] Vapnik, V. N. e A. Ya. Chervonenkis: *On the uniform convergence of relative frequencies of events to their probabilities*. Theory of Probability and its Applications, 16(2):264–280, 1971. 8

[19] Vapnik, Vladimir N.: *Statistical Learning Theory*. Adaptive and learning systems for signal processing, communications, and control. Wiley, 1998, ISBN 9780471030034,0471030031. http://gen.lib.rus.ec/book/index.php?md5=0F9C36EC4D80683A3BE23BE6B8545E96. 8

[20] Awad, Mariette e Rahul Khanna: *Support Vector Regression*, páginas 67–80. janeiro 2015, ISBN 978-1-4302-5989-3. 8

[21] Evgeniou, Theodoros e Massimiliano Pontil: *Support vector machines: Theory and applications*. Volume 2049, páginas 249–257, janeiro 2001. 8

[22] Huh, Myung Hoe: *Kernel-trick regression and classification*. Communications for Statistical Applications and Methods, 22:201–207, março 2015. 9

[23] Scholkopf, B., Peter Bartlett e A. Smola: *Support vector regression with automatic accuracy control*. Proceedings of the 8th International Conference on Artificial Neural Networks, janeiro 2000. 9

[24] Chang, Chih Chung e Chih Jen Lin: *Training v-support vector regression: Theory and algorithms*. Neural Comput., 14(8):1959–1977, agosto 2002, ISSN 0899-7667. https://doi.org/10.1162/089976602760128081. 9

[25] Samui, P. e D.P. Kothari: *Utilization of a least square support vector machine (lssvm) for slope stability analysis*. Scientia Iranica, 18(1):53–58, 2011, ISSN 1026-3098. https://www.sciencedirect.com/science/article/pii/S1026309811000083. 10

[26] Liu, Kun e Bing Yu Sun: *Least squares support vector machine regression with equality constraints*. Physics Procedia, 24:2227–2230, 2012, ISSN 1875-3892. https://www.sciencedirect.com/science/article/pii/S1875389212003707, International Conference on Applied Physics and Industrial Engineering 2012. 10

[27] Schneider, Astrid, Gerhard Hommel e Maria Blettner: *Linear regression analysis part 14 of a series on evaluation of scientific publications*. Deutsches Ärzteblatt international, 107:776–82, novembro 2010. 10

[28] Visa, Sofia, Brian Ramsay, Anca Ralescu e Esther Knaap: *Confusion matrix-based feature selection*. Volume 710, páginas 120–127, janeiro 2011. 11

[29] Ting, Kai Ming: *Precision and Recall*, páginas 781–781. Springer US, Boston, MA, 2010, ISBN 978-0-387-30164-8. https://doi.org/10.1007/978-0-387-30164-8_652. 12, 13

[30] Xu Tao, He Renmu, Wang Peng e Xu Dongjie: *Input dimension reduction for load forecasting based on support vector machines*. Em *2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies. Proceedings*, volume 2, páginas 510–514 Vol.2, 2004. 14

[31] Stuke, Annika, Patrick Rinke e Milica Todorović: *Efficient hyperparameter tuning for kernel ridge regression with bayesian optimization*, abril 2020. 15, 16

[32] Kamble, R. A.: *Short and long term stock trend prediction using decision tree*. Em *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, páginas 1371–1375, 2017. 16

[33] Marek, Patrice e Věra Marková: *Optimization and testing of money flow index*. fevereiro 2020. 18

[34] Murphy, Kevin P.: *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning. The MIT Press, 1ª edição, 2012, ISBN 0262018020,9780262018029. http://gen.lib.rus.ec/book/index.php?md5=8ECFEEB2E1F9A19C770FBA1FF85FA566. 20

Anexo I

Gráficos com Resultados de Métricas

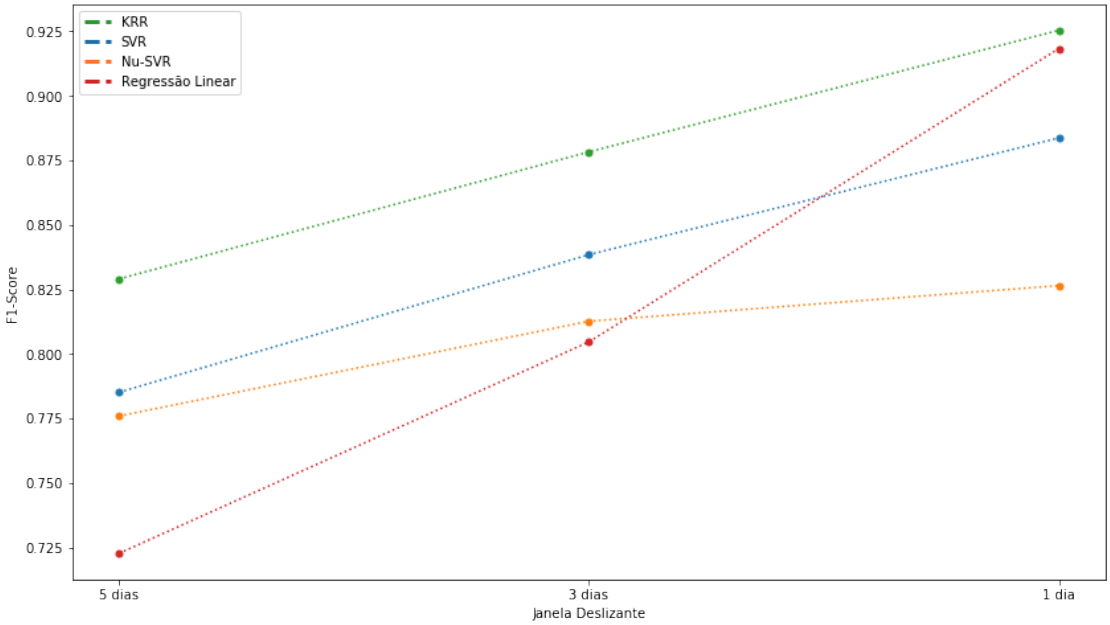

Figura I.1: Janela Deslizante x *F1-Score*

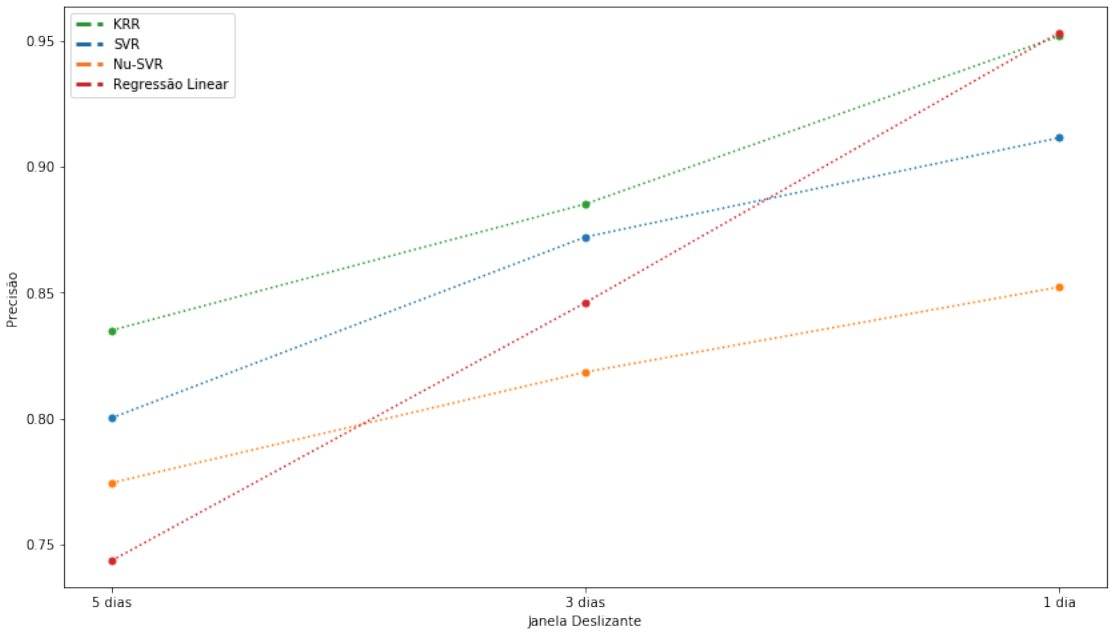

Figura I.2: Janela Deslizante x Precisão

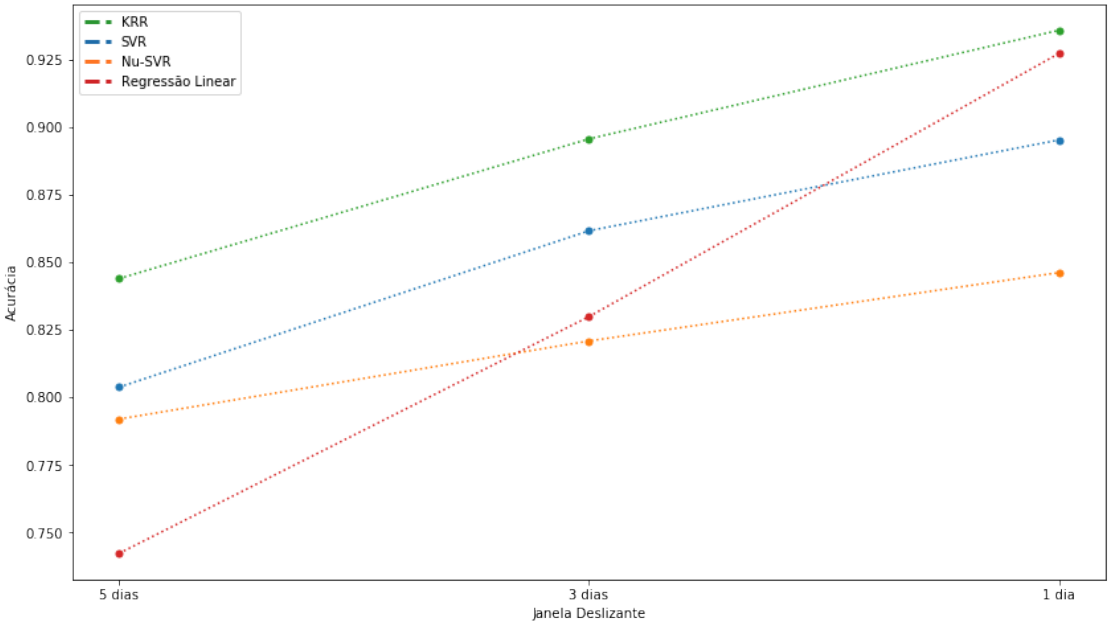

Figura I.3: Janela Deslizante x Acurácia

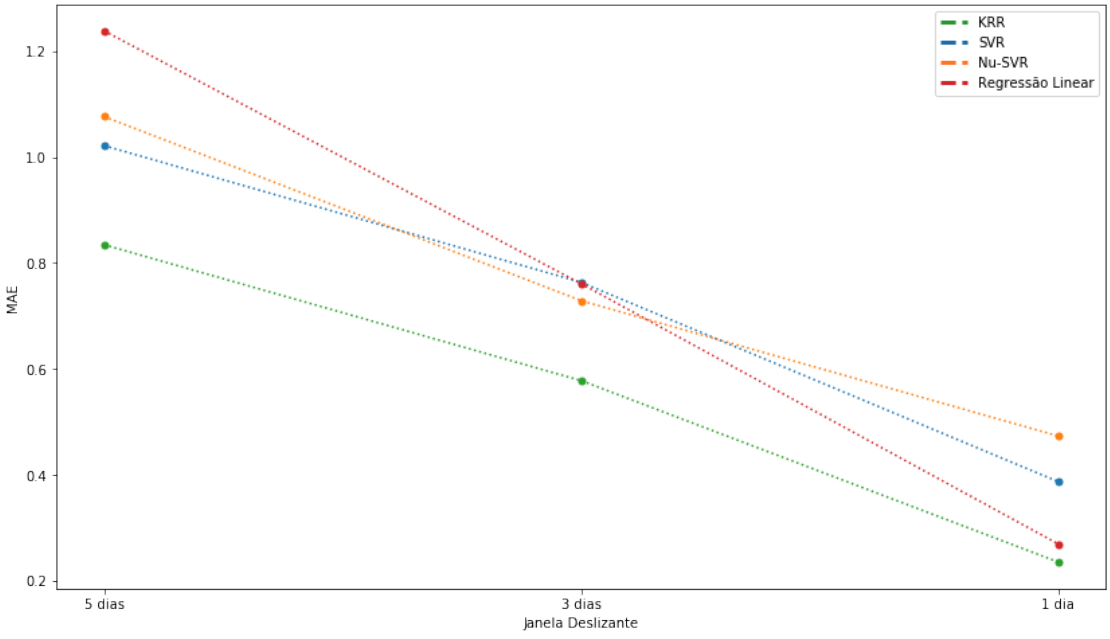

Figura I.4: Janela Deslizante x *MAE*

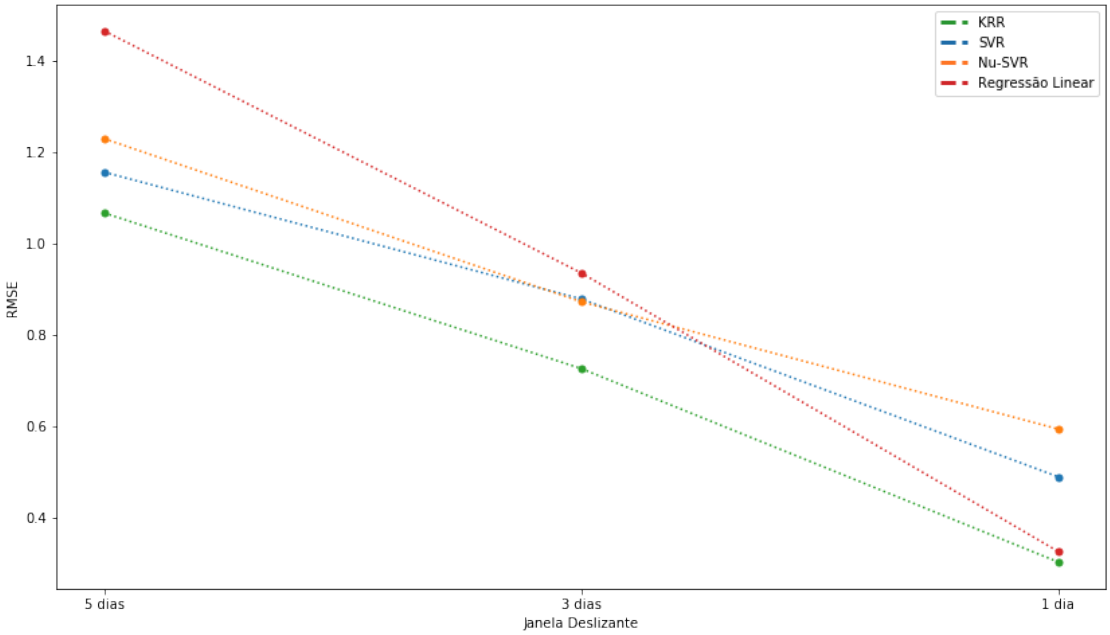

Figura I.5: Janela Deslizante x *RMSE*

Anexo II

Gráficos com Resultados de Valores Estimados

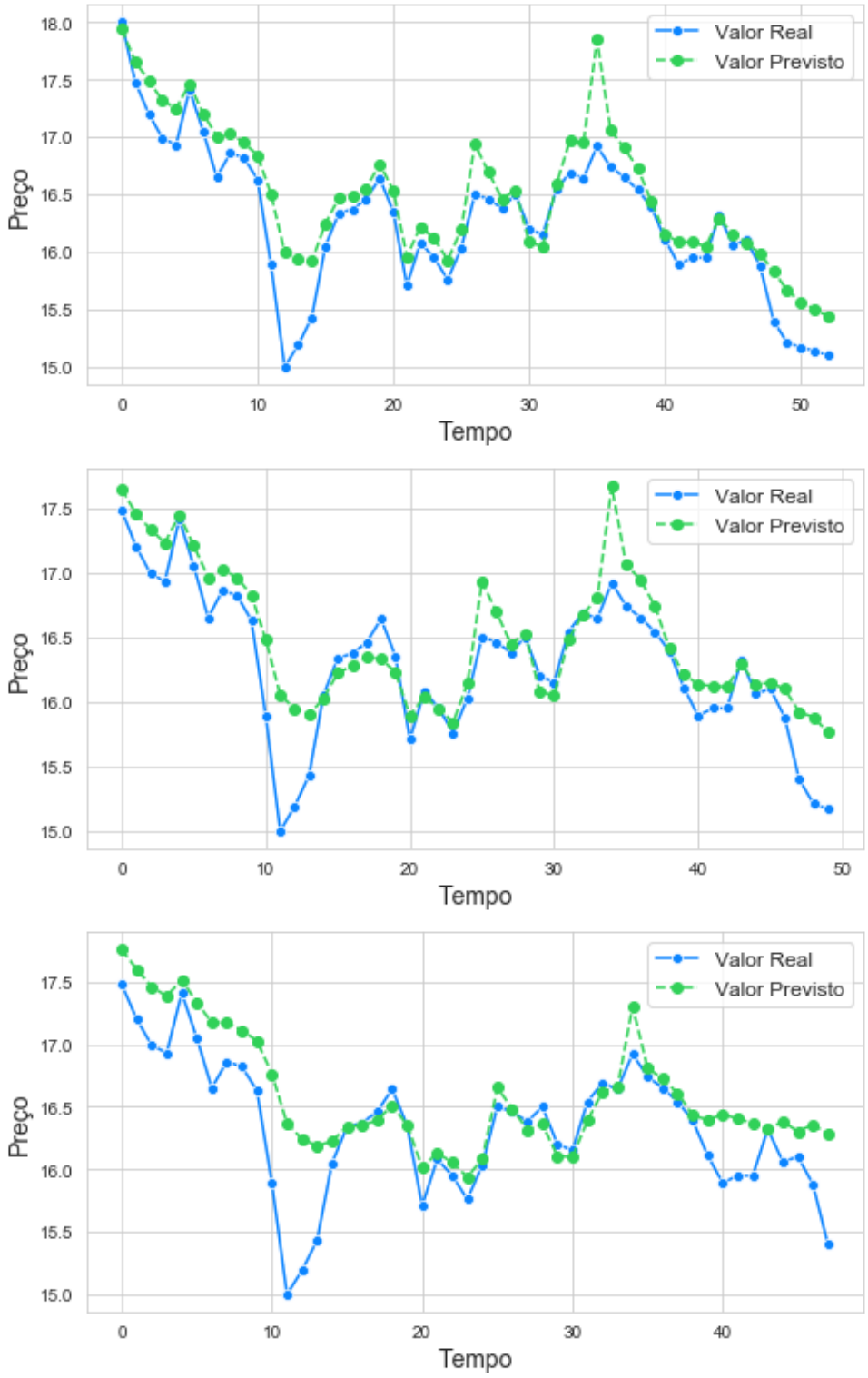

Figura II.1: Gráficos do modelo SVR com janela deslizante de 1, 3 e 5 dias da Ambev

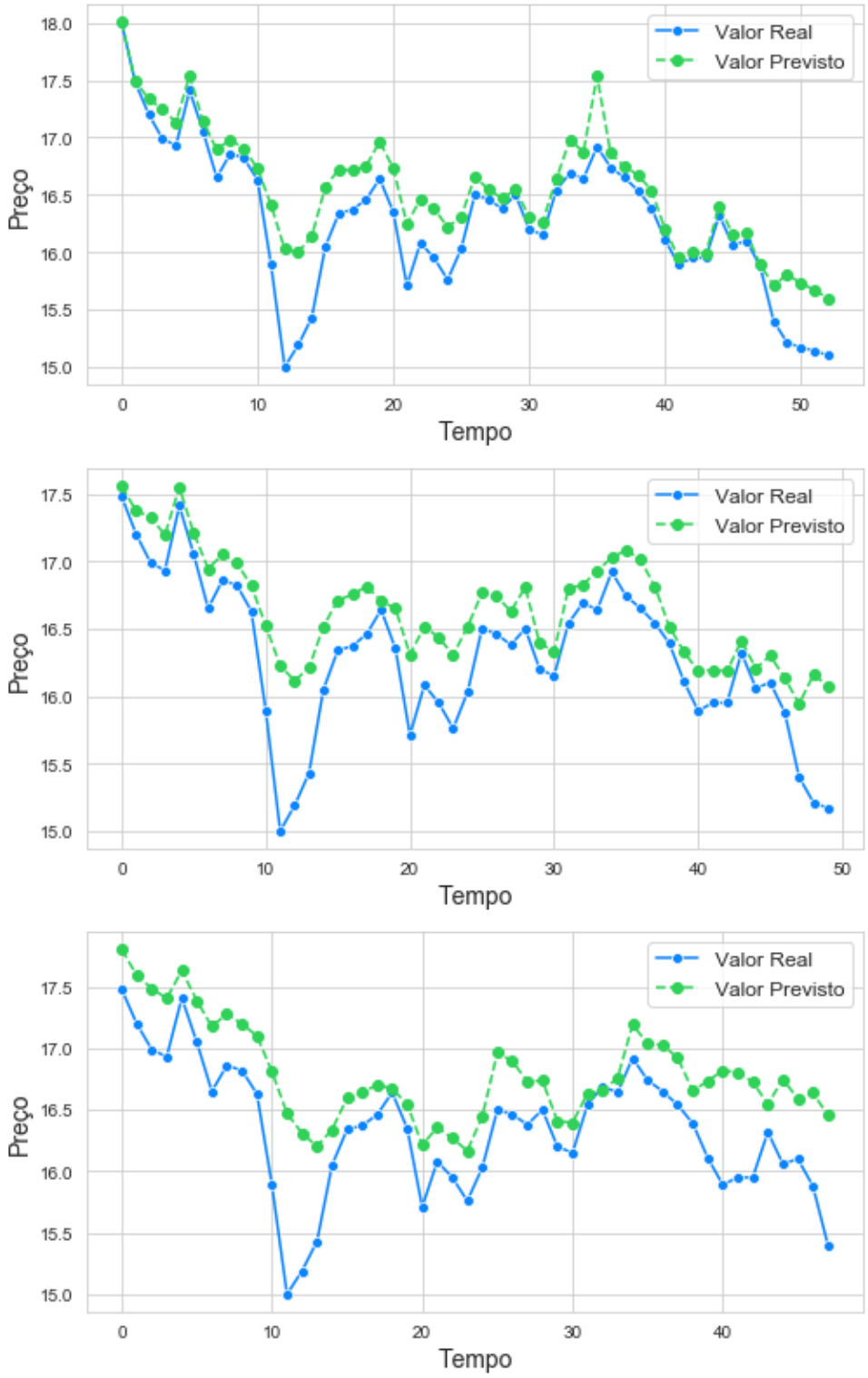

Figura II.2: Gráficos do modelo ν-SVR com janela deslizante de 1, 3 e 5 dias da Ambev

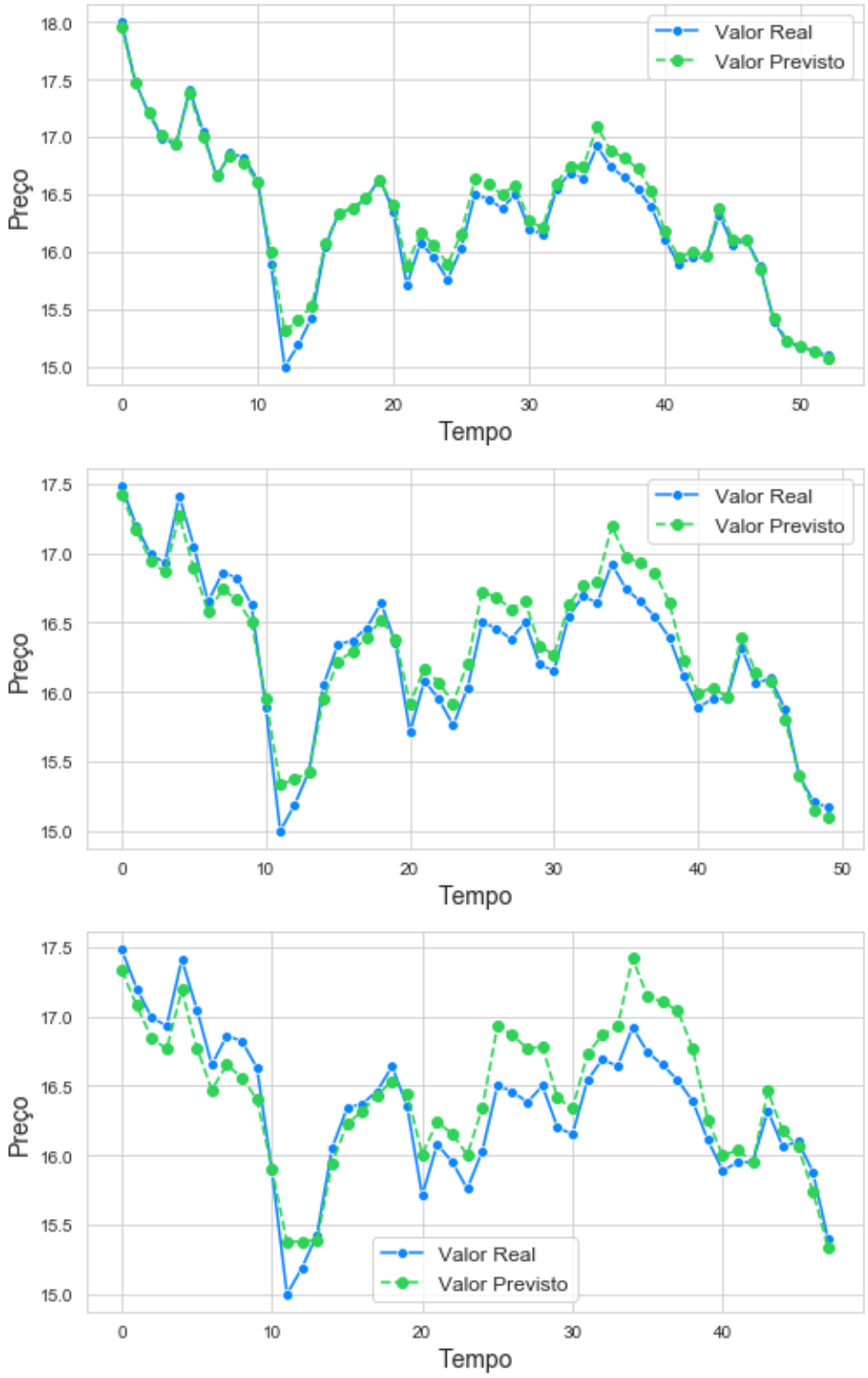

Figura II.3: Gráficos do modelo KRR com janela deslizante de 1, 3 e 5 dias da Ambev

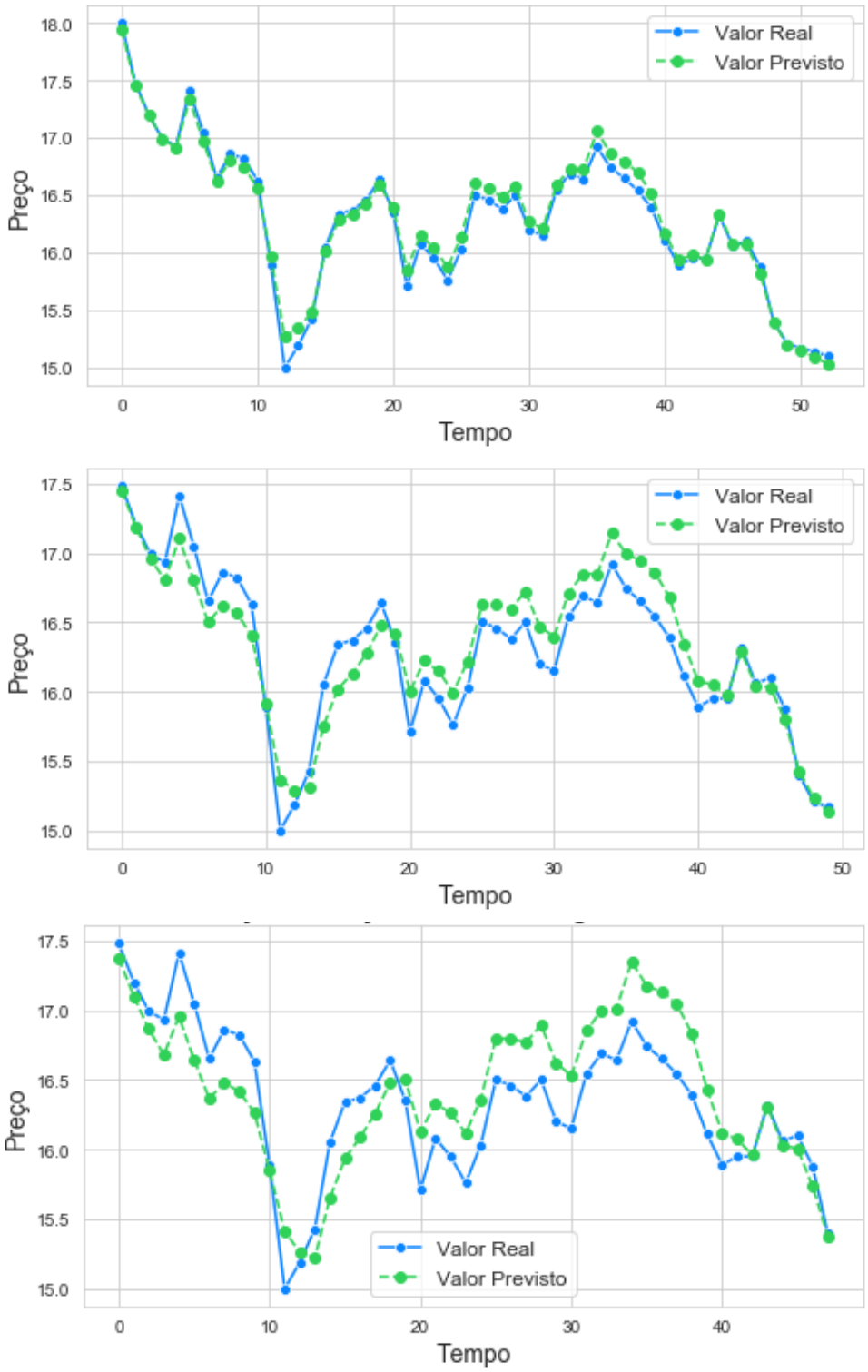

Figura II.4: Gráficos do modelo de Regressão Linear com janela deslizante de 1, 3 e 5 dias da Ambev