



**UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO**



Herculles Hendrius Coutinho Mesquita Silva

Previsão de Preço de Ações de Empresas do Setor Elétrico com Algoritmos de Aprendizado de Máquina

Recife

2024

Herculles Hendrius Coutinho Mesquita Silva

Previsão de Preço de Ações de Empresas do Setor Elétrico com Algoritmos de Aprendizado de Máquina

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Gabriel Alves

Recife

2024

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Ana Catarina Macêdo – CRB-4 1781

S586p Silva, Herculles Hendrius Coutinho Mesquita.
Previsão de preço de ações de empresas do setor elétrico com algoritmos de aprendizado de máquina / Herculles Hendrius Coutinho Mesquita Silva. - Recife, 2025.
30 f.; il.

Orientador(a): Gabriel Alves.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal Rural de Pernambuco, Bacharelado em Sistemas da Informação, Recife, BR-PE, 2025.

Inclui referências.

1. Aprendizado do computador. 2. Bolsa de valores. 3. Algoritmos computacionais . I. Alves, Gabriel, orient. II. Título

CDD 004

Herculles Hendrius Coutinho Mesquita Silva

Previsão de Preço de Ações de Empresas do Setor Elétrico com Algoritmos de Aprendizado de Máquina

Artigo apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 21 de Março de 2025.

BANCA EXAMINADORA

Nome do Orientador (Gabriel Alves)
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Roberta Gouveia
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Agradecimentos

Agradeço, aos meus pais, que sempre acreditaram na importância da educação e me apoiaram incondicionalmente ao longo de toda a minha vida. Por muitas vezes, abriram mão de várias coisas para garantir que eu tivesse acesso a uma educação de qualidade. Sou imensamente grato por todo o suporte, incentivo e amor que me proporcionaram.

Agradeço também à minha esposa, que foi uma fonte constante de motivação durante a minha jornada acadêmica. Mesmo nos momentos em que eu relaxei e perdi o foco, ela esteve ao meu lado, incentivando-me a seguir em frente e concluir o curso de graduação. Além disso, agradeço por tornar cada dia mais especial e por me ajudar a me tornar uma pessoa melhor.

Por fim, expresso minha gratidão ao Professor Gabriel, pelas orientações e dicas valiosas ao longo da elaboração deste trabalho de conclusão de curso. Sua dedicação como orientador e seus ensinamentos ao longo de todo o curso de bacharelado foram fundamentais para o meu crescimento acadêmico e profissional.

A todos, meu sincero agradecimento.

*“A persistência é o caminho do êxito.”
(Charles Chaplin)*

Resumo

O presente trabalho tem como objetivo comparar a eficiência de diferentes algoritmos de aprendizado de máquina na previsão de preços de ações do setor de energia elétrica. Para isso, foram analisados quatro algoritmos: Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Regressão Linear e Random Forest. Os dados utilizados compreendem uma série histórica de preços e indicadores adicionais, como inflação, Índice de energia elétrica (IEE) e variação cambial, que foram processados e utilizados como entrada para os modelos. A análise foi realizada com base em métricas de erro, como Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE) e Raiz do Erro Quadrático Médio (RMSE), bem como pela avaliação da diferença percentual entre os preços previstos e os valores reais. Os resultados mostram que o algoritmo LSTM obteve o melhor desempenho na previsão de preços de fechamento, seguido pela Regressão Linear, enquanto o Random Forest apresentou maior margem de erro e se mostrando inadequado para a aplicação neste problema. Este estudo realiza a aplicação de algoritmos preditivos no mercado financeiro, demonstrando o potencial do aprendizado de máquina como ferramenta para análise e tomada de decisão no setor de energia elétrica.

Abstract

The present study aims to compare the efficiency of different machine learning algorithms in predicting stock prices in the electric energy sector. Four algorithms were analyzed for this purpose: Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Linear Regression, and Random Forest. The dataset includes a historical series of stock prices and additional indicators such as inflation, the Electric Energy Index (IEE), and exchange rate variation, which were processed and used as input for the models. The analysis was conducted based on error metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), as well as the evaluation of the percentage difference between predicted and actual values. The results show that the LSTM algorithm achieved the best performance in forecasting closing prices, followed by Linear Regression, while Random Forest exhibited a higher margin of error, proving unsuitable for this problem. This study applies predictive algorithms in the financial market, demonstrating the potential of machine learning as a tool for analysis and decision-making in the electric energy sector.

Lista de tabelas

Tabela 1 – Metadados do Dataset de Ações da Bolsa de Valores.	17
Tabela 2 – Metadados do Dataset de Inflação (IPCA)	18
Tabela 3 – Metadados da base de dados da Taxa de Câmbio de Dólar para Real (USD/BRL)	18
Tabela 4 – Metadados do Dataset do Índice de Energia Elétrica (IEE)	19
Tabela 5 – Descrição das colunas do dataset consolidado	20
Tabela 6 – Descrição das Colunas para as Previsões das Ações de TAEE4 (TA-ESA)	22
Tabela 7 – Desempenho dos modelos nos diferentes critérios	22

Sumário

1	INTRODUÇÃO	8
1.1	Objetivos	9
2	FUNDAMENTAÇÃO TEÓRICA	10
2.0.1	Regressão Linear	10
2.0.2	Random Forest	10
2.0.3	Support Vector Regression (SVR)	11
2.0.4	LSTM	12
2.1	Impacto de Variáveis Macroeconômicas no Mercado de Ações	12
2.2	Métricas de Avaliação de Modelos	13
2.2.1	MAE	13
2.2.2	MSE	14
2.2.3	RMSE	14
2.2.4	Gráficos de CandleStick	14
2.2.5	Knowledge Discovery in Databases (KDD)	15
2.3	Trabalhos Relacionados	16
3	FERRAMENTAS E MÉTODOS	17
3.1	Seleção de Dados	17
3.2	Processamento e Transformação	19
4	RESULTADOS	21
4.1	Definição do Problema de Previsão	21
4.2	Discussões	24
5	CONCLUSÃO	26
5.1	Trabalhos Futuros	27
	REFERÊNCIAS	28

1 Introdução

O mercado financeiro é um dos pilares mais dinâmicos da economia global, desempenhando um papel crucial na alocação eficiente de recursos e na promoção do crescimento econômico. Dentro desse contexto, a Bolsa de Valores se destaca como um mercado organizado e regulamentado onde são negociados valores mobiliários, como ações, debêntures, fundos de investimento e derivativos. Sua principal função é fornecer um ambiente seguro e transparente para a compra e venda desses ativos, permitindo que empresas de capital aberto captem recursos por meio da emissão de ações e que investidores negociem esses ativos visando retornos financeiros. O comportamento dos preços das ações, porém, é altamente influenciado por fatores econômicos, políticos e especulativos, tornando a previsão de sua dinâmica uma tarefa desafiadora e de grande interesse para investidores e pesquisadores.

As empresas do setor de energia elétrica desempenham um papel relevante no mercado. Esse segmento é essencial para o funcionamento da sociedade e caracteriza-se pela estabilidade da demanda, mesmo durante períodos de incerteza econômica. Como resultado, as ações dessas empresas tendem a apresentar menor volatilidade quando comparadas a outros setores, sendo frequentemente consideradas uma opção segura para investidores em busca de retornos consistentes a longo prazo. No Brasil, esse setor se destaca pela presença de grandes empresas listadas na Bolsa de Valores

A previsão de preços de ações é uma área de estudo que combina conhecimentos de economia, estatística e tecnologia. Nos últimos anos, os avanços na área de aprendizado de máquina têm permitido a utilização de algoritmos sofisticados para modelar padrões complexos e identificar tendências ocultas nos dados financeiros. Essa abordagem tem se mostrado particularmente promissora para a análise de séries temporais, que representam a evolução de uma variável ao longo do tempo, como os preços de ações.

Este trabalho propõe uma análise comparativa de diferentes algoritmos de aprendizado de máquina aplicados à previsão de ações do setor de energia elétrica. Os algoritmos escolhidos incluem Long Short-Term Memory (LSTM), Support Vector Regression (SVR), Linear Regression e Random Forest, cada um com características e capacidades específicas para lidar com a complexidade dos dados financeiros. Além das variáveis intrínsecas do mercado de ações, como preços de abertura, fechamento e volume negociado, este estudo também considera indicadores macroeconômicos, como inflação e taxa de câmbio, ampliando a compreensão dos fatores que influen-

ciam os preços das ações no setor elétrico.

A escolha adequada das métricas de avaliação é essencial para determinar a precisão dos modelos de previsão de preços de ações. O Erro Absoluto Médio (MAE), são utilizados para medir a precisão dos modelos. [Montgomery, Peck e Vining \(2021\)](#), [Hastie, Tibshirani e Friedman \(2009\)](#).

1.1 Objetivos

Este trabalho tem como objetivo principal realizar a previsão de preços de ações de empresas do setor de energia elétrica listadas na B3, a bolsa de valores oficial do Brasil, onde são negociados ativos como ações, fundos imobiliários e derivativos. Analisam-se e comparam-se diferentes algoritmos de aprendizado de máquina, buscando avaliar o desempenho de LSTM, SVR, Linear Regression e Random Forest por meio de métricas como MAE, MSE e RMSE [Vazirani, Sharma e Sharma \(2020\)](#).

- Comparar os valores previstos com os valores reais e avaliar a margem de erro dos algoritmos.

- Melhorar a precisão dos algoritmos utilizando indicadores macroeconômicos, como inflação e taxa de câmbio [Barakat, Elgazzar e Hanafy \(2016\)](#).

- Identificar os algoritmos que apresentam desempenho superior neste cenário, fornecendo recomendações para aplicações futuras no campo da previsão de ações.

A partir dessa análise, espera-se contribuir para a escolha de métodos mais eficazes na previsão de preços de ações para investidores e pesquisadores que atuam no mercado financeiro brasileiro.

2 Fundamentação Teórica

A previsão de preços de ações é uma tarefa desafiadora devido à sua complexidade, alta volatilidade e natureza não-linear. Modelos de aprendizado de máquina têm se mostrado eficazes para lidar com essas características, possibilitando a extração de padrões complexos e a identificação de tendências em grandes volumes de dados [Nabipour et al. \(2020\)](#), [Schölkopf e Smola \(2002\)](#). Entre os algoritmos amplamente utilizados estão Regressão Linear, Random Forest, Support Vector Regression (SVR) e redes neurais profundas, como o Long Short-Term Memory (LSTM) [Goodfellow, Bengio e Courville \(2016\)](#), [James et al. \(2013\)](#).

2.0.1 Regressão Linear

A Regressão Linear é um modelo estatístico clássico que estabelece uma relação linear entre uma variável dependente y e uma ou mais variáveis independentes x_1, x_2, \dots, x_p . A equação geral do modelo é dada por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon,$$

onde β_0 é o intercepto, β_i representa os coeficientes das variáveis independentes e ϵ é o erro residual [Montgomery, Peck e Vining \(2021\)](#). Este modelo é amplamente utilizado devido à sua simplicidade e interpretabilidade.

Técnicas de regularização, como *Lasso* e *Ridge*, são frequentemente aplicadas para melhorar a capacidade do modelo de lidar com multicolinearidade e dados de alta dimensionalidade. O *Lasso* adiciona uma penalidade $\lambda \sum_{i=1}^p |\beta_i|$, promovendo a seleção de variáveis relevantes, enquanto o *Ridge* utiliza $\lambda \sum_{i=1}^p \beta_i^2$, que reduz a magnitude dos coeficientes, ajudando a controlar o overfitting [Hastie, Tibshirani e Friedman \(2009\)](#), [James et al. \(2013\)](#).

2.0.2 Random Forest

O Random Forest é um algoritmo de aprendizado supervisionado que combina múltiplas árvores de decisão para melhorar a precisão preditiva e reduzir o risco de overfitting [Breiman \(2001\)](#). Cada árvore é treinada em um subconjunto aleatório dos dados, e a predição final é obtida pela média (em problemas de regressão) ou pelo voto majoritário (em problemas de classificação).

O modelo é descrito matematicamente como:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x),$$

onde T é o número de árvores no ensemble, $h_t(x)$ é a predição da t -ésima árvore e \hat{y} é a predição final [Hastie, Tibshirani e Friedman \(2009\)](#). A capacidade do Random Forest de lidar com dados não-lineares e fornecer uma medida de importância das variáveis o torna particularmente adequado para cenários financeiros [Nabipour et al. \(2020\)](#).

2.0.3 Support Vector Regression (SVR)

O Support Vector Regression (SVR) é uma extensão do Support Vector Machine (SVM) para tarefas de regressão. Em vez de classificar os dados em diferentes categorias, o SVR busca encontrar uma função $f(x)$ que tenha no máximo uma margem de erro ϵ em relação aos valores reais, minimizando ao mesmo tempo a complexidade do modelo.

O problema de otimização do SVR pode ser formulado como:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*),$$

sujeito a:

$$\begin{aligned} y_i - (w^T x_i + b) &\leq \epsilon + \xi_i, \\ (w^T x_i + b) - y_i &\leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* &\geq 0, \end{aligned}$$

onde: - w representa o vetor de pesos do modelo, - b é o termo de viés (bias), - x_i e y_i são os valores de entrada e saída do conjunto de dados, - ϵ é um hiperparâmetro que define uma margem de tolerância ao erro dentro da qual as previsões não são penalizadas, - ξ_i e ξ_i^* são variáveis de folga para lidar com pontos que ultrapassam essa margem, - C é o parâmetro de regularização que controla o equilíbrio entre a complexidade do modelo e a quantidade de erros permitidos.

A escolha do kernel é um fator essencial no desempenho do SVR, pois define a transformação dos dados no espaço de características. Os principais tipos incluem: - **Linear**: utilizado quando os dados são aproximadamente lineares. - **Polinomial**: adequado para relações mais complexas entre as variáveis. - **Radial Basis Function (RBF)**: amplamente usado para capturar padrões não lineares.

Estudos como [Drucker Christopher J. Burges \(1997\)](#) e [Smola e Schölkopf \(2004\)](#) destacam a eficácia do SVR em tarefas de previsão, incluindo aplicações no mercado financeiro.

2.0.4 LSTM

O LSTM é uma arquitetura de rede neural projetada para capturar dependências temporais em séries temporais. Ele utiliza células de memória, controladas por três portas (de entrada, esquecimento e saída), que determinam quais informações devem ser retidas ou descartadas [Hochreiter e Schmidhuber \(1997\)](#), [Goodfellow, Bengio e Courville \(2016\)](#).

O estado da célula c_t é atualizado pela equação:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t,$$

onde f_t , i_t e \tilde{c}_t são as portas de esquecimento, entrada e memória candidata, respectivamente, e \odot representa a multiplicação elemento a elemento.

Essa capacidade de lidar com longas dependências temporais torna o LSTM particularmente eficaz na previsão de preços de ações, como demonstrado em [Nabipour et al. \(2020\)](#). Sua integração com variáveis exógenas, como taxas de câmbio e inflação, melhora ainda mais sua capacidade preditiva.

2.1 Impacto de Variáveis Macroeconômicas no Mercado de Ações

O comportamento dos mercados de ações é fortemente influenciado por variáveis macroeconômicas, como taxas de câmbio, inflação. Incorporar essas variáveis em modelos preditivos melhora significativamente a precisão das previsões [Barakat, Elgazzar e Hanafy \(2016\)](#). As taxas de câmbio afetam diretamente empresas exportadoras e importadoras, tornando essencial a consideração da volatilidade cambial em modelos preditivos, pois isso possibilita capturar melhor as flutuações dos mercados emergentes [Barakat, Elgazzar e Hanafy \(2016\)](#), [Kaggle \(\)](#). Da mesma forma, a inflação exerce influência sobre os custos operacionais e o poder de compra, o que impacta diretamente os preços das ações. A inclusão de índices como o IPCA auxilia na compreensão da relação entre condições econômicas e o comportamento do mercado acionário [Kaggle \(2023\)](#), [Barakat, Elgazzar e Hanafy \(2016\)](#). Além disso, mudanças na oferta monetária afetam a liquidez e o apetite por risco dos investidores. Modelos que incorporam essa variável podem prever com maior precisão as reações do mercado a diferentes políticas monetárias [Barakat, Elgazzar e Hanafy \(2016\)](#). Dessa forma, a análise dessas variáveis macroeconômicas permite um entendimento mais abrangente

dos fatores que impactam o mercado de ações e contribui para a construção de modelos preditivos mais robustos.

2.2 Métricas de Avaliação de Modelos

A avaliação de modelos é uma etapa essencial no desenvolvimento e validação de sistemas baseados em aprendizado de máquina. No contexto da análise de dados e modelagem estatística, métricas como *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE) e *Root Mean Squared Error* (RMSE) são amplamente utilizadas para quantificar a diferença entre os valores estimados pelo modelo (\hat{y}_i) e os valores reais (y_i) [Chai e Draxler \(2014\)](#). Essas métricas são fundamentais para comparar o desempenho de diferentes algoritmos de aprendizado de máquina, incluindo modelos de regressão e séries temporais, permitindo identificar a abordagem mais adequada para o problema em análise.

2.2.1 MAE

O MAE mede a média das diferenças absolutas entre os valores reais e previstos, sendo calculado como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

onde n é o número total de observações, y_i é o valor real da i -ésima observação e \hat{y}_i é o valor previsto correspondente.

Essa métrica é intuitiva e de fácil interpretação, representando o erro médio em termos absolutos. Por exemplo, se o MAE for 2, isso indica que, em média, o modelo erra por 2 unidades no conjunto de dados avaliado. Uma das vantagens do MAE é que ele trata erros positivos e negativos de forma igual, sem amplificar grandes desvios. Isso o torna ideal para situações em que pequenos e grandes erros têm a mesma importância, porque ele não eleva ao quadrado o erro, como será apresentado no MSE, dando o mesmo peso, para o erro maior e o erro menor. [Montgomery, Peck e Vining \(2021\)](#).

No entanto, o MAE pode não ser tão sensível a grandes desvios, o que pode ser uma desvantagem em contextos onde grandes erros são críticos, como na previsão de preços de ações em mercados voláteis.

2.2.2 MSE

O MSE calcula a média dos quadrados das diferenças entre os valores reais e previstos:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Diferentemente do MAE, o MSE penaliza mais fortemente os grandes erros, pois eleva ao quadrado as diferenças. Essa característica torna o MSE útil para identificar modelos que apresentam grandes desvios em algumas previsões. Por exemplo, um erro de 5 unidades terá um impacto de $5^2 = 25$ no MSE, enquanto no MAE o impacto seria apenas 5.

Por ser mais sensível a grandes desvios, o MSE é frequentemente utilizado em aplicações onde é importante minimizar grandes erros, como em sistemas de negociação financeira que precisam evitar perdas significativas [Hastie, Tibshirani e Friedman \(2009\)](#). No entanto, essa sensibilidade também pode ser uma desvantagem em casos onde outliers estão presentes, já que eles podem distorcer a avaliação do modelo.

2.2.3 RMSE

O RMSE é a raiz quadrada do MSE, calculado como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}.$$

Essa métrica combina a sensibilidade do MSE a grandes erros com a vantagem de manter a unidade dos dados originais, tornando-a mais interpretável. Por exemplo, se os valores previstos são preços em dólares, o RMSE também será expresso em dólares, o que facilita sua análise no contexto do problema [Chai e Draxler \(2014\)](#), [James et al. \(2013\)](#).

O RMSE é amplamente utilizado em finanças, pois é particularmente útil para prever valores absolutos, como preços de ações. Sua sensibilidade a grandes erros é uma vantagem em cenários onde desvios significativos têm implicações financeiras importantes. No entanto, assim como o MSE, o RMSE pode ser influenciado por outliers, o que deve ser levado em consideração ao interpretar os resultados.

2.2.4 Gráficos de CandleStick

O gráfico de Candlestick é uma ferramenta essencial na análise técnica de ativos financeiros, permitindo visualizar a variação dos preços ao longo de um período

específico. Originado no Japão no século XVIII, ele representa quatro valores principais: abertura, fechamento, máxima e mínima do período analisado. O corpo da vela mostra a diferença entre abertura e fechamento, sendo verde ou branca quando há valorização e vermelha ou preta quando há desvalorização. As sombras indicam os preços extremos alcançados.

Além de facilitar a identificação de tendências, os padrões formados pelos candlesticks ajudam a prever movimentos futuros do mercado. Padrões como Doji, Martelo e Engolfo indicam momentos de indecisão, reversão ou continuidade de tendência. Amplamente utilizado por traders e investidores, o gráfico de Candlestick oferece uma leitura clara da dinâmica do mercado e, quando combinado com outras ferramentas técnicas, melhora a precisão das estratégias de investimento.

2.2.5 Knowledge Discovery in Databases (KDD)

O processo de descoberta de conhecimento em bancos de dados, conhecido como Knowledge Discovery in Databases (KDD), é um conjunto de etapas metodológicas que guiam a extração de padrões úteis a partir de grandes volumes de dados. Esse conceito se consolidou como uma base fundamental para a mineração de dados e o aprendizado de máquina, fornecendo diretrizes estruturadas para transformar dados brutos em conhecimento acionável [Fayyad, Piatetsky-Shapiro e Smyth \(1996\)](#).

O KDD é composto por diversas etapas interdependentes, que incluem:

- **Seleção:** escolha das fontes de dados relevantes para a análise.
- **Pré-processamento:** limpeza e tratamento dos dados para eliminar ruídos e inconsistências.
- **Transformação:** conversão dos dados em um formato adequado para a modelagem.
- **Mineração de Dados:** aplicação de algoritmos de aprendizado de máquina ou estatísticos para a extração de padrões significativos.
- **Interpretação e Avaliação:** análise dos resultados obtidos, verificando sua relevância e aplicabilidade no contexto do estudo.

A utilização do KDD neste trabalho permite uma abordagem sistemática para a previsão de preços de ações do setor elétrico, garantindo que os dados passem por um fluxo organizado de processamento antes da aplicação dos modelos de aprendizado de máquina.

2.3 Trabalhos Relacionados

É abordado três estudos relacionados à previsão de mercados de ações utilizando aprendizado de máquina e aprendizado profundo. Além disso, é discutido como cada um desses trabalhos pode contribuir para a análise de ações de energia elétrica no contexto da B3.

O trabalho apresentado por [Nabipour et al. \(2020\)](#) analisou a aplicação de diferentes algoritmos de aprendizado de máquina e aprendizado profundo para prever valores futuros de grupos de ações na Bolsa de Valores de Teerã. Foram avaliados modelos baseados em árvores, como Decision Tree, Bagging, Random Forest, Ada-Boost, Gradient Boosting e Extreme Gradient Boosting (XGBoost), Além de redes neurais como a Rede Neural Artificial (ANN) e a Rede Neural Recorrente (RNN) e LSTM. O estudo utilizou 10 anos de dados históricos, e os resultados mostraram que o modelo LSTM teve o melhor desempenho, devido à sua capacidade de capturar padrões temporais complexos. Este trabalho contribui diretamente para a análise de ações de energia elétrica ao demonstrar a eficácia de algoritmos avançados em séries temporais financeiras, especialmente o LSTM, que pode ser utilizado para capturar padrões de sazonalidade e variações específicas do setor.

O estudo realizado por [Barakat, Elgazzar e Hanafy \(2016\)](#) investigou o impacto de eventos econômicos em mercados financeiros, aplicando técnicas estatísticas e modelos de aprendizado de máquina. Embora o foco tenha sido mercados em desenvolvimento, os autores destacaram a importância de utilizar algoritmos robustos, como Random Forest e Support Vector Machine, para reduzir o impacto de ruídos nos dados. Para a análise de ações do setor elétrico, este estudo reforça a necessidade de incluir variáveis macroeconômicas e setoriais, como políticas regulatórias e indicadores de mercado, para aumentar a precisão preditiva dos modelos.

Por fim, [Vazirani, Sharma e Sharma \(2020\)](#) apresentaram uma análise comparativa de modelos de aprendizado de máquina aplicados à previsão de preços de ações, destacando um modelo híbrido que combina dois algoritmos de aprendizado de máquina, como Linear Regression e Random Forest. O trabalho utilizou métricas como MAE, MSE e RMSE para avaliar a precisão dos modelos e concluiu que a combinação de dois algoritmos de regressão linear resultou em previsões mais precisas, com erro significativamente reduzido. Este estudo é relevante para a análise de ações de energia elétrica, pois reforça a ideia de que modelos híbridos podem melhorar a precisão preditiva ao integrar diferentes abordagens. Além disso, o uso de métricas robustas para avaliação de modelos demonstra a importância de medir a eficácia dos algoritmos no contexto de séries temporais financeiras.

3 Ferramentas e Métodos

Nesta seção, é detalhado as etapas realizadas durante o desenvolvimento deste projeto, com foco na análise e comparação de algoritmos de aprendizado de máquina aplicados ao setor elétrico. A abordagem adotada baseia-se no processo de descoberta de conhecimento em bases de dados Knowledge Discovery in Databases (KDD), conforme descrito no referencial teórico. As etapas são apresentadas de forma prática, com ênfase nos métodos e técnicas utilizados, desde o pré-processamento dos dados até a avaliação dos resultados obtidos. O processo de descoberta de conhecimento em bases de dados KDD surgiu no final da década de 1980 e início dos anos 1990, como uma abordagem estruturada para extrair conhecimento útil a partir de grandes volumes de dados. De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), o KDD é um processo iterativo e interativo que envolve diversas etapas, incluindo a seleção dos dados, pré-processamento, transformação, modelagem e interpretação dos resultados.

3.1 Seleção de Dados

Para a realização deste trabalho, foram utilizados quatro base de dados distintas, que, em conjunto, permitiram analisar o comportamento das ações do setor elétrico, considerando variáveis econômicas e de mercado. Abaixo, será detalhado as características de cada conjunto de dados selecionados. O primeiro dataset utilizado foi obtido com informações sobre ações listadas na Bolsa de Valores Brasileira (B3). Inicialmente, o conjunto de dados continha 894.776 registros abrangendo mais de 400 ações entre os anos de 2010 e 2020. As colunas disponíveis no dataset estão descritas na Tabela 1.

Tabela 1 – Metadados do Dataset de Ações da Bolsa de Valores.

Coluna	Descrição	Tipo	Valores
Date	Data da observação	datetime	04/01/2010 - 31/12/2020
Symbol	Símbolo da ação na bolsa	string	
Adj Close	Preço de fechamento ajustado da ação	float	-44.54 - 44594.76
Close	Preço de fechamento da ação	float	0.01 - 49200.49
High	Preço mais alto da ação no dia	float	0.00 - 49500.50
Low	Preço mais baixo da ação no dia	float	0.01 - 45600.46
Open	Preço de abertura da ação	float	0.01 - 49200.49
Volume	Quantidade de ações negociadas no dia	float	0.00 - 698950612.00

1

O segundo dataset utilizado foi baseado no Índice Nacional de Preços ao Consumidor Amplo (IPCA), que mede a inflação no Brasil. Esse conjunto de dados contém informações mensais entre os anos de 2010 e 2020, com colunas que permitem análises temporais detalhadas. As descrições das colunas estão apresentadas na Tabela 2.

Tabela 2 – Metadados do Dataset de Inflação (IPCA)

Coluna	Descrição	Tipo	Valores
ano	Ano de referência	Integer	2010 - 2020
mes	Mês de referência	String	janeiro - dezembro
indice	Valor do índice acumulado	Float	3040,22 - 5560,59
no_mes	Varição do índice no mês	Float	-0,38 - 1,35
3_meses	Varição acumulada em 3 meses	Float	-0,62 - 3,83
6_meses	Varição acumulada em 6 meses	Float	0,10 - 6,17
ano.1	Varição acumulada no ano	Float	-0,16 - 10,67
12_meses	Varição acumulada nos últimos 12 meses	Float	1,88 - 10,71

2

O terceiro dataset contém informações diárias sobre o câmbio de Dólar Americano (USD) para o Real Brasileiro (BRL), abrangendo o período de 2010 a 2020. Este conjunto de dados foi utilizado para adicionar uma perspectiva econômica às análises realizadas. As colunas e suas descrições estão detalhadas na Tabela 3.

Tabela 3 – Metadados da base de dados da Taxa de Câmbio de Dólar para Real (USD/BRL)

Coluna	Descrição	Tipo	Valores
datetime	Data da observação	datetime	04/01/2010 - 31/12/2020
usd_brl	Valor do dólar americano em reais	float	1.5345 - 5.9372

3

O quarto dataset contém informações sobre o Índice de Energia Elétrica (IEE), utilizado para representar o comportamento do setor elétrico na bolsa de valores brasileira. Este conjunto de dados inclui informações diárias entre 2010 e 2020, permitindo análises detalhadas de mercado. As colunas disponíveis estão descritas na Tabela 4.

¹ Disponível em: <<https://www.kaggle.com/datasets/andrewmvd/brazilian-stock-market>>. Acesso em: 05 mar. 2025.

² Disponível em: <<https://www.kaggle.com/datasets/brenojesusfernandes/ipca-index>>. Acesso em: 05 mar. 2025.

³ Disponível em: <<https://www.kaggle.com/datasets/felsal/ibovespa-stocks?select=usd2brl.csv>>. Acesso em: 05 mar. 2025.

Tabela 4 – Metadados do Dataset do Índice de Energia Elétrica (IEE)

Coluna	Descrição	Tipo	Valores
lee_Data	Data da observação	Date	04/01/2010 - 30/12/2020
lee_Ultimo	Valor de fechamento do índice	Float	20873,27 - 82845,94
lee_Abertura	Valor de abertura do índice	Float	20877,14 - 82493,31
lee_Maxima	Valor máximo do índice no dia	Float	21080,43 - 83052,71
lee_Minima	Valor mínimo do índice no dia	Float	20671,65 - 82211,84
lee_Variacao	Variação percentual no dia	Float	-11,59 - 8,99

⁴ para análise A combinação desses quatro conjuntos de dados permitiu realizar análises robustas sobre o comportamento das ações do setor elétrico, incorporando variáveis de mercado, inflação e câmbio para avaliar a influência de fatores econômicos nas tendências observadas.

3.2 Processamento e Transformação

O processo de processamento e transformação seguiu as etapas do KDD, que envolvem a seleção, pré-processamento, transformação, mineração e interpretação dos dados. Inicialmente, os dados relevantes foram selecionados, seguidos pelo pré-processamento, que incluiu a remoção de registros faltantes e a adaptação das variáveis para uma estrutura compatível. Na transformação, os dados foram normalizados e ajustados à granularidade necessária, de maneira diária.

Para garantir a utilização uniforme e padronizada dos diferentes bases de dados selecionadas, foi necessário realizar diversas etapas de processamento e transformação dos dados. Inicialmente, identificou-se que alguns registros apresentavam datas faltantes, especialmente em períodos mais antigos. Para alinhar os dados de forma consistente, todos os conjuntos de dados foram filtrados para o intervalo entre 4 de janeiro de 2010 e 31 de dezembro de 2020, período no qual todos os datasets estavam completos. O dataset final resultante contém 642.456 linhas, abrangendo 294 ações diferentes. É importante ressaltar que a granularidade dos dados é diária. Porém, há dados apenas nos dias em que a bolsa de valores está aberta para compra e venda, o que ocorre em dias úteis.

Em prol de uma análise mais robusta e menos suscetível a oscilações extraordinárias do mercado, optou-se por concentrar a análise no setor de energia elétrica. Este setor é tradicionalmente mais estável, com menor sensibilidade a quedas ou subidas abruptas em comparação a outros setores econômicos, como tecnologia ou finanças.

⁴ Disponível em: <<https://br.investing.com/indices/electric-power-historical-data>>. Acesso em: 05 mar. 2025.

Essa abordagem proporcionou maior confiabilidade nas análises realizadas, especialmente ao comparar diferentes algoritmos de aprendizado de máquina.

Os dados foram organizados de forma que cada registro representasse um único dia, garantindo uma estrutura temporal uniforme para as análises. Os datasets de câmbio, IEE e B3 já estavam disponíveis com granularidade diária, permitindo sua incorporação direta ao modelo. No entanto, o IPCA estava disponível apenas em frequência mensal, o que exigiu sua adaptação. Para compatibilizar essa informação com a estrutura diária dos demais dados, o valor mensal foi distribuído proporcionalmente entre os dias do respectivo mês. Esse ajuste assegurou a integração consistente das informações, permitindo uma melhor correlação entre os diferentes fatores analisados e evitando discrepâncias decorrentes de periodicidades distintas. Dessa forma, a padronização contribuiu para tornar os modelos mais alinhados à dinâmica do mercado ao longo do tempo.

A base de dados final consolidado contém as colunas apresentadas no Quadro 5:

Quadro 5 – Descrição das colunas do dataset consolidado

Coluna	Descrição
Date	Data da observação.
Symbol	Símbolo da ação na bolsa.
Adj Close	Preço de fechamento ajustado da ação.
Close	Preço de fechamento da ação.
High	Preço mais alto da ação no dia.
Low	Preço mais baixo da ação no dia.
Open	Preço de abertura da ação.
Volume	Quantidade de ações negociadas no dia.
usd_brl	Valor do dólar americano em reais no dia.
lee_Ultimo	Valor de fechamento do Índice de Energia Elétrica (IEE).
lee_Abertura	Valor de abertura do Índice de Energia Elétrica (IEE).
lee_Maxima	Valor máximo do Índice de Energia Elétrica (IEE) no dia.
lee_Minima	Valor mínimo do Índice de Energia Elétrica (IEE) no dia.
lee_Variacao	Variação percentual diária do Índice de Energia Elétrica (IEE).
no_mes	Variação percentual do IPCA no mês.
12_meses	Variação percentual acumulada do IPCA nos últimos 12 meses.

O processo de integração e transformação garantiu que todos os dados fossem harmonizados e prontos para aplicação dos algoritmos de aprendizado de máquina, oferecendo uma base sólida e confiável para análises subsequentes.

4 Resultados

Neste capítulo, são apresentados e analisados os resultados obtidos a partir da aplicação de modelos de aprendizado de máquina na previsão dos preços das ações. Para exemplificar a análise dos dados, foram utilizadas as ações da empresa Transmissora Aliança de Energia Elétrica S.A. (TAESA), uma das principais companhias do setor de transmissão de energia elétrica no Brasil. A ação da TAESA com código TAEE4 representa as ações preferenciais da TAESA, que oferecem prioridade no pagamento de dividendos, que consiste em todas as ações com o número 4 no final. Onde não possuem direito a voto nas decisões da empresa. Esse tipo de ativo é amplamente negociado no mercado financeiro e atrai investidores interessados em retornos consistentes por meio de dividendos periódicos.

4.1 Definição do Problema de Previsão

O objetivo principal é desenvolver modelos de aprendizado de máquina capazes de prever o comportamento diário das ações da TAESA (TAEE4), utilizando dados históricos do mercado financeiro e variáveis macroeconômicas.

A previsão abrange o preço de fechamento (Close), o preço de abertura (Open), o preço máximo do dia (High), o preço mínimo do dia (Low) e o volume de negociações (Volume). A escolha dessas variáveis se justifica pela sua relevância na análise de investimentos, auxiliando na tomada de decisões estratégicas no mercado acionário. Além disso, a previsão do volume de negociações é necessária, pois essa variável influencia diretamente a dinâmica dos preços ao longo do tempo.

O processo de modelagem considera a relação entre essas variáveis e fatores externos, como câmbio (usd_brl), Índice de Energia Elétrica (IEE) e inflação (IPCA), que exercem influência sobre os preços das ações. Para garantir que as previsões sejam avaliadas de forma objetiva, foram utilizadas métricas de erro amplamente reconhecidas na literatura, como erro médio absoluto (MAE), erro quadrático médio (MSE) e raiz do erro quadrático médio (RMSE), além da diferença percentual média entre os valores previstos e os valores reais.

A seguir, são apresentados os resultados obtidos para cada um dos algoritmos testados, analisando seu desempenho na previsão dos preços das ações. A avaliação foi realizada com base na diferença percentual entre os valores previstos e os valores reais, permitindo uma comparação objetiva do desempenho de cada modelo.

Quadro 6 – Descrição das Colunas para as Previsões das Ações de TAEE4 (TAESA)

Coluna	Descrição
Date	Data correspondente ao registro.
Close	Valor original do preço de fechamento.
Close_Prev	Previsão do algoritmo para o preço de fechamento.
Close_Diff	Diferença percentual entre o valor previsto e o valor real do preço de fechamento.
Open	Valor original do preço de abertura.
Open_Prev	Previsão do algoritmo para o preço de abertura.
Open_Diff	Diferença percentual entre o valor previsto e o valor real do preço de abertura.
High	Valor original do preço máximo do dia.
High_Prev	Previsão do algoritmo para o preço máximo do dia.
High_Diff	Diferença percentual entre o valor previsto e o valor real do preço máximo do dia.
Low	Valor original do preço mínimo do dia.
Low_Prev	Previsão do algoritmo para o preço mínimo do dia.
Low_Diff	Diferença percentual entre o valor previsto e o valor real do preço mínimo do dia.
Volume	Volume total de transações registrado no dia.

Os valores gerados no quadro 6 foram utilizados para comparar os algoritmos de aprendizado de máquina, analisando principalmente as variáveis de diferença percentual (Diff), que indicam o desvio entre os valores previstos e os valores reais. Essa abordagem permitiu avaliar a precisão de cada modelo na previsão dos preços das ações e entender suas limitações em relação às variações diárias do mercado.

Na tabela 7 estão os resultados para todos os algoritmos, foi selecionado o intervalo entre 14 de novembro de 2020 e 14 de dezembro de 2020. estes dados foram selecionados para comparar a precisão dos modelos de aprendizado de máquina.

Tabela 7 – Desempenho dos modelos nos diferentes critérios

Métrica	LSTM	SVR	Regressão Linear	Random Forest
Close_Diff (%)	-2,69	-18,08	-2,92	416,60
Open_Diff (%)	-20,08	-18,08	-20,19	325,79
High_Diff (%)	-20,51	-17,61	-20,57	321,69
Low_Diff (%)	-19,24	-18,32	-19,48	330,60
MAE	0.1371	0.0706	0.10	1.8233
MSE	0.0322	0.0094	0.02	8.0102
RMSE	0.1794	0.0969	0.12	2.8302

Apresentamos os gráficos das ações da TAESA (TAEE4) analisadas nos últimos 60 dias, este intervalo foi selecionado para obter uma visualização melhor dos dados utilizando os gráficos. Esses gráficos representam a movimentação dos preços,

utilizando o formato de candlestick. A coluna de Preço a unidade é em Real Brasileiro e o volume a quantidade de ações negociadas em bolsa.

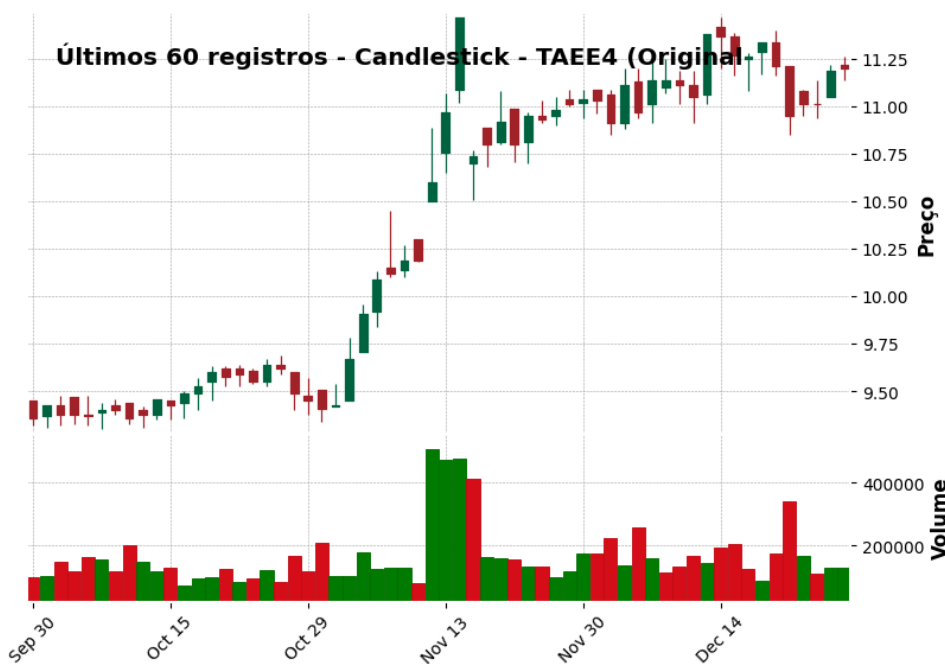
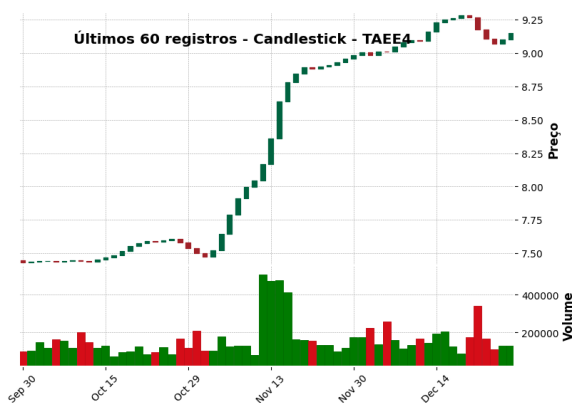
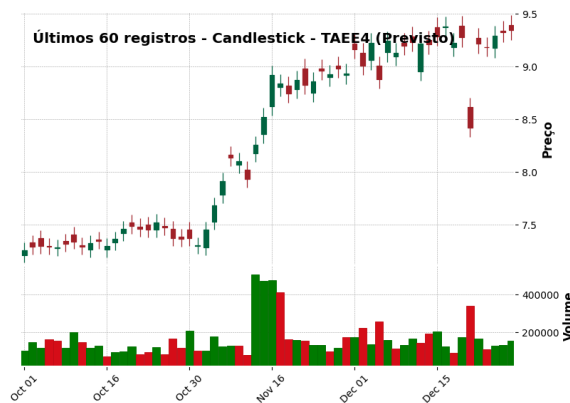


Gráfico 1 – Gráfico candlestick dos últimos 60 dias (valores originais das ações TAE4).

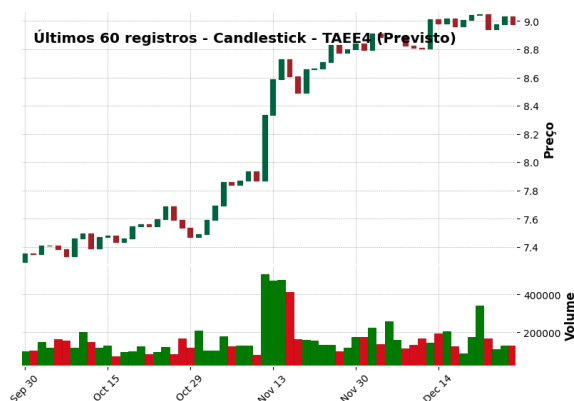


(a) LSTM

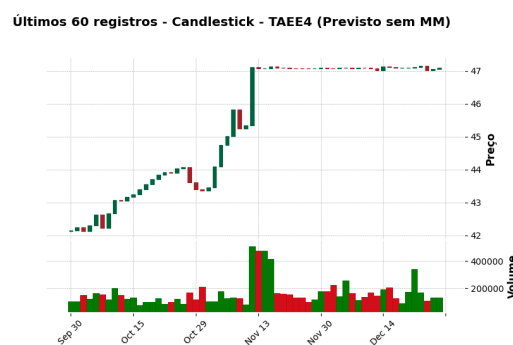


(b) SVR

Gráfico 2 – Gráficos candlestick gerados pelos modelos LSTM e SVR.



(a) Regressão Linear



(b) Random Forest

Gráfico 3 – Gráficos candlestick gerados pelos modelos Regressão Linear e Random Forest.

4.2 Discussões

Após a análise dos resultados obtidos para cada um dos modelos testados, é possível identificar diferenças significativas no desempenho das previsões. A avaliação das métricas de erro e das diferenças percentuais entre os valores previstos e os valores reais permitiu uma comparação direta entre os algoritmos e sua aplicabilidade na previsão de preços de ações.

Os resultados mostraram que o modelo LSTM apresentou o melhor desempenho, com uma diferença percentual média de -2,69% para o preço de fechamento. Esse resultado sugere que o modelo foi capaz de capturar padrões temporais e dependências não lineares nos dados, tornando-o mais eficiente para a previsão de séries temporais financeiras. A regressão linear também obteve um desempenho satisfatório, com uma diferença percentual média de -2,92%, evidenciando sua capacidade de modelar relações lineares entre as variáveis.

O SVR, por outro lado, apresentou uma diferença percentual média de -18,08%, indicando dificuldades na captura da complexidade dos dados financeiros. Isso pode ser atribuído à escolha do kernel e à sensibilidade do modelo a hiperparâmetros, o que pode limitar seu desempenho quando aplicado a séries temporais com alta volatilidade.

O modelo Random Forest foi o que apresentou os piores resultados. Com uma diferença percentual média de 416,60% para o preço de fechamento e desvios semelhantes para as demais variáveis, ficou evidente que o modelo teve dificuldades extremas em gerar previsões coerentes. Esse comportamento pode estar relacionado ao overfitting do modelo, que pode ter memorizado padrões espúrios nos dados de treinamento, resultando em previsões extremamente inflacionadas. Dada essa limitação, o uso do Random Forest para previsão de preços de ações não se mostrou viável sem

ajustes significativos, como a regularização dos hiperparâmetros ou a reavaliação das features utilizadas no treinamento.

A partir dessas análises, é possível concluir que os modelos baseados em aprendizado profundo, como o LSTM, apresentam vantagens na previsão de séries temporais complexas, capturando padrões de longo prazo que algoritmos mais tradicionais, como regressão linear e SVR, podem não identificar com a mesma precisão. No entanto, cada modelo possui suas particularidades e pode ser mais adequado dependendo do contexto da previsão e do nível de complexidade dos dados.

Dessa forma, os resultados obtidos neste estudo reforçam a importância da escolha criteriosa do modelo a ser utilizado para previsão de preços de ações. O ajuste de hiperparâmetros, o uso de dados adicionais e a exploração de modelos híbridos são estratégias que podem ser consideradas para aprimorar a precisão das previsões e reduzir a sensibilidade a variações bruscas do mercado.

5 Conclusão

Este trabalho investigou a aplicação de diferentes algoritmos de aprendizado de máquina na previsão de preços de ações do setor elétrico, utilizando dados históricos da TAESA (TAEE4). O objetivo foi avaliar a capacidade desses modelos de prever o comportamento diário do mercado, considerando variáveis como preços de abertura, fechamento, máxima, mínima e volume de negociações.

A metodologia adotada seguiu o processo de descoberta de conhecimento em banco de dados (KDD), passando pelas etapas de coleta, processamento, transformação e modelagem dos dados. A avaliação foi conduzida com base em métricas de erro amplamente utilizadas na literatura, como erro médio absoluto (MAE), erro quadrático médio (MSE) e raiz do erro quadrático médio (RMSE), além da análise da diferença percentual entre os valores previstos e os valores reais.

Os resultados obtidos reforçam a importância da escolha adequada do modelo para a previsão de séries temporais financeiras, uma vez que diferentes algoritmos apresentaram desempenhos distintos na previsão dos preços das ações. A análise das previsões revelou que alguns modelos foram mais eficientes na captura de padrões temporais, enquanto outros demonstraram dificuldades em representar corretamente as variações do mercado.

Este estudo também evidencia a necessidade de melhorias contínuas nos modelos preditivos. O refinamento de hiperparâmetros, o uso de dados complementares e a exploração de arquiteturas mais avançadas combinando diferentes algoritmos, como Linear regression junto com o random forest, ou SVR, podem contribuir para aumentar a precisão das previsões.

Como perspectiva futura, a aplicação desses modelos em ambientes de negociação em tempo real pode representar um avanço significativo, permitindo uma análise dinâmica do mercado e a adaptação a novas condições. Além disso, a investigação de técnicas mais recentes, como modelos baseados em aprendizado profundo e arquiteturas como Transformer, pode trazer novas oportunidades para o aprimoramento das previsões.

Dessa forma, este trabalho contribui para a compreensão do uso de algoritmos de aprendizado de máquina na análise de séries temporais financeiras, destacando os desafios e oportunidades para futuras pesquisas na área de previsão do mercado de ações.

5.1 Trabalhos Futuros

Os resultados deste estudo indicam diversas possibilidades de aprimoramento para pesquisas futuras. Uma das principais direções envolve o refinamento dos modelos utilizados, especialmente por meio do ajuste de hiperparâmetros e da inclusão de novas variáveis que possam capturar melhor os fatores que influenciam os preços das ações, como a taxa SELIC e outros indicadores econômicos, como o produto interno bruto (PIB). Indicadores econômicos adicionais e fontes alternativas de dados, como análise de sentimento em redes sociais e notícias financeiras, podem contribuir para previsões mais precisas. Outro aspecto relevante é a aplicação desses modelos em cenários de negociação em tempo real, testando sua viabilidade prática no mercado financeiro. Além disso, técnicas de interpretabilidade, como Shapley Additive Explanations (SHAP) e Local Interpretable Model-agnostic Explanations (LIME), podem ser empregadas para tornar os modelos mais compreensíveis, permitindo uma melhor avaliação dos fatores que impactam as previsões.

Os resultados deste estudo indicam diversas possibilidades de aprimoramento para pesquisas futuras. Uma das principais direções envolve o refinamento dos modelos utilizados, especialmente por meio do ajuste de hiperparâmetros e da inclusão de novas variáveis que possam capturar melhor os fatores que influenciam os preços das ações, como a taxa SELIC e outros indicadores econômicos, como o produto interno bruto (PIB). Indicadores econômicos adicionais e fontes alternativas de dados, como análise de sentimento em redes sociais e notícias financeiras, podem contribuir para previsões mais precisas. Outro aspecto relevante é a aplicação desses modelos em cenários de negociação em tempo real, testando sua viabilidade prática no mercado financeiro. Além disso, técnicas de interpretabilidade, como Shapley Additive Explanations (SHAP) e Local Interpretable Model-agnostic Explanations (LIME), podem ser empregadas para tornar os modelos mais compreensíveis, permitindo uma melhor avaliação dos fatores que impactam as previsões.

Além disso, o Modelo de Precificação de Ativos de Capital (CAPM) pode ser integrado aos modelos preditivos, oferecendo uma abordagem adicional para avaliar o risco e o retorno esperado das ações. A utilização do CAPM pode ajudar a melhorar as previsões, ao considerar o comportamento das ações em relação ao mercado, oferecendo insights valiosos para a análise financeira e decisões de investimento.

Dessa forma, este estudo abre caminho para novas pesquisas que possam expandir a precisão e aplicabilidade dos modelos preditivos, contribuindo para o avanço das técnicas de aprendizado de máquina no mercado financeiro.

Referências

- BARAKAT, M. R.; ELGAZZAR, S. H.; HANAFY, K. M. Impact of macroeconomic variables on stock markets: Evidence from emerging markets. *International Journal of Economics and Finance*, Canadian Center of Science and Education, v. 8, n. 1, p. 195–207, 2016. Citado 3 vezes nas páginas 9, 12 e 16.
- BREIMAN, L. Random forests. *Machine Learning*, Springer, v. 45, n. 1, p. 5–32, 2001. Citado na página 10.
- CHAI, T.; DRAXLER, R. R. Root mean square error (rmse) or mean absolute error (mae) arguments against avoiding rmse in the literature. *Geoscientific Model Development*, Copernicus GmbH, v. 7, n. 3, p. 1247–1250, 2014. Citado 2 vezes nas páginas 13 e 14.
- DRUCKER CHRISTOPHER J. BURGESS, L. K. A. S. V. V. H. Support vector regression machines. In: *Advances in Neural Information Processing Systems (NIPS)*. [S.l.: s.n.], 1997. p. 155–161. Citado na página 12.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, n. 3, p. 37–54, 1996. Disponível em: <<https://ojs.aaai.org/index.php/aimagazine/article/view/1230>>. Citado na página 15.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Citado 2 vezes nas páginas 10 e 12.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning*. [S.l.]: Springer, 2009. Citado 4 vezes nas páginas 9, 10, 11 e 14.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 12.
- JAMES, G. et al. *An Introduction to Statistical Learning*. [S.l.]: Springer, 2013. Citado 2 vezes nas páginas 10 e 14.
- KAGGLE. *Dados de Conversão USD/BRL*. Acesso em: 19 de dezembro de 2024. Disponível em: <<https://www.kaggle.com/datasets/felsal/ibovespa-stocks?select=usd2brl.csv>>. Citado na página 12.
- KAGGLE. *IPCA Index Data*. 2023. Disponível em: <<https://www.kaggle.com/datasets/brenojesusfernandes/ipca-index>>. Citado na página 12.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to Linear Regression Analysis*. [S.l.]: John Wiley & Sons, 2021. Citado 3 vezes nas páginas 9, 10 e 13.
- NABIPOUR, M. et al. Deep learning for stock market prediction. *Entropy*, MDPI, v. 22, n. 8, p. 840, 2020. Citado 4 vezes nas páginas 10, 11, 12 e 16.
- SCHÖLKOPF, B.; SMOLA, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. [S.l.]: MIT Press, 2002. Citado na página 10.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and Computing*, v. 14, p. 199–222, 2004. Citado na página 12.

VAZIRANI, S.; SHARMA, A.; SHARMA, P. Analysis of various machine learning algorithm and hybrid model for stock market prediction using python. In: IEEE. *International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*. [S.l.], 2020. p. 204–207. Citado 2 vezes nas páginas 9 e 16.