

Previsão de preços de ações e ETF na bolsa de valores B3 aplicando técnicas de *machine learning*

 <https://doi.org/10.47236/2594-7036.2026.v10.1879>

Eduardo Luiz Zanotto¹
Carlos Amaral Hölbig²

Data de submissão concluída: 10/10/2025. Data de aprovação: 16/12/2025. Data de publicação: 19/1/2026.

Resumo – É crescente o número de novos investidores na bolsa de valores brasileira. Muitos desses investidores buscam uma rentabilidade maior, muitas vezes sem saber analisar as oportunidades e os perigos. Este estudo visa ao desenvolvimento de um modelo de LSTM com o objetivo de prever os preços de ações e ETF na bolsa de valores B3. O modelo foi treinado com dados históricos dos últimos 10 anos de um grupo de cinco ações (Banco do Brasil, Itaú, Vale, Petrobras, Caixa Seguridade) e de dois ETF (BOVA11 e FIND11). Os resultados demonstraram que o modelo é eficaz na previsão dos ativos dentro do mercado financeiro brasileiro, validados por meio de algumas métricas, com os resultados dentro das seguintes faixas: RMSE (0,30 a 1,98), MAE (0,23 a 1,54), MAPE (0,98 a 3,40) e R^2 (0,77 a 0,99). Entretanto, ativos com pouco histórico de dados, como a Caixa Seguridade, apresentaram maior variação nas previsões, indicando limitações do modelo.




Palavras-chave: Aprendizagem de máquina. Bolsa de valores. LSTM. Séries temporais.




Predicting stock prices and ETF on the B3 stock exchange using machine learning techniques

Abstract – The number of new investors in the Brazilian stock market is increasing. Many of these new investors seek higher returns, often without the necessary skills to analyze the opportunities and dangers. This study aims to develop an LSTM model with the objective of predicting the prices of stocks and ETF on the B3 stock exchange. The model was trained with historical data from the last 10 years of a group of five stocks (Banco do Brasil, Itaú, Vale, Petrobras, and Caixa Seguridade) and two ETF (BOVA11 and FIND11). The results demonstrated that the model is effective in predicting assets within the Brazilian financial market, validated through some metrics, with results within the following ranges: RMSE (0.30 to 1.98), MAE (0.23 to 1.54), MAPE (0.98 to 3.40), and R^2 (0.77 to 0.99). However, assets with little historical data, such as Caixa Seguridade, showed greater variation in the forecasts, indicating limitations of the model.

Keywords: Machine learning. Stock market. LSTM. Time series.

Predicción de precios de acciones y ETF en la bolsa B3 utilizando técnicas de aprendizaje automático

¹ Bacharel em Ciência da Computação pela Universidade de Passo Fundo. Passo Fundo, Rio Grande do Sul, Brasil.  eduardoluizzanoto@gmail.com  <https://orcid.org/0009-0004-9479-9841>  <http://lattes.cnpq.br/1732542745002387>.

² Doutor em Ciência da Computação pela Universidade Federal do Rio Grande do Sul. Professor Titular do Programa de Pós-Graduação em Computação Aplicada da Universidade de Passo Fundo. Passo Fundo, Rio Grande do Sul, Brasil.  holbig@upf.br  <https://orcid.org/0000-0002-3126-344X>  <http://lattes.cnpq.br/5419646313109789>.

Resumen – El número de nuevos inversores en la bolsa brasileña está aumentando. Muchos de esos nuevos inversores buscan mayores rendimientos, a menudo sin las habilidades necesarias para analizar las oportunidades y los peligros. Este estudio tiene como objetivo desarrollar un modelo LSTM con el propósito de predecir los precios de las acciones y ETF en la bolsa de valores B3. El modelo fue entrenado con datos históricos de los últimos 10 años de un grupo de cinco acciones (Banco do Brasil, Itaú, Vale, Petrobras y Caixa Seguridade) y dos ETF (BOVA11 y FIND11). Los resultados demostraron que el modelo es efectivo para predecir activos dentro del mercado financiero brasileño, validado a través de algunas métricas, con resultados dentro de los siguientes rangos: RMSE (0.30 a 1.98), MAE (0.23 a 1.54), MAPE (0.98 a 3.40) y R^2 (0.77 a 0.99). Sin embargo, los activos con pocos datos históricos, como Caixa Seguridade, mostraron una mayor variación en las previsiones, lo que indica limitaciones del modelo.

Palabras clave: Aprendizaje automático. Mercado de valores. LSTM. Series temporales.

Introdução

De acordo com um estudo realizado pela Bolsa de Valores Brasileira, conhecida como B3 (2024), em 2023, o número de investidores na bolsa de valores cresceu 23 % entre pessoas físicas e jurídicas. Dessa forma, observa-se que a população brasileira está demonstrando mais interesse em investir seu dinheiro. Ainda conforme esse estudo, nota-se que a quantidade de contas vem aumentando ao longo dos anos juntamente com o número de investidores em renda variável, com exceção do ano de 2023, quando a renda fixa atraiu mais investidores, uma vez que a taxa de juros Selic esteve em alta.

Com o aumento de novos investidores em renda variável, é importante entender como esse mercado se comporta. Através da análise dos indicadores de cada ativo, deve-se conhecer os fatores macroeconômicos que influenciam no preço dos ativos. Além disso, é fundamental analisar a influência da especulação e do sentimento do mercado nos preços. Porém, segundo a própria B3 (2024), há uma carência de conhecimento entre os investidores, pois muitos não fazem análises ou seguem dicas sem embasamento técnico.

Analisar os principais indicadores de ativos antes de comprar é fundamental para uma decisão informada. O Dividend Yield (DY) é uma métrica que contabiliza os pagamentos de dividendos dos últimos 12 meses em relação ao preço atual das ações, refletindo os lucros distribuídos pela empresa aos acionistas. O Preço sobre Valor Patrimonial (P/VP) é utilizado para avaliar se uma empresa está sendo negociada a um preço considerado alto, justo ou baixo em relação ao seu valor patrimonial, indicando se as ações estão supervalorizadas ou subvalorizadas. Já o Preço sobre Lucro (P/L) indica quanto uma empresa está custando em relação ao seu lucro, mostrando quantos anos seriam necessários para o investidor recuperar o dinheiro investido com base nos lucros atuais. A Dívida Líquida também deve ser levada em consideração, pois esse indicador revela o quanto uma empresa está endividada, aspecto crucial para avaliar sua saúde financeira. Por fim, a Margem Líquida deve ser analisada, pois mostra a eficácia da empresa na geração de lucro a partir das vendas; nesse caso, quanto maior a margem, melhor.

Além desses principais indicadores, há fatores macroeconômicos que influenciam o preço dos ativos (Machado; Gartner; Machado, 2018). Destacam-se a taxa de juros, que afeta diretamente o custo de aquisição de financiamentos; a

inflação, que impacta o poder de compra do dinheiro ao longo do tempo; e o crescimento econômico. Quando um país apresenta saúde financeira controlada, as empresas tendem a se valorizar, sendo que o oposto também ocorre em períodos de instabilidade. Além disso, o sentimento de mercado, no qual a especulação dos investidores afeta diretamente o preço dos ativos muitas vezes sem justificativa fundamentada e originada de reações a notícias, exerce grande influência. Enfim, eventos geopolíticos, como guerras ou pandemias, também podem impactar significativamente os preços dos ativos.

Quando se fala em ativos de bolsas de valores, está-se referindo a ações, debêntures, fundos imobiliários, *Exchange-Traded Funds* (ETF), entre outros. Este trabalho irá focar nos ativos ações e ETF. Os ETF, conforme a B3 (2024), são fundos negociados em bolsa que representam uma comunhão de recursos destinados à aplicação em uma carteira de ativos, que buscam retornos que correspondam, de forma geral, à performance, antes de taxas e despesas, de um índice de referência.

Com o aumento de novos investidores e a necessidade de conhecimento ao analisar ativos, aliado ao crescimento no uso de Inteligência Artificial (IA), surge a seguinte questão: é possível prever o preço de ações e ETF acompanhando o desempenho de um índice ou de uma cesta de ativos na bolsa de valores utilizando dados históricos e técnicas de aprendizado de máquina?

Com o intuito de responder essa questão, este trabalho visa a aplicar o modelo de *machine learning* LSTM (*Long Short-Term Memory*) (Da Silva *et al.*, 2021; Hochreiter, Schmidhuber, 1997) para a análise das ações do Banco do Brasil (BBAS3), Itaú (ITUB4), Vale (VALE3), Petrobras (PETR4) e Caixa Seguridade (CXSE3) e dois ETF (BOVA11 e FIND11). LSTM é uma arquitetura de rede neural recorrente (RNN) que possui a capacidade de “lembrar” valores em intervalos arbitrários. Ela destaca-se devido a sua eficácia em lidar com séries temporais cujos intervalos de tempo podem ter duração desconhecida e alta variação. A escolha dessas ações e das ETF deve-se aos fatores do seu grande volume de transações na bolsa e a natureza jurídica distinta entre as empresas, permitindo realizar uma análise mais ampla do mercado de ações e de variáveis diversas que podem afetar o funcionamento desse mercado, como pode ser visto no trabalho de Lopes *et al.* (2021).

De forma complementar, este trabalho visa a destacar a importância da previsão dos preços de ações e ETF na B3 para os investidores que enfrentam um ambiente de mercado volátil e complexo. A utilização de técnicas de IA e de análise de dados de séries temporais pode oferecer aos investidores uma ferramenta efetiva que auxilie na tomada de decisões e na previsão do preço de ativos na bolsa de valores brasileira com uma alta assertividade. Em última análise, o estudo busca contribuir para uma melhor compreensão e aplicação da IA no contexto financeiro, visando a melhorar a precisão das previsões e, consequentemente, proporcionar maior rentabilidade aos investidores. A falta de software e ferramentas acessíveis para essa finalidade reforça a necessidade desta pesquisa e o seu potencial impacto positivo na área financeira.

Revisão de literatura

Devido ao aumento da capacidade de computação e ao dos volumes de dados disponíveis, a IA apresentou um notável crescimento nos últimos anos, utilizando técnicas avançadas de aprendizado de máquina, como *machine learning* e *deep learning* (Goodfellow *et al.*, 2016), que processam volumes massivos de dados, nas mais diversas áreas do conhecimento (Almeida; Magalhães, 2021). Com isso, a IA é capaz de distinguir padrões e aprender com eles, oferecendo informações valiosas e

úteis. A previsão dos preços de ações e ETF é uma tarefa desafiadora devido à natureza complexa e altamente volátil do mercado financeiro. No entanto, o avanço das técnicas de IA, como o aprendizado de máquina, oferece novas perspectivas para lidar com essa questão.

Bishop (2011, p. 606) destaca o potencial das redes neurais artificiais e de outros métodos de aprendizado de máquina na análise de dados financeiros. Como ele aponta, “as redes neurais artificiais oferecem uma abordagem flexível e poderosa para modelar padrões não lineares nos dados, permitindo previsões mais precisas e informadas sobre os movimentos do mercado”. Por meio desses métodos, é possível identificar padrões complexos nos dados históricos do mercado de ações, capturando relações não lineares e potencialmente prevendo tendências futuras com maior precisão.

Como Goodfellow *et al.* (2016, p. 168) observam, “as redes neurais profundas têm o potencial de aprender automaticamente características hierárquicas dos dados, permitindo a representação de informações em diferentes níveis de abstração”. Essa capacidade das redes neurais profundas é relevante na previsão de preços de ações, tema deste trabalho, pois os dados de séries temporais podem conter uma variedade de padrões e relações complexas. Ao utilizar técnicas de *deep learning*, como redes neurais convolucionais (CNN) e RNN, é possível capturar essas nuances e padrões sutis nos dados históricos do mercado de ações, melhorando, assim, a capacidade de prever tendências futuras de preços. Por meio da análise das representações aprendidas pelas redes neurais profundas, é possível identificar indicadores-chave e correlações ocultas nos dados, que podem ser utilizados para tomar decisões de investimento mais estratégicas.

A análise de séries temporais é uma área fundamental da estatística que lida com a modelagem e a previsão de dados que evoluem ao longo do tempo. Conforme argumentam Box *et al.* (2015), as análises de séries temporais são utilizadas para identificar padrões, tendências e estruturas subjacentes, apoiando uma tomada de decisão sólida. Hyndman e Athanasopoulos (2021) referem-se a esse processo como “previsão de séries temporais”. Dessa forma, ao utilizar algoritmos de *machine learning* e *deep learning*, os investidores podem tomar decisões mais embasadas e estratégicas, contribuindo para uma análise mais sofisticada e uma tomada de decisão mais eficaz no momento de compra e venda de ações. Portanto, é de extrema importância prever séries temporais para informar a tomada de decisões estratégicas em vários campos, como finanças e saúde pública.

O estudo de Silveira (2021) explorou a aplicação de modelos preditivos na B3 utilizando dados históricos, como preços diários de abertura, fechamento, máximo, mínimo e volume de negociações, focando especificamente na ação PETR4, devido ao seu alto volume de negociações. Após a etapa de pré-processamento dos dados, foram testados os modelos LSTM e Random Forest. O autor destacou que ambos os modelos são eficazes para a previsão de preços no contexto do mercado brasileiro de ações, mas salienta a escassez de estudos voltados para a B3.

Já Alves e Prado (2022), em seu estudo, fizeram uma comparação de algoritmos de aprendizado de máquina para prever o volume de negociação no mercado financeiro, usando dados históricos da *New York Stock Exchange*. O trabalho avalia três modelos: regressão linear, redes neurais convencionais e RNN – todos implementados em Python com pacotes de aprendizado de máquina e análise de séries temporais. Utilizando a métrica R^2 , os autores obtêm os resultados que, após normalizações e ajustes nos dados, indicam que a regressão linear apresentou desempenho competitivo com RNN em termos de eficiência e precisão preditiva.

De acordo com Santos (2022), os modelos Random Forest e LightGBM aplicados a séries temporais de ativos da bolsa B3 demonstraram eficácia, especialmente ao utilizar dados como preço e volume de negociação. O trabalho adotou duas estratégias de particionamento dos dados (temporal e aleatória) e avaliou métricas como acurácia e precisão com o LightGBM, destacando-se por seu desempenho superior, mesmo sem ajustes de hiperparâmetros. Isso sugeriu um potencial significativo para decisões de investimento mais eficazes no mercado brasileiro, evidenciando as vantagens desses modelos em relação à baixa complexidade de customização e à capacidade de capturar padrões não lineares.

Com o intuito de ajudar os investidores, Nascimento, Santos e Ferreira (2022) desenvolveram o protótipo *The Seer*, uma aplicação web baseada em IA. Utilizando modelos ARIMA, Prophet e LSTM para prever tendências de preços na B3, os autores testaram suas previsões em ativos e ETF com dados históricos, descobrindo que o LSTM alcançou a maior precisão para períodos de até 90 dias, enquanto ARIMA e Prophet se destacaram em previsões de curto prazo.

Através de seu estudo, Lin *et al.* (2021) implementou um modelo de previsão de tendências de ações que combinou gráficos de *candlestick* com um esquema de engenharia de recursos baseado nos padrões “oito trigramas”. Esse sistema utiliza um conjunto de algoritmos de aprendizado de máquina, incluindo Random Forest, GBDT e LSTM, para prever a direção dos preços de fechamento. Com dados de 3445 ações da bolsa chinesa de 2000 a 2017, o modelo demonstrou uma precisão superior a 60 % em determinados padrões de tendência, principalmente nos de alta volatilidade.

Fazendo uma revisão sistemática da literatura sobre o uso de aprendizado de máquina na previsão de ações, Mintarya *et al.* (2023) apontaram que as redes neurais são os modelos mais utilizados, destacando-se pela capacidade de identificar padrões não lineares em dados financeiros. O estudo observou um aumento significativo na adoção do LSTM em previsões financeiras a partir de 2015, enquanto métodos como *Support Vector Machine* (SVM) e *K-Nearest Neighbors* (KNN) são menos recorrentes. A pesquisa ofereceu uma visão abrangente das abordagens atuais, sugerindo que modelos como LSTM estão ganhando força para análise e previsão de tendências do mercado de ações.

Com foco em prever o risco de quedas nos preços de ações, Jiang, Ma e Zhu (2024) se concentraram no mercado de ações chinês. O estudo avaliou 55 características específicas de empresas, utilizando modelos de aprendizado de máquina como redes neurais e Random Forest, e concluiu que essas técnicas são eficazes em capturar a volatilidade, principalmente em empresas estatais e em períodos de baixa incerteza econômica. Os autores destacaram que métricas de lucratividade e valor *versus* crescimento são determinantes cruciais, sugerindo que o aprendizado de máquina oferece uma vantagem significativa para a previsão de quedas de preços ao integrar uma análise complexa de interações financeiras e fatores específicos da empresa.

Ren, Li e Zhang (2024) propuseram um modelo avançado para a previsão de riscos extremos no mercado de ações americano, com foco na S&P 500 entre 2005 e 2022. Para enfrentar a distribuição desbalanceada dos dados e a deriva de conceito, os autores incorporaram pesos de classe e de tempo ao algoritmo AdaBoost e implementaram um framework de aprendizado ativo com anotação automatizada. Os testes empíricos demonstraram que essa abordagem melhorou substancialmente a precisão na classificação de eventos de risco extremo, oferecendo uma ferramenta

valiosa para a formulação de políticas macroeconômicas e o suporte à estabilidade financeira.

Materiais e métodos

Para o desenvolvimento do projeto desta pesquisa, foi utilizada a linguagem Python (Python Software Foundation, 2024), devido à sua ampla utilização em *machine learning*, *deep learning* e manipulação de dados. O ambiente de desenvolvimento utilizado foi o Google Colab (Google, 2024), por ser possível aproveitar os recursos de *Graphics Processing Unit* (GPU) para acelerar o treinamento do modelo. Diversas bibliotecas foram integradas ao projeto: TensorFlow/Keras (TensorFlow, 2024), utilizadas para construção e treinamento do modelo LSTM, fundamental para as tarefas de previsão; Yahoo Finance (yfinance), que consiste em uma *Application Programming Interface* (API) para obtenção dos dados históricos de bolsas de valores utilizados para treinamento e teste do modelo (Yfinance, 2024); Pandas, empregado na manipulação e pré-processamento dos dados (Pandas Development Team 2024); Matplotlib, que proporcionou a visualização gráfica dos resultados, facilitando a interpretação dos dados (Matplotlib, 2024); scikit-learn, utilizado para a normalização dos dados e no cálculo de métricas de desempenho, permitindo a avaliação de desempenho do modelo (Scikit-learn, 2024). Por fim, o Neptune.ai, que é um rastreador de experimentos para treinamento de modelos de base, utilizado para o monitoramento do treinamento e armazenamento de resultados, permitindo um acompanhamento em tempo real completo e detalhado (Neptune.ai, 2024).

Coletas dos dados

Neste projeto, foram utilizados dados históricos diários de ações e ETF da B3, obtidos do yfinance. Os ativos específicos foram: BBAS3, ITUB4, BOVA11, FIND11, VALE3, PETR4 e CXSE3. O período de dados utilizado para o modelo é de 10 anos, contando a partir da data em que o modelo é executado. Essa escolha de período é baseada em estudos anteriores que demonstraram bons resultados com esse intervalo (Ayyildiz; Iskenderoglu, 2024). Caso o ativo não possua 10 anos de histórico disponível, o código ajusta a data de início para a primeira data disponível nos registros históricos, evitando-se, dessa forma, a inclusão de dados inexistentes, garantindo a veracidade e a integridade dos dados temporais utilizados. Os dados diários coletados incluem preço de abertura, máxima, mínima, fechamento, fechamento ajustado e volume de negociações, obtidos utilizando a biblioteca yfinance. Esses dados são fundamentais para a análise temporal e para a construção do modelo de previsão, permitindo ao modelo aprender e capturar padrões históricos e tendências que influenciam o comportamento futuro dos preços das ações.

Preparação dos dados

Foi utilizado o preço de fechamento ajustado, pois ele ajusta os preços de acordo com eventos societários promovidos pelas companhias, como desdobramentos, grupamentos, bonificações, subscrições e pagamento de dividendos, os quais impactam diretamente em reajustes nos preços das ações. Como o yfinance fornece apenas o preço de fechamento ajustado, foi necessário ajustar também os preços de abertura, máxima e mínima para refletir esses eventos.

Para realizar esse ajuste, com base no estudo de Santana (2022), primeiro, determina-se o fator de ajuste. O fator de ajuste de ações é um coeficiente usado para calcular o preço ajustado de um ativo, que reflete com precisão o retorno total do

investidor ao longo do tempo, incorporando eventos como o pagamento de proventos e eventos corporativos como desdobramentos ou grupamentos. Esse fator é obtido conforme a equação (1),

$$fator\ de\ ajuste = \frac{valor\ de\ fechamento\ na\ data - valor\ do\ evento\ por\ ação}{valor\ de\ fechamento\ na\ data} \quad (1)$$

onde o valor de fechamento menos o valor do evento por ação é o valor de fechamento ajustado. Esse fator representa a proporção pela qual os preços devem ser ajustados para refletir os eventos corporativos. Em seguida, aplica-se o fator de ajuste para os preços de abertura, máxima e mínima, ajustando todos os preços históricos de forma consistente. Essa uniformização dos dados é importante para o modelo. Após o ajuste, os preços originais foram substituídos pelos preços ajustados, garantindo que todas as análises subsequentes fossem baseadas em dados consistentes e coerentes com a atualidade.

Entretanto, além dos ajustes nos preços, também foi necessário tratar os dados faltantes, seja em dias que a bolsa não abre, como em finais de semana e feriados, seja por problemas técnicos. Para não causar ruídos nos dados, para os casos em que os preços de abertura, fechamento ajustado, máxima e mínima não tinham valores, foi utilizada a interpolação linear, a qual estima os valores ausentes com base nos pontos anteriores e posteriores conhecidos. Esse método preserva a tendência dos dados e minimiza a introdução de vieses. Já os dados em que o volume de negociações era zero foram substituídos pelo último valor conhecido, utilizando o método de preenchimento *forward fill*, assegurando, com isso, a continuidade dos dados, evitando falsas avaliações do modelo.

A média móvel exponencial (do inglês *exponential moving average* – EMA) também foi incorporada neste trabalho (Ayyildiz; Iskenderoglu, 2024; Chen, 2025). EMA é um indicador técnico que suaviza as flutuações de curto prazo nos preços e destaca as tendências de mercado. Diferentemente da média móvel simples (do inglês *simple moving average* – SMA), a EMA atribui pesos maiores aos preços mais recentes, o que a torna mais sensível às mudanças recentes no mercado. Para todas as ações avaliadas por este trabalho, a EMA foi calculada com um período de 60 dias, período comum para identificar tendências de médio prazo, equilibrando a sensibilidade a novas informações com a estabilidade necessária para evitar ruídos causados pela volatilidade de curto prazo.

A inclusão da EMA como uma coluna adicional no conjunto de dados visa a enriquecer as informações fornecidas ao modelo LSTM. Ao adicionar esse indicador, o modelo tem acesso a uma representação das tendências subjacentes nos preços das ações, melhorando consideravelmente sua capacidade de prever movimentos futuros, segundo estudos recentes de Ayyildiz e Iskenderoglu (2024). Dessa forma, a EMA é crucial para a precisão das previsões financeiras.

Pré-processamento dos dados

Na implementação deste trabalho, a normalização visa a padronizar as variáveis de entrada, garantindo que todas estejam na mesma escala e contribuam de forma equilibrada durante o treinamento do modelo LSTM. As variáveis de entrada apresentam escalas e unidades distintas. Por exemplo, os preços das ações ou ETF estão geralmente na ordem de dezenas ou centenas de reais, enquanto o volume de negociações pode atingir milhões. Essa disparidade pode levar o modelo a atribuir maior peso às variáveis com magnitudes maiores, influenciando negativamente o processo de aprendizagem.

Baseado no trabalho de Jiang, Ma e Zhu (2024), optou-se pela técnica de padronização ou normalização Z-score utilizando a classe `StandardScaler` da biblioteca `scikit-learn`. A padronização transforma os dados para que cada variável tenha média zero e desvio padrão um. Para evitar o vazamento de dados (*data leakage*), os parâmetros de média e desvio padrão foram calculados exclusivamente no conjunto de treinamento. Isso assegura que o modelo não seja influenciado por informações do conjunto de testes durante o aprendizado.

Na sequência, cada variável (coluna) do conjunto de dados foi normalizada individualmente, incluindo preços ajustados de abertura, máxima, mínima, fechamento, volume de negociações e indicadores técnicos como a EMA. Isso garante que as características específicas de cada variável sejam mantidas, ao mesmo tempo que todas contribuem de maneira equilibrada. É importante que os parâmetros de normalização obtidos do conjunto de treinamento sejam aplicados ao conjunto de teste, para que, dessa maneira, a consistência seja mantida, a fim de que o modelo possa realizar previsões precisas em dados não vistos anteriormente. Os escaladores usados na normalização foram armazenados, facilitando a inversão da transformação (desnormalização) dos valores previstos de volta à escala original para interpretação.

Portanto, a normalização contribui para um processo de treinamento mais estável, evitando problemas como gradientes explosivos ou desaparecidos, comuns em redes neurais profundas com dados não normalizados, proporcionando uma convergência mais rápida do modelo devido às escalas equilibradas das *features*, necessitando de menos épocas para alcançar um desempenho satisfatório e, consequentemente, um melhor desempenho preditivo, evidenciando métricas aprimoradas com menor RMSE (*root mean squared error*) e MAE (*mean absolute error*).

É importante destacar que a divisão dos dados em conjuntos de treinamento e teste é uma etapa essencial para avaliar a capacidade de generalização do modelo preditivo. Além disso, de acordo com Box *et al.* (2015), em séries temporais, é fundamental manter a sequência temporal dos dados. Baseado nisso, este trabalho manteve a sequência temporal dos dados, que foram divididos em 80 % para treinamento e 20 % para teste. Destes, o conjunto de treinamento compreende os primeiros 80 % dos dados cronologicamente ordenados, e o conjunto de testes é constituído pelos últimos 20 % dos dados, representando os períodos mais recentes. Dessa forma, o conjunto de treinamento não utilizou os dados utilizados nos testes. Essa divisão foi realizada com base nos estudos de Ayyildiz e Iskenderoglu (2024), bem como no de Najem *et al.* (2024), que utilizaram abordagens semelhantes para garantir a integridade temporal e a representatividade dos dados mais recentes para a validação, refletindo o cenário real de previsão no mercado financeiro.

Além disso, foi definida uma janela de observação (*window size*), que é um parâmetro essencial no pré-processamento dos dados para modelos de séries temporais, especialmente redes neurais como LSTM. Esse parâmetro determina o número de períodos anteriores considerados pelo modelo para prever valores futuros, permitindo capturar padrões e tendências temporais nos dados. Ao definir esse parâmetro com um número fixo, o modelo é capaz de aprender tendências e comportamentos recorrentes que influenciam as variações futuras dos preços das ações, sem sobrecarregá-lo com dados excessivos ou irrelevantes.

A utilização da janela de observação, assim como seu respectivo valor, foi baseada em estudos anteriores, como os de Gülmez (2023) e de Li (2024), além de testes realizados durante o desenvolvimento do modelo. Um período de 50 dias abrange aproximadamente dois meses de negociações, permitindo capturar

movimentos e tendências de curto a médio prazo no mercado financeiro, além de manter um bom equilíbrio entre desempenho preditivo e eficiência computacional. Entretanto, esse valor pode ser alterado, visando a obter melhores resultados. É importante ressaltar que o *window size* é diferente do EMA *span*, que foi definido em 60 dias para o cálculo da EMA. Enquanto o *window size* determina a quantidade de dados históricos usados pelo modelo LSTM, o EMA *span* estabelece o período sobre o qual a EMA é calculada, servindo como uma *feature* adicional que representa tendências de mercado.

Arquitetura do modelo

Diante dos pontos e trabalhos relacionados descritos anteriormente, o modelo escolhido e construído neste trabalho é uma rede neural recorrente do LSTM, implementada utilizando a biblioteca TensorFlow/Keras para a previsão dos preços de ações e ETF na B3. A escolha pela arquitetura LSTM deve-se à sua comprovada capacidade de capturar tendências e padrões em séries temporais, características essenciais para a previsão de preços de ativos financeiros, conforme pode ser visto nos trabalhos de Gupta, Nachappa e Parmanandham (2025) e de Kundu e Pinsky (2025). Além disso, na revisão sistemática de literatura elaborada por Mintarya *et al.* (2023), é demonstrada uma tendência no uso maior de LSTM em relação a outros modelos de *machine learning*, como, por exemplo, árvores de decisão, Random Forest, KNN, naive bayes, regressão linear e outros modelos de redes neurais, os quais seus resultados são apresentados e comparados com os obtidos pelo modelo LSTM (Gupta, Nachappa, Parmanandham, 2025; Gülmez, 2023; Kundu, Pinsky, 2025; Nascimento, Santos, Ferreira, 2022), que obteve o melhor resultado, de acordo com as métricas utilizadas.

O modelo é estruturado em várias camadas. A camada de entrada recebe sequências de dados com dimensões correspondentes ao tamanho da janela de observação e ao número de *features*. O *window size* representa o número de períodos anteriores considerados pelo modelo, enquanto o número de *features* inclui as variáveis de entrada utilizadas, como preços ajustados de abertura, máxima, mínima, fechamento, volume e indicadores técnicos como a EMA.

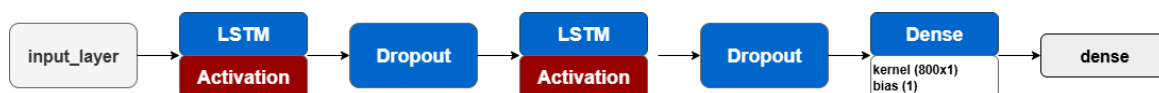
Em seguida, o modelo incorpora duas camadas LSTM e duas camadas *dropout* com taxa de 30 %. A primeira camada LSTM possui 500 unidades e está configurada com *return_sequences=True*, permitindo que a saída seja uma sequência para cada passo de tempo. Isso é fundamental para que a próxima camada LSTM receba a sequência completa de informações temporais. Após essa camada, uma camada *Dropout* é aplicada para reduzir o *overfitting*, desativando aleatoriamente uma fração das unidades durante o treinamento e melhorando a generalização do modelo.

A segunda camada LSTM também conta com 500 unidades, mas com *return_sequences=False*, resultando em uma única saída, que representa o estado final da sequência. Segue-se outra camada *dropout* com taxa de 30 %, que auxilia na síntese das informações acumuladas e na captura dos padrões temporais significativos. Por fim, a camada de saída é composta por um único neurônio com função de ativação linear, responsável por gerar a previsão final do preço da ação ou ETF para o próximo dia. Na figura 1, é ilustrada essa arquitetura, destacando as camadas mencionadas e suas conexões, do *input* inicial à camada de previsão final.

É importante destacar que o número de unidades nas camadas LSTM, definido como 500 para todos os ativos deste estudo, não precisa ser fixo para todos os ativos, podendo ser ajustado para otimizar o desempenho do modelo. A quantidade de unidades é um hiperparâmetro que influencia diretamente a capacidade do modelo de

aprender padrões complexos nos dados. Valores maiores podem aumentar a capacidade de aprendizado, mas também podem levar a um tempo de treinamento mais longo e a um possível *overfitting*. Portanto, a escolha do número de unidades deve equilibrar a complexidade do modelo e sua capacidade de generalização, podendo ser alterada conforme necessário para obter melhores resultados.

Figura 1 - Arquitetura do modelo LSTM utilizado para a previsão de preços de ações e ETF na B3, composto por uma camada de entrada, duas camadas LSTM intercaladas com *dropout* para evitar *overfitting*, e uma camada densa para a previsão final



Fonte: elaborado pelos autores (2025)

Treinamento do modelo

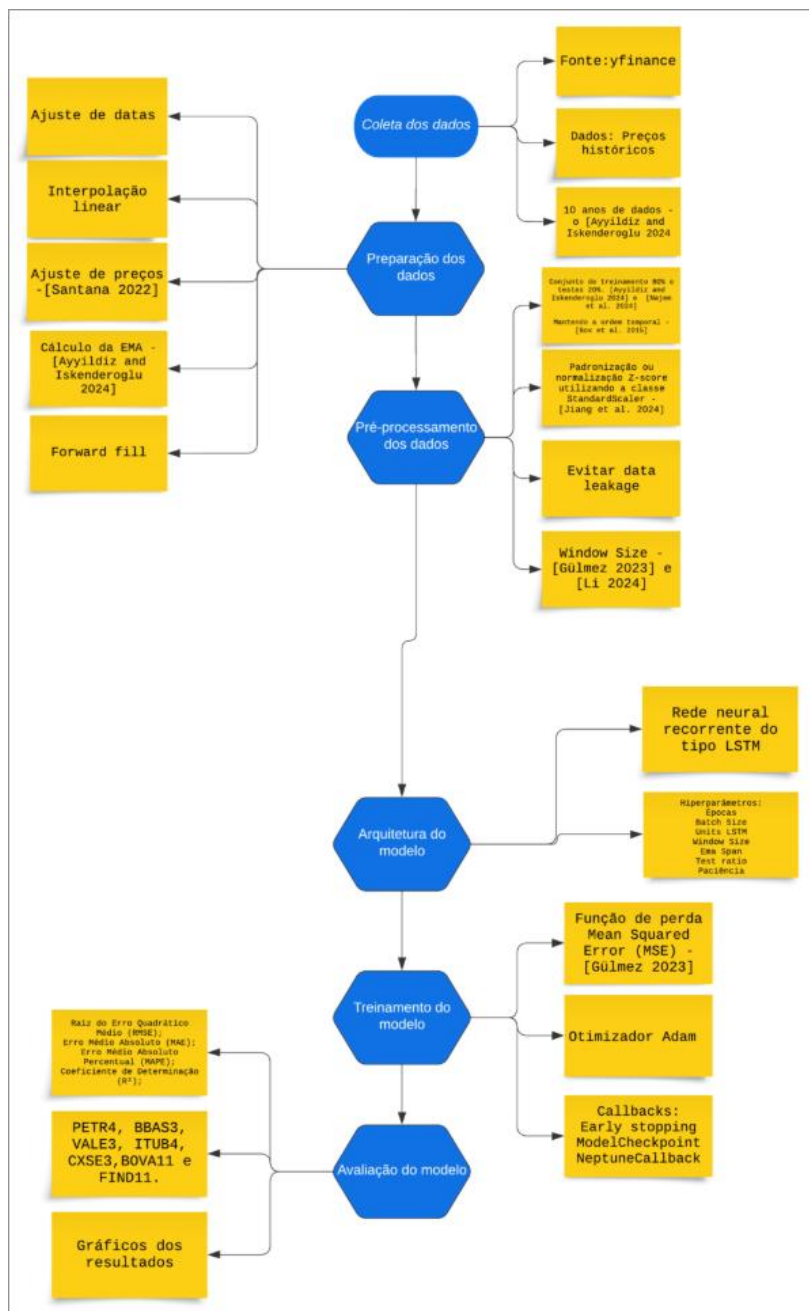
Durante o treinamento do modelo LSTM desenvolvido neste trabalho, a configuração adequada dos hiperparâmetros, a implementação de técnicas de regularização e o monitoramento do desempenho através do Neptune.ai foram importantes para chegar a um resultado satisfatório. Inicialmente, o modelo foi compilado utilizando a função de perda *mean squared error* (MSE), que, de acordo com o estudo de Gülmez (2023), é apropriada para problemas de regressão por penalizar grandes desvios entre os valores previstos e os reais. O otimizador escolhido foi o Adam (*Adaptive Moment Estimation*), reconhecido por sua eficiência computacional e por ajustar dinamicamente as taxas de aprendizado durante o treinamento.

Os hiperparâmetros necessários para o funcionamento do modelo são número de épocas, tamanho do batch, unidades LSTM, *window size*, *EMA span*, *test ratio* e paciência do *early stopping*. Conforme explicação anterior, o conjunto de dados foi dividido em 80 % treino e 20 % teste, preservando as dependências sequenciais. Além disso, existem as sequências de entrada que foram preparadas utilizando *window size* variável de acordo com o melhor resultado, permitindo que o modelo considere as sequências temporais dentro desse período para prever o preço do dia seguinte.

O *early stopping* foi implementado através do *callback* do TensorFlow, em que o objetivo principal é monitorar o treinamento e interrompê-lo caso a perda do conjunto de validação não apresente melhora após 10 épocas consecutivas. Com essa técnica, pode-se prevenir que o modelo se ajuste demais aos dados e evitar o tempo de treinamento desnecessário quando o modelo não apresente melhorias. Também, foi utilizado o *callback* ModelCheckpoint para salvar automaticamente os pesos do modelo sempre que uma melhoria na perda de validação fosse observada, assegurando, assim, que a melhor versão do modelo fosse utilizada em previsões futuras.

A integração do Neptune.ai, que é configurada através de um *token* gerado por ele e passado no *script* do modelo LSTM, permite o acompanhamento em tempo real das métricas de desempenho durante treinamento e teste, facilitando o registro dos hiperparâmetros, a visualização de gráficos e do *dataset* utilizado e a taxa de aprendizado, além de manter um histórico detalhado dos experimentos realizados.

Figura 2 - Fluxograma que resume todas as etapas do trabalho, mostrando a sequência lógica da coleta e preparação dos dados, pré-processamento, arquitetura do modelo, treinamento e avaliação



Fonte: elaborado pelos autores (2025)

Por fim, foi utilizado o método *fit* do TensorFlow, sem embaralhamento dos dados, para manter a ordem temporal. Com o uso do *early stopping*, permitiu-se que o modelo convergisse antes das 50 épocas, evitando o *overfitting*. Após o treinamento, o modelo foi avaliado no conjunto de teste utilizando as métricas RMSE, MAE, MAPE e R^2 , o que proporcionou uma análise abrangente de seu desempenho. Para facilitar a visualização, foi gerado um gráfico, em que é apresentada a comparação dos valores reais e previstos, incluindo a previsão para o próximo dia, permitindo uma avaliação visual da eficácia do modelo. Com todas as etapas descritas, na figura 2, é sintetizado o fluxo completo do projeto, desde a coleta dos dados até a avaliação dos resultados, auxiliando na visualização geral do processo.

Resultados e discussões

O modelo LSTM desenvolvido neste trabalho teve por objetivo fazer a previsão do preço de ações e ETF para um dia a partir de sua execução, buscando prever o preço do próximo dia com base em dados passados. O modelo foi implementado deixando a livre escolha ao usuário de digitar o *ticker* do ativo de que deseja realizar a previsão; dessa forma, o modelo se torna uma ferramenta versátil para análises preditivas. Entretanto, para a validação do modelo, foram escolhidas as ações PETR4, BBAS3, VALE3, ITUB4, CXSE3 e as ETF BOVA11 e FIND11. Essa escolha permitiu avaliar a previsão do modelo entre diferentes ativos e setores, de modo que fosse possível entender como o modelo se comporta com empresas ou ETF mais recentes ou que sofrem forte impacto de fatores externos.

A avaliação do desempenho do modelo foi realizada utilizando métricas amplamente reconhecidas para a previsão de séries temporais, como erro quadrático médio (MSE), que tira a média dos quadrados das diferenças entre os valores previstos e os reais, penalizando grandes desvios; raiz do erro quadrático médio (RMSE), que é a raiz quadrada do MSE e traz o erro para a mesma unidade dos dados originais, facilitando a interpretação; erro médio absoluto (MAE), que calcula a média dos valores absolutos das diferenças entre as previsões e os valores reais, fornecendo uma medida clara da precisão média sem penalizar excessivamente grandes erros; erro médio absoluto percentual (MAPE), que representa o erro médio absoluto em termos percentuais em relação aos valores reais, permitindo avaliar a precisão relativa das previsões entre ativos com diferentes escalas de preço; e, por fim, o coeficiente de determinação (R^2), que indica a proporção da variância dos dados reais, sendo um indicador da qualidade do ajuste, em que valores próximos de 1 (um) sugerem que o modelo explica bem a variabilidade dos dados. Essas métricas proporcionaram um entendimento detalhado do desempenho do modelo sob diferentes perspectivas, permitindo identificar não apenas a magnitude média dos erros, mas também a eficácia do modelo em capturar a variabilidade dos dados e a precisão relativa das previsões. Os resultados das métricas de avaliação são apresentados na tabela 1.

Tabela 1 - Resultados de métricas dos ativos referentes a avaliação do modelo implementado

Ativo	RMSE	MAE	MAPE (%)	R^2
PETR4	1,3247	1,0350	3,4042	0,9692
BBAS3	0,5293	0,4360	1,8372	0,9752
VALE3	1,3175	1,0259	1,6758	0,8564
ITUB4	0,4073	0,3246	1,1167	0,9917
CXSE3	0,3077	0,2347	1,6045	0,7713
BOVA11	1,4713	1,1625	0,9839	0,9738
FIND11	1,9832	1,5468	1,3176	0,9705

Fonte: elaborado pelos autores (2025)

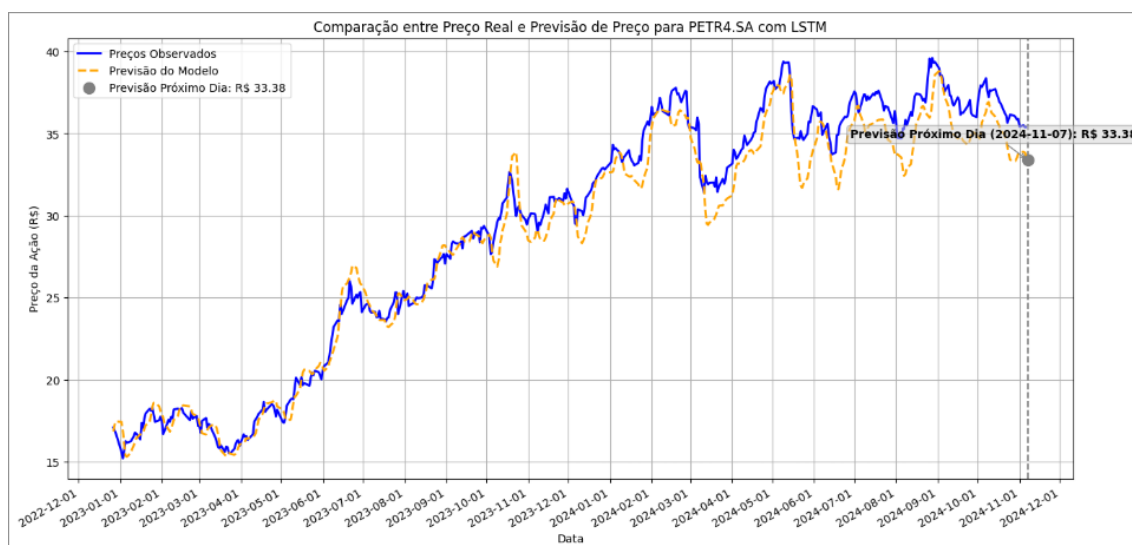
A fim de complementar a análise quantitativa, foram gerados gráficos comparando os valores reais e previstos para cada ativo. Nesses gráficos, é possível visualizar o desempenho do modelo ao longo do tempo, identificando padrões ou discrepâncias nas previsões, além de visualizar a previsão para o próximo dia a partir da data de execução. Nos gráficos apresentados, o eixo X (Data) representa o tempo, com as datas dos preços de fechamento ajustados de cada ação ou ETF ao longo do período analisado. O eixo Y (Preço da Ação) indica o preço de fechamento ajustado em reais (R\$). A linha azul do gráfico reflete os preços reais observados, enquanto a linha laranja representa as previsões do modelo LSTM para o ativo analisado. Por fim,

o ponto em cinza mostra a previsão feita pelo modelo para o próximo dia de negociação, que não é conhecido e não possui dados.

Para o ativo PETR4 (Petrobras S.A.), analisado no período de 10 de novembro de 2014 a 6 de novembro de 2024, o modelo demonstrou boa precisão, com um RMSE de 1,3247 e MAE de 1,035, refletindo erros médios baixos nas previsões. O MAPE de 3,4042 % confirma a proximidade entre os valores previstos e os reais, enquanto o R^2 de 0,9692 indica que o modelo capturou 96,92 % da variabilidade dos preços.

Conforme pode ser visto na figura 3, as previsões acompanharam de perto a trajetória dos preços reais, apresentando pequenas discrepâncias em períodos de maior volatilidade. Tais divergências são esperadas, pois o modelo não incorpora fatores de especulação ou sentimento de mercado, especialmente relevantes em empresas de commodities. A previsão para o próximo dia a partir da execução, 7 de novembro de 2024, foi de R\$ 33,38, com base nas tendências recentes. No entanto, o preço de fechamento real nesse dia foi de R\$ 35,51, resultando em uma discrepância entre o valor previsto e o valor efetivo. Esse desvio pode ser atribuído à influência de fatores não considerados pelo modelo, tais como especulação, eventos inesperados ou variações de sentimento no mercado.

Figura 3 - Comparação entre os valores reais e preditos das ações da Petrobras (PETR4) no período de 1º de dezembro de 2022 a 6 de junho de 2024, com previsão para o dia 7 de novembro de 2024



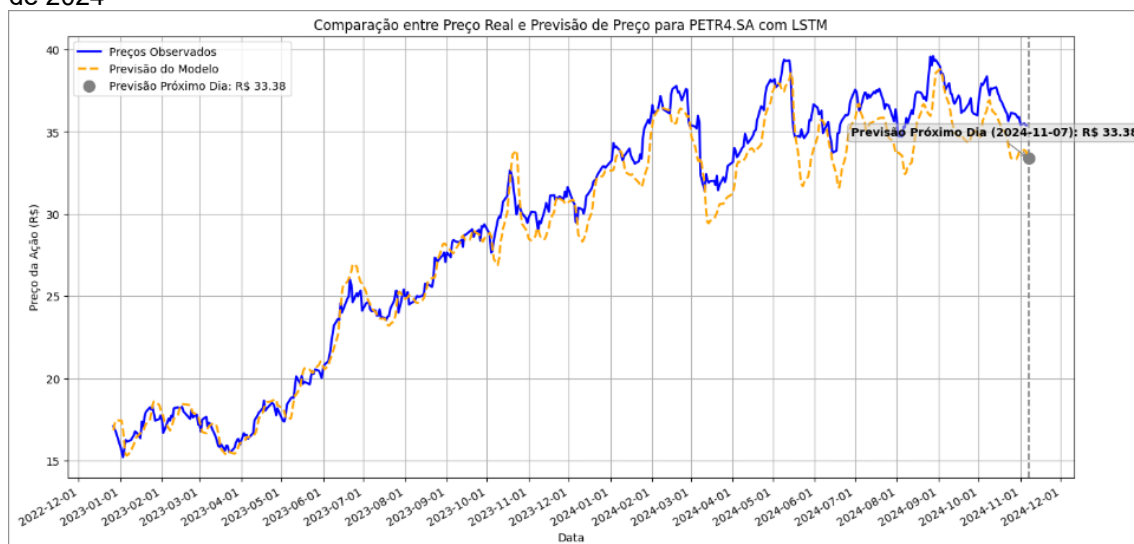
Fonte: elaborado pelos autores (2025)

Para o ativo BBAS3 (Banco do Brasil S.A.), analisado no período de 10 de novembro de 2014 a 6 de novembro de 2024, o modelo apresentou excelente precisão, com RMSE de 0,5293 e MAE de 0,4360, indicando erros médios muito baixos nas previsões. O MAPE de 1,8372 % evidenciou a significativa proximidade entre os valores previstos e observados, e o R^2 de 0,9752 mostrou que o modelo capturou 97,52 % da variabilidade dos preços da ação, reforçando sua alta acurácia.

No gráfico da figura 4, observa-se que a previsão acompanha de perto os preços reais, com desvios mínimos mesmo em períodos de maior volatilidade. É importante considerar que, sendo o Banco do Brasil uma empresa estatal, seu preço está sujeito a influências políticas e econômicas específicas do governo, o que pode impactar as previsões em cenários de mudanças repentinas no contexto econômico. A previsão para o próximo dia a partir da execução é 7 de novembro de 2024, com

valor previsto de R\$ 26,66, fornecendo uma estimativa atualizada. No entanto, o preço de fechamento real nesse dia foi de R\$ 26,19, resultando em uma pequena diferença em relação ao valor estimado. Essa discrepância pode ser atribuída a fatores de mercado não considerados pelo modelo, reforçando a importância de levar em conta variáveis adicionais ao interpretar as previsões.

Figura 4 - Comparação entre os valores reais e preditos das ações do Banco do Brasil (BBAS3) no período de 1º de dezembro de 2022 a 6 de novembro de 2024, com previsão para o dia 7 de novembro de 2024

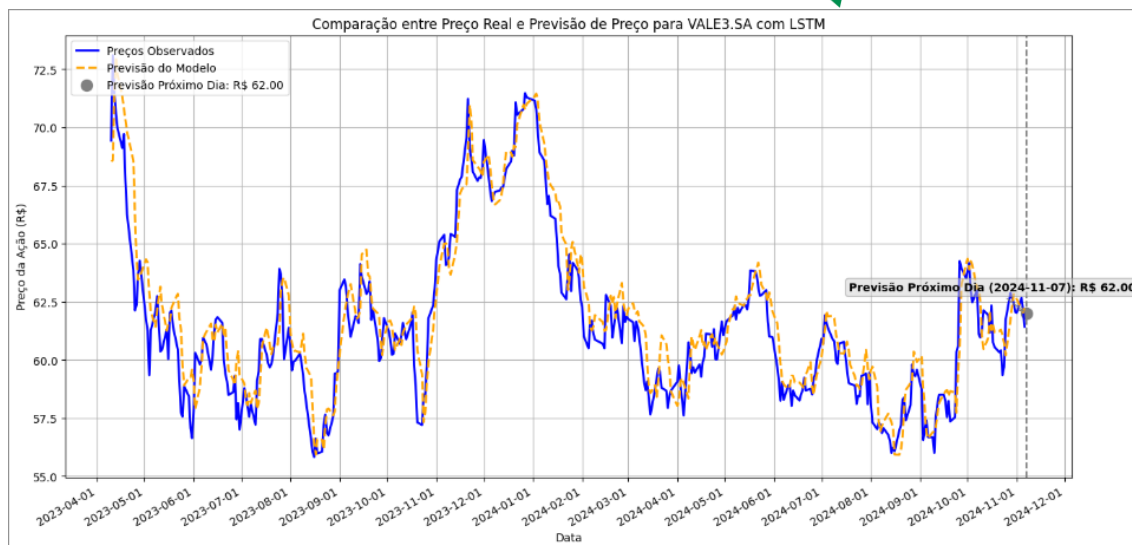


Fonte: elaborado pelos autores (2025)

Para o ativo VALE3 (Vale S.A.), analisado no período de 10 de novembro de 2014 a 6 de novembro de 2024, o modelo apresentou desempenho satisfatório, com RMSE de 1,3175 e MAE de 1,0259, indicando erros médios moderados. O MAPE de 1,6758 % sugere boa proximidade entre os valores previstos e reais, embora o R^2 de 0,8564 revele que o modelo capturou aproximadamente 85,64 % da variabilidade dos preços, evidenciando alguma dificuldade em prever movimentos mais voláteis.

Conforme o gráfico da figura 5, as previsões seguem o padrão geral dos preços reais, mas com desvios mais perceptíveis em períodos de alta oscilação. Isso pode ser atribuído à exposição da empresa a fatores externos, como flutuações no preço do minério de ferro e mudanças na economia global. Sendo a Vale uma empresa de commodities, seu desempenho é fortemente influenciado por eventos internacionais e pela volatilidade das commodities, o que explica algumas das limitações observadas. A previsão para o próximo dia a partir da execução é 7 de novembro de 2024, com valor previsto de R\$ 62,00, refletindo a tendência recente capturada pelo modelo. Entretanto, o preço de fechamento real nesse dia foi de R\$ 63,00, apresentando um desvio de R\$ 1,00 em relação à estimativa. Essa diferença reforça a influência de fatores não modelados, como a sensibilidade às flutuações internacionais do minério de ferro e eventos externos, demonstrando que o modelo, embora útil, não captura todos os elementos que impactam o preço da ação.

Figura 5 - Comparação entre os valores reais e preditos das ações da Vale (VALE3) no período de 1º de abril de 2023 a 6 de novembro de 2024, com previsão para o dia 7 de novembro de 2024



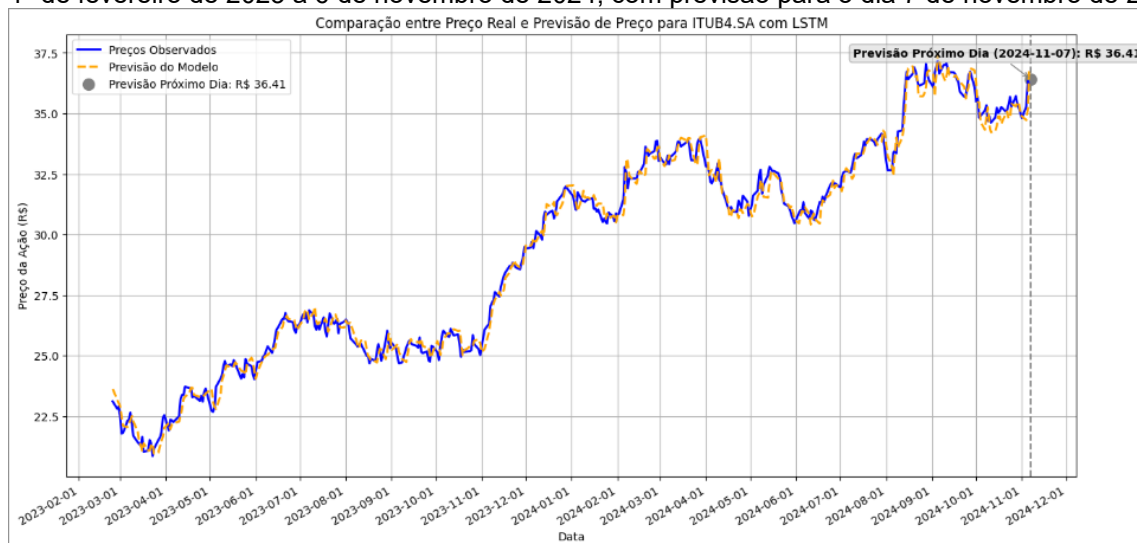
Fonte: elaborado pelos autores (2025)

Para o ativo ITUB4 (Itaú Unibanco S.A.), analisado no período de 10 de novembro de 2014 a 6 de novembro de 2024, o modelo apresentou excelente desempenho, com RMSE de 0,4070 e MAE de 0,3246, indicando erros médios muito baixos. O MAPE de 1,1167 % demonstra alta precisão, e o R^2 de 0,9917 revela que o modelo capturou 99,17 % da variabilidade dos preços reais, evidenciando alta acurácia. No gráfico da figura 6, percebe-se que as previsões acompanham de perto os preços observados, com pequenas discrepâncias em momentos pontuais. Sendo o Itaú o maior banco privado da América Latina, seu desempenho tende a ser influenciado por fatores macroeconômicos internos e externos, mas, diferentemente de empresas estatais, está menos sujeito a interferências governamentais diretas, o que pode contribuir para a estabilidade das previsões. A previsão para o próximo dia a partir da execução é 7 de novembro de 2024, com valor previsto de R\$ 36,41, refletindo a tendência atual capturada pelo modelo. Entretanto, o preço de fechamento real nesse dia foi de R\$ 35,64, indicando uma pequena diferença em relação ao valor estimado. Esse desvio, embora limitado, reforça a ideia de que fatores não contemplados pelo modelo, como alterações nas condições macroeconômicas e mudanças no sentimento de mercado, podem influenciar o preço final do ativo.

Para o ativo CXSE3 (Caixa Seguridade S.A.), analisado no período de 30 de abril de 2021 a 5 de novembro de 2024, o modelo apresentou desempenho mediano, com RMSE de 0,3077 e MAE de 0,2347, refletindo erros médios baixos, porém, o R^2 de 0,7713 indica que capturou apenas 77,13 % da variabilidade dos preços. Isso sugere maior dificuldade em comparação com os outros ativos, possivelmente devido à menor quantidade de dados históricos disponíveis, já que a empresa realizou sua IPO (Oferta Pública Inicial) em 2021. O MAPE de 1,6045 % ainda reflete precisão razoável, mas, conforme o gráfico na figura 7, o modelo teve dificuldades em capturar variações abruptas nos preços, especialmente em períodos de alta volatilidade. Como uma empresa relativamente nova no mercado e sendo estatal, a Caixa Seguridade está sujeita a volatilidade adicional associada à sua fase inicial de crescimento e a influências políticas, fatores que impactam a previsão do modelo. A previsão para o próximo dia a partir da execução é 6 de novembro de 2024, com valor previsto de R\$ 14,54, refletindo a tendência recente identificada. Entretanto, o preço de fechamento real nesse dia foi de R\$ 14,46, apresentando uma pequena diferença em relação ao valor estimado. Essa discreta discrepância ressalta a importância de incorporar variáveis adicionais e estratégias mais robustas, a fim de lidar

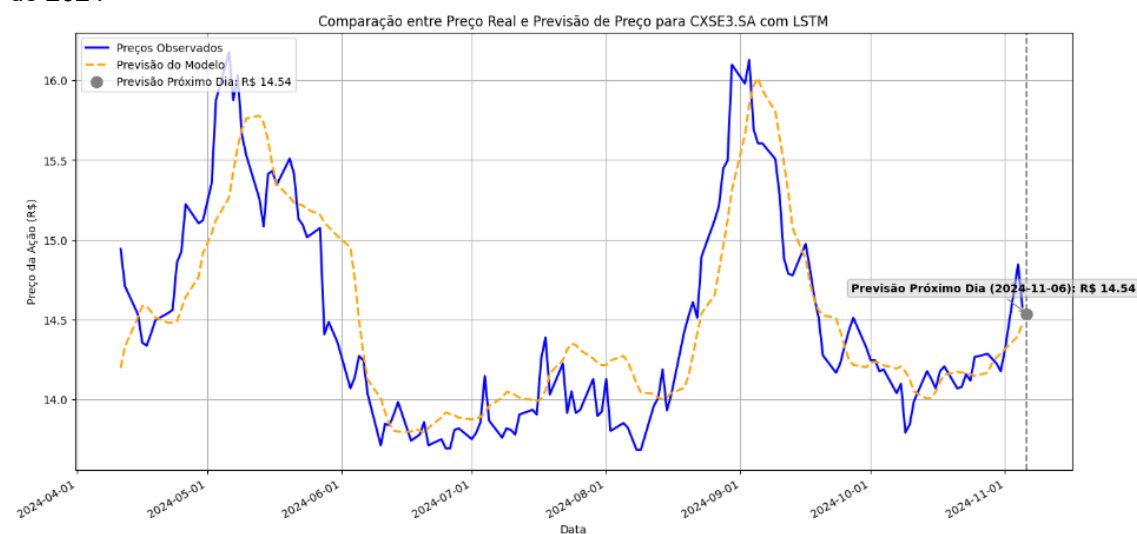
adequadamente com a natureza volátil e as limitações de dados históricas associadas a ativos recém-introduzidos no mercado.

Figura 6 - Comparação entre os valores reais e preditos das ações do Itaú (ITUB4) no período de 1º de fevereiro de 2023 a 6 de novembro de 2024, com previsão para o dia 7 de novembro de 2024



Fonte: elaborado pelos autores (2025)

Figura 7 - Comparação entre os valores reais e preditos das ações da Caixa Seguridade (CXSE3) no período de 30 de abril de 2021 a 5 de novembro de 2024, com previsão para o dia 6 de novembro de 2024

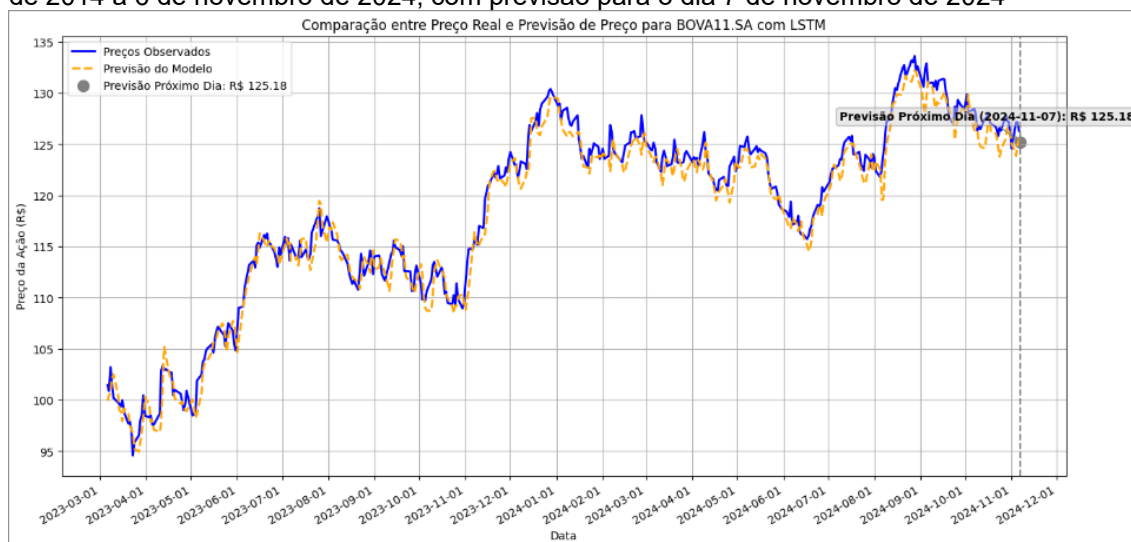


Fonte: elaborado pelos autores (2025)

Para o ETF BOVA11, analisado no período de 10 de novembro de 2014 a 6 de novembro de 2024, o modelo apresentou desempenho satisfatório, com RMSE de 1,4713 e MAE de 1,1625, indicando precisão razoável nas previsões. O MAPE de 0,9839 % mostra grande proximidade entre os valores previstos e reais, enquanto o R^2 de 0,9738 indica que o modelo capturou 97,38 % da variabilidade dos preços, reforçando sua capacidade preditiva. No gráfico da figura 8, observa-se que as previsões seguem de perto os preços reais, com pequenas discrepâncias em períodos de maior volatilidade. Como o BOVA11 é um ETF que replica o desempenho do índice Bovespa, está exposto a uma série de variáveis macroeconômicas e eventos que influenciam o mercado de ações brasileiro como um todo. Em períodos de incerteza econômica, como oscilações nas taxas de juros e mudanças nas políticas

econômicas, é esperado aumento na volatilidade do BOVA11, o que pode impactar a precisão das previsões. A previsão para o próximo dia a partir da execução é 7 de novembro de 2024, com valor previsto de R\$ 125,18, refletindo a tendência recente capturada pelo modelo. Entretanto, o preço de fechamento real nesse dia foi de R\$ 126,00, mostrando uma pequena diferença em relação ao valor estimado, o que reforça a importância de considerar fatores externos não modelados.

Figura 8 - Comparação entre os valores reais e preditos do ETF BOVA11 no período de 10 de novembro de 2014 a 6 de novembro de 2024, com previsão para o dia 7 de novembro de 2024

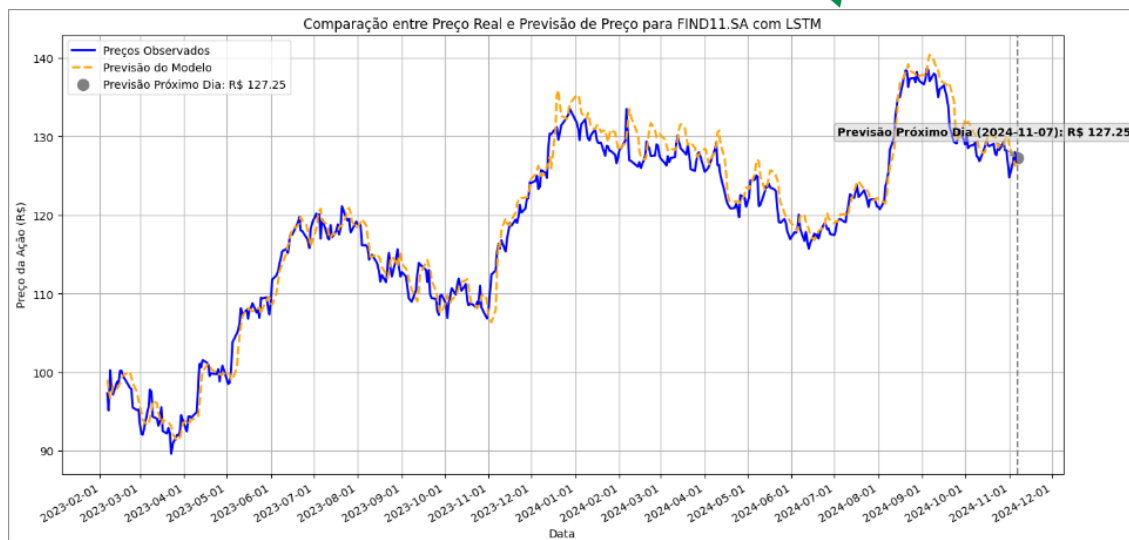


Fonte: elaborado pelos autores (2025)

O ETF FIND11, que replica o IFNC (Índice Financeiro), acompanha o desempenho médio das empresas com maior representatividade no setor financeiro brasileiro, como bancos, fintechs e seguradoras, analisado no período de 10 de novembro de 2014 a 6 de novembro de 2024. O modelo demonstrou desempenho consistente, com RMSE de 1,9832 e MAE de 1,5468, indicando erro médio um pouco mais elevado em comparação com outros ativos. O MAPE de 1,3176 % ainda reflete boa proximidade entre os valores previstos e observados, e o R^2 de 0,9705 indica que o modelo capturou 97,05 % da variabilidade dos preços.

No gráfico da figura 9, percebe-se que as previsões acompanham de perto os preços reais, com pequenas discrepâncias em períodos de maior volatilidade. Por se tratar de um ETF, o FIND11 está exposto a fatores que afetam todo o mercado financeiro brasileiro, incluindo variações macroeconômicas e políticas internas que podem gerar oscilações. Essas influências podem explicar algumas das variações e limitações observadas nas previsões. A previsão para o próximo dia a partir da execução é 7 de novembro de 2024, com valor previsto de R\$ 127,25, refletindo a tendência atual identificada pelo modelo. Entretanto, o preço de fechamento real nesse dia foi de R\$ 126,23, mostrando uma leve diferença em relação ao valor estimado, possivelmente decorrente de eventos de curto prazo ou de incertezas de mercado não capturadas pelo modelo.

Figura 9 - Comparação entre os valores reais e preditos do ETF FIND11 no período de 10 de novembro de 2014 a 6 de novembro de 2024, com previsão para o dia 7 de novembro de 2024



Fonte: elaborado pelos autores (2025)

Comparação dos resultados com trabalhos similares

Visando a demonstrar a eficácia deste estudo, uma comparação com o modelo LSTM de Silveira (2021) para o ativo PETR4 foi realizada, observando-se que o modelo deste estudo apresentou um RMSE de 1,3247, enquanto o modelo de Silveira (2021) alcançou um RMSE de 1,7194. Isso representa uma redução de aproximadamente 22,96 % no erro de previsão, indicando uma melhoria significativa na precisão do modelo. Essa melhoria pode ser atribuída aos ajustes nos hiperparâmetros e às técnicas de pré-processamento aplicadas, como a normalização adequada dos dados e a manutenção da sequência temporal durante a divisão dos conjuntos de treinamento e teste. A definição apropriada da janela de observação permitiu ao modelo capturar padrões e tendências significativas no comportamento do mercado sem sobrecarregá-lo com informações irrelevantes. Portanto, os resultados deste estudo demonstram maior eficácia em relação ao modelo de Silveira (2021), evidenciando que os ajustes metodológicos implementados resultaram em previsões mais precisas. Já o trabalho de Nascimento, Santos e Ferreira (2022) avalia três ativos: PETR4, ITUB4 e BOVA11. Nele, são aplicados os métodos ARIMA, Prophet e LSTM, e a avaliação foi realizada pelas métricas MAPE e RMSE. O modelo LSTM obteve os melhores resultados para todos os ativos, com o MAPE variando de 1,69 a 2,75 e o RMSE de 0,54 a 4,07. Já o trabalho desta pesquisa, na análise dos resultados de seus 7 ativos, obteve um RMSE variando de 0,30 a 1,98, MAE de 0,23 a 1,54 e MAPE de 0,98 a 3,40, o que demonstra a qualidade dos resultados obtidos e sua aplicabilidade ao contexto da Bolsa brasileira.

Além do resultado numérico, a escolha dos trabalhos de Silveira (2021) e de Nascimento, Santos e Ferreira (2022) para comparação justifica-se pela similaridade de contexto e de abordagem. Assim como neste estudo, ambos os trabalhos focaram na B3, empregando dados históricos de ativos e validando o modelo sobre os ativos nacionais. Ademais, ambas as pesquisas adotaram métricas semelhantes, como MAPA, RMSE e MSE, e utilizaram arquiteturas baseadas em RNN do tipo LSTM. Essa equivalência de escopo, dados, métricas e técnicas confere relevância à comparação, pois garante que as melhorias observadas não sejam fruto de condições contextuais diferentes, mas, sim, decorram dos aperfeiçoamentos metodológicos implementados. Dessa forma, a avaliação comparativa reforça a validade dos resultados alcançados neste estudo.

Considerações finais

Neste estudo, implementou-se um modelo LSTM para prever os preços de ações e ETF na B3, utilizando dados históricos de preços ajustados e a média móvel exponencial (EMA) como característica adicional para capturar padrões temporais. Os resultados demonstram que o modelo é eficaz na previsão dos ativos dentro do mercado financeiro brasileiro, com métricas como RMSE, MAE, MAPE e R^2 indicando alta precisão. Entretanto, ativos com pouco histórico de dados, como a Caixa Seguridade, apresentaram maior variação nas previsões, revelando limitações do modelo. Pode-se concluir que o modelo se mostrou promissor, aplicando boas técnicas de processamento dos dados e configuração dos hiperparâmetros para a análise de tendências de preços, embora ainda existam oportunidades de aprimoramento para torná-lo mais adaptável às complexidades do mercado financeiro brasileiro. Assim, este trabalho contribui para o avanço das técnicas de previsão financeira no Brasil, indicando caminhos promissores para pesquisas futuras. Este trabalho destaca-se por focar sua análise em ativos da bolsa de valores brasileira, avaliando em sua análise os aspectos econômicos e as incertezas que regem o mercado de ações do Brasil. Esse é ponto positivo do trabalho em relação àqueles que focam o mercado de ações internacionais e acabam sendo específicos a situações de certos países ou regiões.

Para trabalhos futuros, visando à melhoria deste modelo, podem ser adicionadas variáveis para lidar com momentos macroeconômicos e fatores externos que influenciam o mercado, como notícias, especulação e mudanças políticas, ou ainda combinar o LSTM com outras arquiteturas de aprendizado de máquina. Além disso, a implementação de técnicas de otimização automatizada de hiperparâmetros, como *grid search*, *random search* ou otimização bayesiana, pode aprimorar o desempenho do modelo. A incorporação de modelos híbridos, como CNN-LSTM ou Transformers, também pode aumentar a capacidade de capturar padrões complexos nos dados. A implementação dessas melhorias permitirá ao modelo considerar uma gama mais ampla de fatores que influenciam os preços, aumentando significativamente sua precisão.

Referências

- ALMEIDA, Tiago da Silva; MAGALHÃES, Hudson Pena. A precipitation forecasting in central ecotone region in Brazil using artificial neural networks and public climatic data. **Revista Sítio Novo**, Palmas, v. 5, n. 2, p. 131-146, 2021. DOI: 10.47236/2594-7036.2021.v5.i2.131-146p. Disponível em: <https://sitionovo.iftto.edu.br/index.php/sitionovo/article/view/919>. Acesso em: 15 dez. 2025.
- ALVES, P. C. P.; PRADO, S. C. C. Estudo comparativo entre algoritmos de *machine learning* aplicados à previsão de séries temporais do mercado financeiro. In: CONGRESSO DE TECNOLOGIA, Faculdade de Tecnologia de Mococa, 2022, Mococa. **Anais** [...]. Mococa: Fatec, 2022. v. 6, n. 2.
- AYYILDIZ, Nazif; ISKENDEROGU, Omer. How effective is machine learning in stock market predictions? **Heliyon**, v. 10, n. 2, p. e24123, 2024. DOI: 10.1016/j.heliyon.2024.e24123. Disponível em: <https://doi.org/10.1016/j.heliyon.2024.e24123>. Acesso em: 8 out. 2025.

B3. **Perfil de investidores pessoa física**. 2024. Disponível em: https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/. Acesso em: 17 out. 2024.

BISHOP, Christopher M. **Pattern recognition and machine learning**. Berlin: Springer, 2011.

BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C.; LJUNG, Greta M. **Time series analysis forecasting and control**. 5. ed. San Francisco: Wiley, 2015.

CHEN, James. Exponential moving average (EMA): definition, formula, and usage. **Investopedia**, 2025. Disponível em: <https://www.investopedia.com/terms/e/ema.asp>. Acesso em: 14 dez. 2025.

DA SILVA, Jeremias Fontinele; RODRIGUES, Carlos Alberto de Sousa Parente; TOLENTINO, Carlos Henrique Corrêa; MACIEL, Wandro Bequiman. Modelo de rede neural para avaliação desportiva. **Revista Sítio Novo**, Palmas, v. 5, n. 3, p. 34-44, 2021. DOI: 10.47236/2594-7036.2021. Disponível em: <https://sitionovo.iftto.edu.br/index.php/sitionovo/article/view/952>. Acesso em: 15 dez. 2025.

GOODFELLOW, Ian *et al.* **Deep learning**. Cambridge: The MIT Press, 2016. (Adaptive Computation and Machine Learning Series).

GOOGLE. **Google Colaboratory**. 2024. Disponível em: <https://colab.research.google.com/>. Acesso em: 15 maio 2024.

GÜLMEZ, Burak. Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. **Expert Systems with Application**, v. 227, p.120346, 2023. DOI: 10.1016/j.eswa.2023.120346. Disponível em: <https://doi.org/10.1016/j.eswa.2023.120346>. Acesso em: 15 out. 2024.

GUPTA, Samraj; NACHAPPA, Sanchal; PARMANANDHAM, Nirmala. Stock market time series forecasting using comparative machine learning algorithms. **Procedia Computer Science**. v. 252, n. C, p. 893-904, 2025. DOI: 10.1016/j.procs.2025.01.050. Disponível em: <https://doi.org/10.1016/j.procs.2025.01.050>. Acesso em: 14 dez. 2025.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. **Neural Computation**. v. 9, n. 8, p. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735. Disponível em: <https://doi.org/10.1162/neco.1997.9.8.1735>. Acesso em: 15 dez. 2025.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: principles and practice**. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 14 dez. 2025.

JIANG, Fuwei; MA, Tian; ZHU, Feifei. Fundamental characteristics, machine learning, and stock price crash risk. **Journal of Financial Markets**, v. 69, p.100908,

2024. DOI: 10.1016/j.finmar.2024.100908. Disponível em: <https://doi.org/10.1016/j.finmar.2024.100908>. Acesso em: 15 out. 2024.

KUNDU, Triparna; PINSKY, Eugene. Predicting daily stock price directions with deep learning models. **Machine Learning with Applications**, v. 22, p. 100744, 2025. DOI: 10.1016/j.mlwa.2025.100744. Disponível em: <https://doi.org/10.1016/j.mlwa.2025.100744>. Acesso em: 14 dez. 2025.

LI, Katherine Yi. Machine learning for stock price prediction. **Neptune Blog**, 2024. Disponível em: <https://neptune.ai/blog/predicting-stock-prices-using-machine-learning>. Acesso em: 16 out. 2024.

LIN, Yaohu *et al.* Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme. **IEEE Access**, v. 9, p. 101433-101446, 2021.

LOPES, Alex Oliveira *et al.* Factors influencing Petrobras Stock Prices (PETR4) between 2009 and 2020. **Research, Society and Development**, [s. l.], v. 10, n. 7, p. e13410716294, 2021. DOI: 10.33448/rsd-v10i7.16294. Disponível em: <https://rsdjournal.org/rsd/article/view/16294>. Acesso em: 12 dez. 2025.

MACHADO, Michele Rílany Rodrigues; GARTNER, Ivan Ricardo; MACHADO, Lúcio de Souza. Relationship between Ibovespa and macroeconomic variables: evidence from a Markov-Switching model. **Brazilian Review of Finance**, [s. l.], v. 15, n. 3, p. 435-468, 2018. DOI: 10.12660/rbfin.v15n3.2017.59406. Disponível em: <https://periodicos.fgv.br/rbfin/article/view/59406>. Acesso em: 12 dez. 2025.

MATPLOTLIB. **Matplotlib**. 2024. Disponível em: <https://matplotlib.org/>. Acesso em: 20 abr. 2025.

MINTARYA, Latrisha N. *et al.* Machine learning approaches in stock market prediction: a systematic literature review. **Procedia Computer Science**, v. 216, p.96-102, 2023. DOI: 10.1016/j.procs.2022.12.115. Disponível em: <https://doi.org/10.1016/j.procs.2022.12.115>. Acesso em: 20 abr. 2025.

NAJEM, Rihab *et al.* Advancements in Artificial Intelligence and machine learning for stock market prediction: a comprehensive analysis of techniques and case studies. **Procedia Computer Science**, v. 231, p.198-204, 2024. DOI: 10.1016/j.procs.2023.12.193. Disponível em: <https://doi.org/10.1016/j.procs.2023.12.193>. Acesso em: 20 abr. 2025.

NASCIMENTO, Oberis S.; SANTOS, Felipe G.; FERREIRA, Karl Hansimuller A. Previsão de preços de ações utilizando Inteligência Artificial. *In*: BRAZILIAN WORKSHOP ON ARTIFICIAL INTELLIGENCE IN FINANCE (BWAIF), 1., 2022, Niterói, RJ. **Anais eletrônicos** [...]. Porto Alegre: SBC, 2022. Disponível em: <https://sol.sbc.org.br/index.php/bwaif/article/view/20485>. Acesso em: 20 abr. 2025.

NEPTUNE.AI. **Neptune documentation**. 2024. Disponível em: <https://neptune.ai/>. Acesso em: 16 out. 2024.

PANDAS DEVELOPMENT TEAM. **Pandas documentation**. 2024. Disponível em: <https://pandas.pydata.org/>. Acesso em: 16 out. 2024.

PYTHON SOFTWARE FOUNDATION. **Python**. 2024. Disponível em: <https://www.python.org/>. Acesso em: 15 out. 2024.

REN, Tingting; LI, Shaofang; ZHANG, Siying. Stock market extreme risk prediction based on machine learning: evidence from the american market. 2024. **The North American Journal of Economics and Finance**, v. 74, p. 102241. DOI: 10.1016/j.najef.2024.102241. Disponível em: <https://doi.org/10.1016/j.najef.2024.102241>. Acesso em: 20 abr. 2025.

SANTANA, Daniel. **Ajustando o preço das ações Yahoo Finance**. 2022. Disponível em: <https://medium.com/@sant1/ajustando-o-preço-das-ações-yahoo-finance-9bd5520bf941>. Acesso em: 16 out. 2024.

SANTOS, Júlio Cesar. **Abordagens de aprendizado de máquina para séries temporais financeiras**. 2022. Trabalho de Conclusão de Curso (MBA em Inteligência Artificial e Big Data) – Universidade de São Paulo, São Paulo, 2022.

SCIKIT-LEARN. **Scikit-learn documentation**. 2024. Disponível em: <https://scikit-learn.org/stable/>. Acesso em: 16 out. 2024.

SILVEIRA, Rafael Bourscheid. **Avaliação de modelos preditivos para o mercado de ações utilizando *machine* e *deep learning***. 2021. Trabalho de Conclusão de Curso (Ciência da Computação) – Universidade de Caxias do Sul. Disponível em: <https://repositorio.ucs.br/11338/9714>. Acesso em: 20 abr. 2025.

TENSORFLOW. **TensorFlow**. 2024. Disponível em: <https://www.tensorflow.org/>. Acessos em: 16 out. 2024; 20 abr. 2025.

YFINANCE. **Yfinance library documentation**. 2024. Disponível em: <https://pypi.org/project/yfinance/>. Acesso em: 16 out. 2024.

Informações complementares

Descrição		Declaração
Financiamento		Não se aplica.
Aprovação ética		Não se aplica.
Conflito de interesses		Não há.
Disponibilidade dos dados de pesquisa subjacentes		O trabalho não é um <i>preprint</i> e os conteúdos subjacentes ao texto da pesquisa estão contidos neste artigo.
Uso de Inteligência Artificial		Não há.
CrediT	Eduardo Luiz Zanotto	Funções: Conceitualização, análise formal, metodologia, programas, validação e escrita – rascunho original.
	Carlos Amaral Hölbig	Funções: Conceitualização, análise formal, supervisão, escrita – revisão e edição.

Avaliadores: Os avaliadores optaram pela avaliação fechada e pelo anonimato.
Revisora do texto em português: Jéssica Rejane Lima.
Revisora do texto em inglês: Adriana de Oliveira Gomes Araújo.

Revisora do texto em espanhol: Jéssica Rejane Lima.

Como citar:

ZANOTTO, Eduardo Luiz; HOLBIG, Carlos Amaral. Previsão de preços de ações e ETF na bolsa de valores B3 aplicando técnicas de machine learning. **Revista Sítio Novo**, Palmas, v. 10, p. e1879, 2026. DOI: <https://doi.org/10.47236/2594-7036.2026.v10.1879>. Disponível em: <https://sitionovo.ifto.edu.br/index.php/sitionovo/article/view/1879>.