

# Inteligência artificial e deep learning na previsão de preços de ações e etfs na B3

Eduardo L. Zanotto<sup>1</sup>, Carlos Amaral Holbig<sup>2</sup>

<sup>1</sup>Instituto de Tecnologia – Universidade de Passo Fundo (UPF)  
Caixa Postal 611 – CEP 99001-970 – Passo Fundo – RS – Brasil

179530@upf.br, holbig@upf.br

**Abstract.** *The number of new investors on the Brazilian stock exchange is growing. People who seek greater profitability, often without knowing how to analyze opportunities and dangers. This study seeks to predict stock prices on the stock exchange using advanced artificial intelligence and data analysis techniques. Aiming to deliver high assertiveness in the operations carried out. For prediction to be possible, the algorithm used was: LSTM, which was trained separately with historical data from five stocks and two ETFs, including: Banco do Brasil, Itau, Vale, Petrobras, Caixa Seguridade, BOVA11 e FIND11.*  
**Keywords:** *b3, lstm, temporal series.*

**Resumo.** *É crescente o número de novos investidores na bolsa de valores brasileira. Pessoas que buscam uma rentabilidade maior, muitas vezes sem saber analisar as oportunidades e os perigos. Este estudo busca prever os preços de ações na bolsa de valores utilizando técnicas avançadas de inteligência artificial e análise de dados. Visando entregar uma alta assertividade nas operações realizadas. Para a previsão ser possível o algoritmo utilizado foi o LSTM sendo este treinado de forma separada com dados históricos de cinco ações e dois ETFs, dentre elas: Banco do Brasil, Itaú, Vale, Petrobras, Caixa Seguridade BOVA11 e FIND11.*

**Palavras chave:** *b3, lstm, temporal series.*

## 1. INTRODUÇÃO

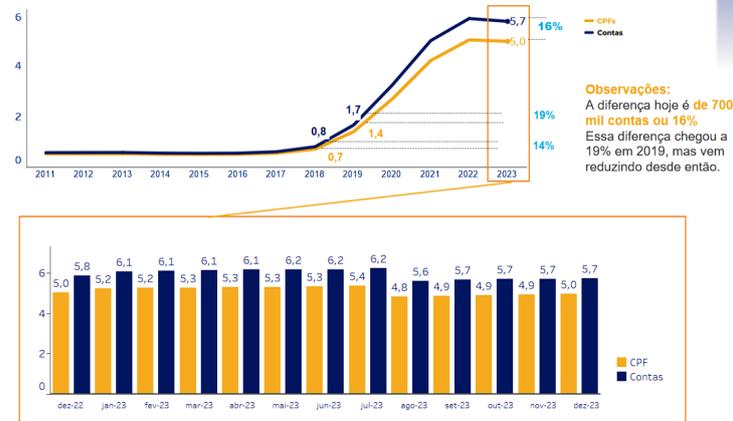
De acordo com um estudo realizado pela bolsa de valores brasileira B3 [B3 2024] em 2023, o número de investidores cresceu 23% entre pessoas físicas e jurídicas. Dessa forma, é perceptível que a população brasileira está demonstrando mais interesse em investir seu dinheiro. Conforme apresentado na Figura 1, pode-se perceber que a quantidade de contas vem aumentando ao longo dos anos. O número de investidores em renda variável também está crescendo, como mostrado na Figura 2, com exceção do ano de 2023, quando a renda fixa atraiu mais investidores, uma vez que a taxa SELIC esteve em alta.

### Número de contas e número de investidores

O número de Investidores PFs na negociação aumentou consideravelmente nos últimos anos. Desde 2020, a média de investidores que fazem ao menos um negócio no mês está acima de 1 milhão.

#### Evolução da Posição das Pessoas Físicas na B3

Conceito de Contas x Investidores por CPFs Distintos | em milhões



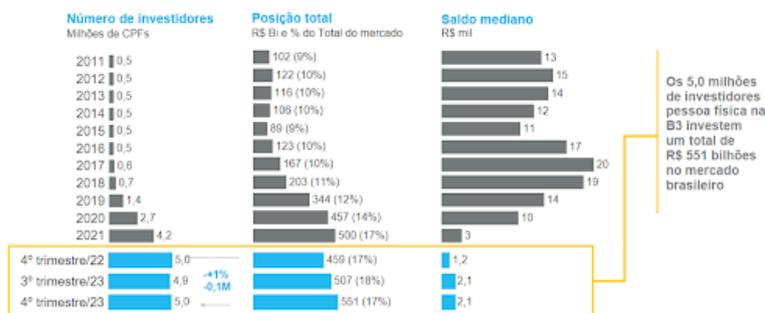
Discâmer: CPF = um investidor / Conta = investidor na corretora

Figura 1. Número de Pessoas contas na Bolsa Brasileira (B3) Dez/2023 [B3 2024]

### Estabilidade de investidores em Renda Variável no último ano

Nos últimos 12 meses tivemos uma estabilidade na quantidade de investidores pessoa física no mercado de capitais. E estes representam 17% do total de recursos investido em equities<sup>2</sup> na B3.

#### Evolução da Posição das Pessoas Físicas na B3



<sup>2</sup> considera-se Ações à Vista, FI, ETF, SDR e outros produtos de equities

Figura 2. Número de investidores na Bolsa Brasileira (B3) em Dez/2023 [B3 2024]

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, há uma carência de conhecimento entre os investidores, conforme a Figura 3, onde muitos não fazem análises ou seguem dicas sem embasamento. Como diz um dos maiores investidores da atualidade:

“Se você não consegue entender o negócio, então não invista nele”  
 - Warren Buffet

### Como ocorre o processo de escolha e análise de investimentos



**Figura 3. Processo de escolha e análise de investimentos. [B3 2024]**

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. 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 dos principais indicadores, há fatores macroeconômicos que influenciam o preço dos ativos. 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.

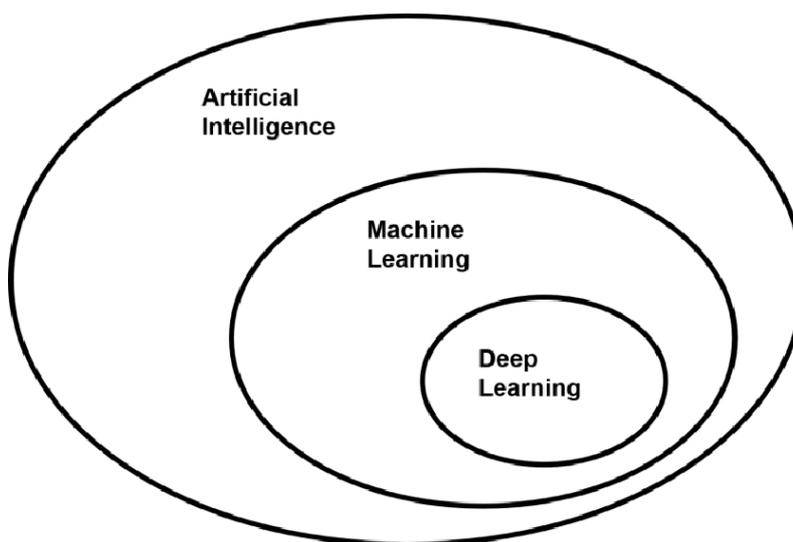
Com o aumento de novos investidores e a necessidade de conhecimento ao analisar ativos, além do crescimento no uso de inteligência artificial nos dias atuais, surge a seguinte questão: é possível prever o preço de ações e ETFs (Exchange-Traded Funds, fundos de investimento negociados em bolsa como ações, que acompanham o desempenho de um índice ou uma cesta de ativos) na bolsa de valores utilizando dados históricos e técnicas de aprendizado de máquina? Para responder a essa questão, será utilizado o modelo Long Short-Term Memory (LSTM), uma arquitetura de rede neural recorrente (RNN) que possui a capacidade de "lembrar" valores em intervalos arbitrários. Essa técnica, pertencente ao domínio do deep learning, destaca-se devido a sua eficácia em lidar com séries temporais cujos intervalos de tempo podem ter duração desconhecida e

alta variação.

Este trabalho tem como objetivo geral destacar a importância da previsão dos preços de ações e ETFs na B3 para os investidores, que enfrentam um ambiente de mercado volátil e complexo. A utilização de técnicas de inteligência artificial e análise de dados de séries temporais permite oferecer aos investidores um software efetivo que auxilie na tomada de decisões, prevendo o preço de ativos na bolsa de valores brasileira com alta assertividade. Em última análise, o estudo busca contribuir para uma melhor compreensão e aplicação da inteligência artificial no contexto financeiro, visando melhorar a precisão das previsões e, conseqüentemente, proporcionar maior rentabilidade aos investidores. A falta de softwares e ferramentas acessíveis para essa finalidade reforça a necessidade deste trabalho e seu potencial impacto positivo na área financeira.

## 2. Fundamentação Teórica

Devido ao desenvolvimento da capacidade de computação e ao grande aumento dos volumes de dados disponíveis, a Inteligência Artificial (IA) apresentou um notável crescimento nos últimos anos. Alan Turing, considerado o pai da computação e da IA por seu trabalho pioneiro, propôs em seu artigo "Computing Machinery and Intelligence" que as máquinas poderiam operar de maneira semelhante aos humanos, resolvendo problemas e tomando decisões com base em informações disponíveis [Turing 1950]. A IA utiliza técnicas avançadas de aprendizado de máquina, como machine learning e deep learning, que processam volumes massivos de dados. Com isso, a IA é capaz de distinguir padrões e aprender com eles, oferecendo informações valiosas e úteis. A Figura 4 ilustra a relação hierárquica entre Inteligência Artificial, Machine Learning e Deep Learning, mostrando como o Deep Learning é um subconjunto do Aprendizado de Máquina, que, por sua vez, está contido no conjunto da IA.



**Figura 4. Relação entre Inteligência Artificial, Aprendizado de Máquina e Deep Learning. Adaptado de Chollet (2021).**

A previsão dos preços de ações e ETFs é uma tarefa desafiadora devido à natureza complexa e altamente volátil do mercado financeiro. No entanto, o avanço das técnicas de inteligência artificial (IA), como o aprendizado de máquina, oferece novas

perspectivas para lidar com essa questão. Christopher M. Bishop, em seu livro "Pattern Recognition and Machine Learning", destaca o potencial das redes neurais artificiais e de outros métodos de aprendizado de máquina na análise de dados financeiros. Como Bishop 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" [Bishop 2011]. 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.

Yoshua Bengio, uma figura proeminente no campo do aprendizado profundo, discute em seu livro "Deep Learning", da série "Adaptive Computation and Machine Learning", a importância das redes neurais profundas na extração de representações complexas dos dados. Como Bengio observa, "As redes neurais profundas têm o potencial de aprender automaticamente características hierárquicas dos dados, permitindo a representação de informações em diferentes níveis de abstração" [Goodfellow et al. 2016]. Essa capacidade das redes neurais profundas é relevante na previsão de preços de ações, tema deste trabalho, onde os dados de séries temporais podem conter uma variedade de padrões e relações complexas. Ao utilizar técnicas de deep learning, como redes neurais convolucionais (CNNs) e redes neurais recorrentes (RNNs), 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 e Jenkins em seu livro "Time Series Analysis: Forecasting and Control", as análises de séries temporais são utilizadas para identificar padrões, tendências e estruturas subjacentes, apoiando uma tomada de decisão sólida [Box et al. 2015]. Hyndman e Athanasopoulos referem-se a esse processo como "previsão de séries temporais" [Hyndman and Athanasopoulos 2021]. 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], explora a aplicação de modelos preditivos na B3, utilizando dados históricos, como preços diários de: abertura, fechamento, máximo, mínimo e volume de negociações. Focando especificamente em 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 destaca que ambos os modelos são eficazes para a previsão de preços no contexto do mercado brasileiro de ações, e a escassez de estudos voltados para a B3.

Já [Alves and Prado 2022] em seu estudo, fazem 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 Redes Neurais Recorrentes, ambas implementadas 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, a Regressão Linear apresentou desempenho competitivo com RNN em termos de eficiência e precisão preditiva.

De acordo com [dos Santos 2022], os modelos Random Forest e LightGBM aplicados em séries temporais de ativos da bolsa B3 demonstram eficácia, especialmente ao utilizar dados como preço e volume de negociação. O trabalho adota duas estratégias de particionamento dos dados (temporal e aleatória) e avalia métricas como acurácia e precisão, com o LightGBM se destacando por seu desempenho superior, mesmo sem ajustes de hiperparâmetros. Isso sugere 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 et al. 2022], desenvolveram o protótipo The Seer, uma aplicação web baseada em inteligência artificial. Utilizando modelos ARIMA, Prophet e LSTM para prever tendências de preços na B3, os autores testaram suas previsões em ativos e ETFs 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 2021] implementa um modelo de previsão de tendências de ações que combina gráficos de candlestick com um esquema de engenharia de recursos baseado nos padrões "oito trigramas". Este 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 3.445 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] aponta que as redes neurais são os modelos mais utilizados, destacando-se pela capacidade de identificar padrões não lineares em dados financeiros. O estudo observa um aumento significativo na adoção do LSTM em previsões financeiras a partir de 2015, enquanto métodos como SVM e KNN são menos recorrentes. A pesquisa oferece 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 et al. 2024] focaram no mercado de ações chinês. O estudo avalia 55 características específicas de empresas, utilizando modelos de aprendizado de máquina como: Redes Neurais e Random Forest e conclui que essas técnicas são eficazes em capturar a volatilidade, principalmente em empresas estatais e em períodos de baixa incerteza econômica. Os autores destacam 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 et al. 2024] propõem 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 incorporam pesos de classe e de tempo ao algoritmo AdaBoost e implementam um framework de aprendizado ativo com anotação automatizada. Os testes empíricos demonstraram que essa abordagem melhora 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.

### 3. Materiais e métodos

Para o desenvolvimento do projeto a linguagem utilizada foi 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 GPU para acelerar o treinamento do modelo. Diversas bibliotecas foram integradas ao projeto: TensorFlow/Keras [TensorFlow 2024] foi utilizada para a construção e treinamento do modelo LSTM, fundamental para as tarefas de previsão; yfinance consiste em uma API para obtenção dos dados históricos utilizados para treinamento e teste do modelo [Yfinance 2024]; pandas foi empregado na manipulação e pré-processamento dos dados [Pandas Development Team 2024]; matplotlib proporcionou a visualização gráfica dos resultados, facilitando a interpretação dos dados [Matplotlib 2024]; scikit-learn foi utilizado para a normalização dos dados e no cálculo de métricas de desempenho, permitindo a avaliação de desempenho do modelo [Scikit-learn 2024]. E o neptune.ai que é um rastreador de experimentos para treinamento de modelos de base, foi utilizado para o monitoramento do treinamento e armazenamento de resultados, permitindo um acompanhamento em tempo real completo e detalhado [Neptune.ai 2024].

#### 3.1. Coleta dos dados

Neste trabalho, foram utilizados dados históricos diários de ações e ETFs da B3, obtidos do Yahoo Finance. Os ativos específicos foram: BBAS3, ITUB4, BOVA11, FIND11, VALE3, PETR4 e CXSE3, no entanto o usuário é livre para digitar qualquer ticker. O período de dados utilizado para o modelo é de 10 anos, contando a partir da data em que o modelo é executado, escolha esta baseada em estudos anteriores que demonstraram bons resultados com esse intervalo [Ayyildiz and 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. Dessa forma, evita-se 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 ajustado e volume de negociações, obtidos utilizando a biblioteca *yfinance*<sup>1</sup>. 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.

---

<sup>1</sup><https://pypi.org/project/yfinance/>

### 3.2. 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, estes impactam diretamente em reajustes nos preços das ações, como o yfinance fornece apenas o preço de fechamento ajustado, é necessário ajustar os preços de: abertura, máxima e mínima para refletir estes eventos.

Com base no estudo de [Santana 2022], primeiro, determina-se o fator de ajuste dividindo o preço de fechamento ajustado pelo preço de fechamento não ajustado. Este fator representa a proporção pela qual os preços devem ser ajustados para refletir os eventos corporativos, em seguida se aplica 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 finais de semana, feriados ou 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, ela 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 para os dados em que o volume de negociações era zero, estes foram substituídos pelo último valor conhecido, utilizando o método de preenchimento forward fill. Dessa forma assegurando a continuidade dos dados, evitando falsas avaliações do modelo.

A média móvel exponencial (EMA) também foi incorporada neste trabalho. 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 (MA), a EMA atribui pesos maiores aos preços mais recentes, o que a torna mais sensível às mudanças recentes no mercado. Neste trabalho, a EMA foi calculada com um período de 60 dias podendo ser alterada em alguns casos, como quando as empresas são recentes. Este período é comum para identificar tendências de médio prazo. Dessa forma equilibra 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 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 and Iskenderoglu 2024]. Dessa forma, a EMA é crucial para a precisão das previsões financeiras.

### 3.3. Pré-processamento dos dados

Na implementação deste trabalho, a normalização visa 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 ETFs 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 et al. 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.

Dando continuidade, 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 em que todas contribuem de maneira equilibrada. É importante que os parâmetros de normalização obtidos do conjunto de treinamento fossem aplicados ao conjunto de teste, para que dessa maneira a consistência seja mantida para 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. Prosseguindo com uma convergência mais rápida do modelo devido às escalas equilibradas das features, necessitando de menos épocas para alcançar um desempenho satisfatório, e conseqüentemente em um melhor desempenho preditivo, evidenciando métricas aprimoradas com menor RMSE e MAE.

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, assegurando que o modelo seja treinado em dados do passado e testado em dados do futuro, refletindo o cenário real de previsão no mercado financeiro.

Neste trabalho, os dados 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. Esta divisão foi realizada com base nos estudos de [Ayyildiz and Iskenderoglu 2024], bem como de [Najem et al. 2024], que utilizaram abordagens semelhantes para garantir a integridade temporal e a representatividade dos dados mais recentes para a validação.

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 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.

### 3.4. Arquitetura do modelo

Diante dos pontos e trabalhos relacionados descritos anteriormente, o modelo escolhido e construído neste trabalho é uma rede neural recorrente do tipo LSTM (Long Short-Term Memory), implementada utilizando a biblioteca TensorFlow Keras para a previsão dos preços de ações e ETFs 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.

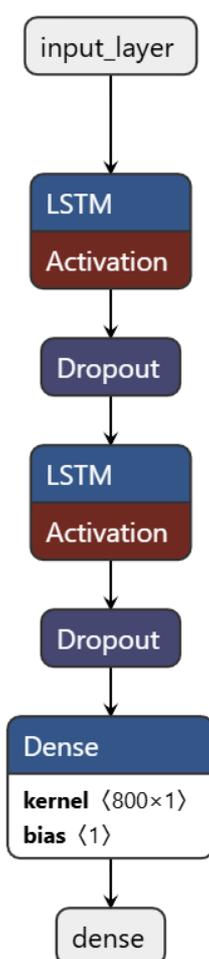
O modelo é estruturado em várias camadas. A camada de entrada recebe sequências de dados com dimensões correspondentes ao tamanho da janela de observação (Window Size) e ao número de features. O Window Size, conforme explicado na seção de pré-processamento dos dados, representa o número de períodos anteriores 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.

A Figura 5 ilustra 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 neste estudo, não é fixo para todos os ativos, e pode ser ajustado para otimizar o desempenho do modelo. A quantidade de unidades é um hiperparâmetro que influencia diretamente a capacidade do modelo em 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 5. Arquitetura do modelo LSTM utilizado para a previsão de preços de ações 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**

### 3.5. 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 foram importantes para chegar em um

resultado satisfatório. Inicialmente, o modelo foi compilado utilizando a função de perda Mean Squared Error (MSE), de acordo com o estudo de Gülmez. B. (2023) esta função 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.

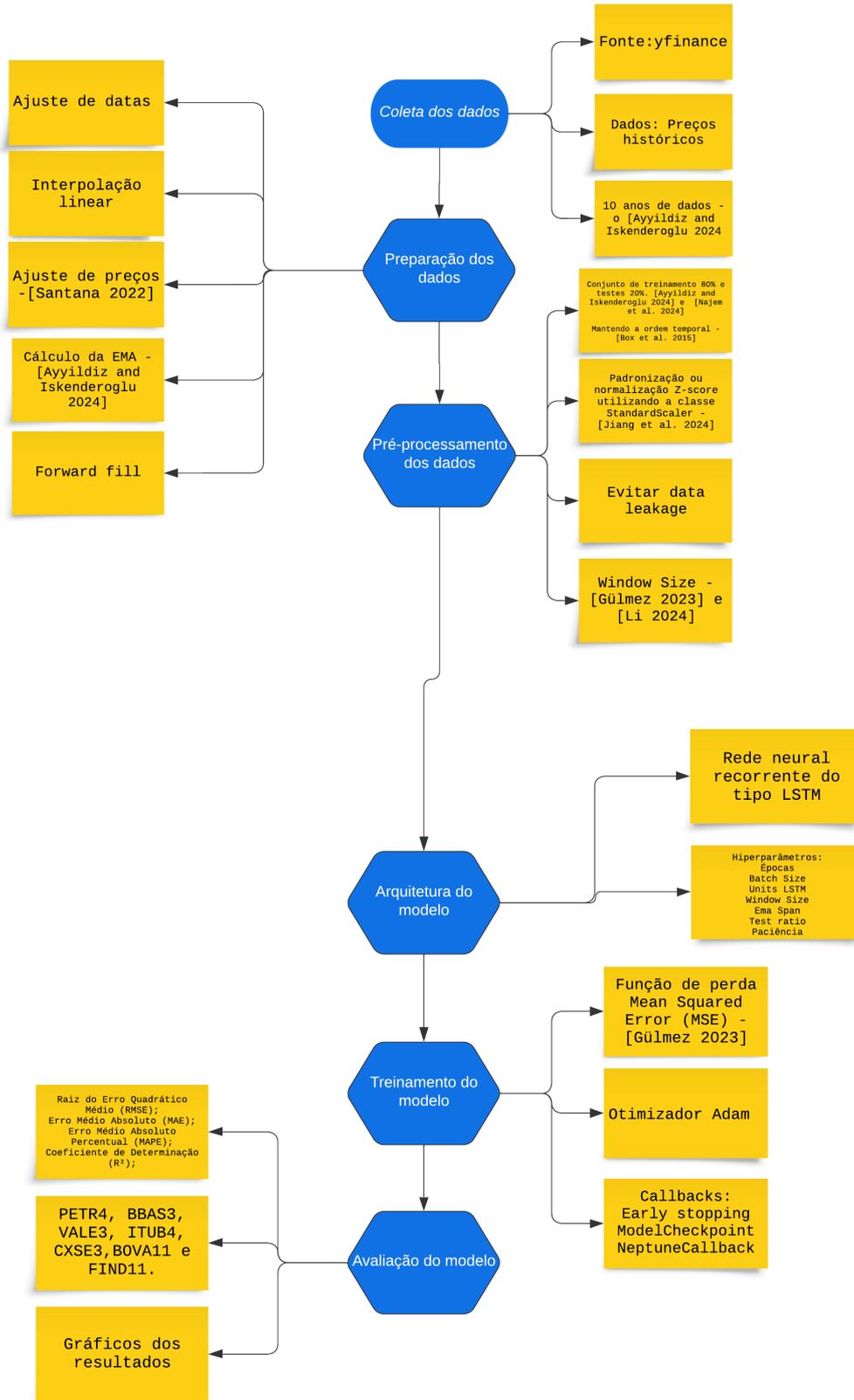
Os hiperparâmetros necessários para o funcionamento do modelo são: Número de épocas, tamanho do batch, unidades lstm, window size, ema span, test ration e paciência do early stopping. Conforme explicação anterior, o conjunto de dados foi dividido em 80% treino e 20% teste, preservando as dependências 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, onde 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 evita o tempo de treinamento desnecessário quando o modelo não apresenta melhorias. Também foi utilizado o callback ModelCheckpoint para salvar automaticamente os pesos do modelo sempre que uma melhoria na perda de validação é observada. Assegurando assim que a melhor versão do modelo seja utilizada em previsões futuras.

A integração do neptune, que é configurada através de um token gerado pelo mesmo e passado no script do modelo LSTM, permite o acompanhamento em tempo real das métricas de desempenho durante o treinamento e teste, facilitou o registro dos hiperparâmetros, visualização de gráficos, visualização do dataset utilizado, taxa de aprendizado e manteve um histórico detalhado dos experimentos realizados.

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 implementado o gráfico que mostra 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, a Figura 6 sintetiza o fluxo completo do trabalho, desde a coleta dos dados até a avaliação dos resultados, auxiliando na visualização geral do processo.



**Figura 6.** 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.

#### 4. Resultados e experimentos

O modelo LSTM desenvolvido neste trabalho tem por objetivo fazer a previsão do preço de ações e ETFs 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 digitar o ticker do ativo 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 seguintes ações: PETR4, BBAS3, VALE3, ITUB4, CXSE3 e ETFs: BOVA11 e FIND11. Escolha esta que permite avaliar a previsão do modelo entre diferentes ativos e setores, de modo que seja possível entender como o modelo se comporta com empresas ou ETFs 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 mede 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, onde valores próximos de 1 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.

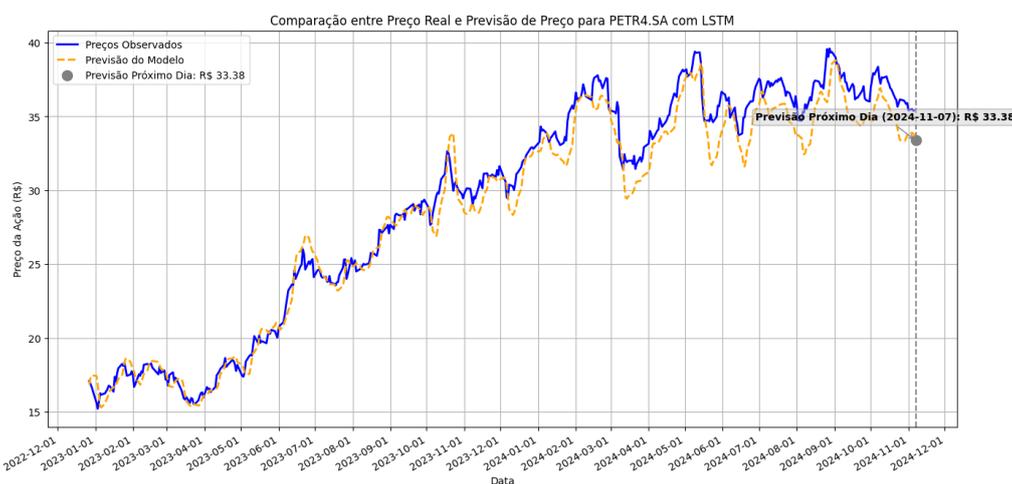
<b>Ativo</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAPE (%)</b>	<b><math>R^2</math></b>
PETR4	1.3247	1.035	3.4042	0.9692
BBAS3	0.5293	0.4360	1.8372	0.9752
VALE3	1.3175	1.0259	1.6758	0.8564
ITUB4	0.4073	0.3246	1.1167	0.9917
CXSE3	0.3077	0.2347	1.6045	0.7713
BOVA11	1.4713	1.1625	0.9839	0.9738
FIND11	1.9832	1.5468	1.3176	0.9705

**Tabela 1. Resultados de métricas dos ativos**

#### 4.1. Análise dos resultados obtidos

A fim de complementar a análise quantitativa, foram gerados gráficos comparando os valores reais e previstos para cada ativo. Esses gráficos permitem visualizar o desempenho do modelo ao longo do tempo, identificando padrões ou discrepâncias nas previsões, além de exibir 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, cujo este não é conhecido e não possui dados.

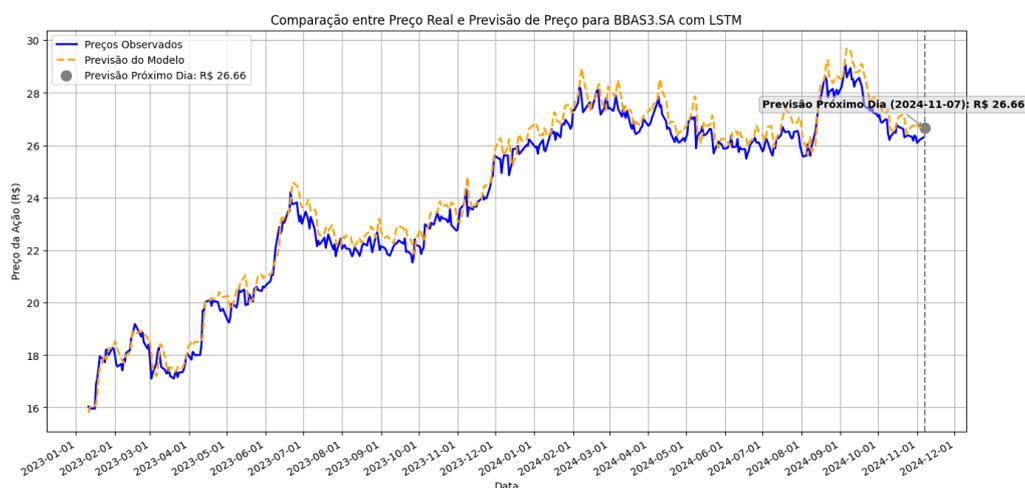
Para o ativo PETR4 (Petrobras S.A.), analisado no período de 10/11/2014 a 06/11/2024, o modelo demonstrou boa precisão, com um RMSE de 1,3247 e MAE de 1,035, refletindo erros médios baixos nas previsões. O MAPE de 3,4042% confirma a proximidade entre os valores previstos e os reais, enquanto o  $R^2$  de 0,9692 indica que o modelo capturou 96,92% da variabilidade dos preços. Conforme o gráfico da Figura 7, as previsões acompanham 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 07/11/2024 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 7. Comparação entre os valores reais e preditos das ações da Petrobras (PETR4), no período de 10/11/2014 a 06/11/2024, com previsão para o dia 07/11/2024**

Para o ativo BBAS3 (Banco do Brasil S.A.), analisado no período de 10/11/2014 a 06/11/2024, 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% evidencia a significativa proximidade entre os valores previstos e observados, e o  $R^2$  de 0,9752

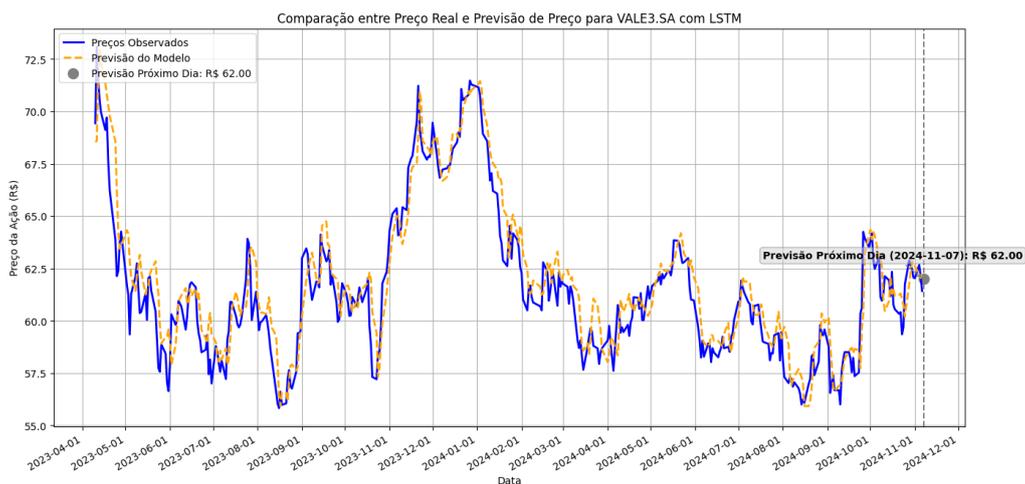
mostra que o modelo capturou 97,52% da variabilidade dos preços da ação, reforçando sua alta acurácia. No gráfico da Figura 8, observa-se que a previsão acompanha de perto os preços reais, com desvios mínimos mesmo em períodos de maior volatilidade. É importante considerar que, sendo o Banco do Brasil 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 07/11/2024 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 8. Comparação entre os valores reais e preditos das ações do Banco do Brasil (BBAS3), no período de 10/11/2014 a 06/11/2024, com previsão para o dia 07/11/2024**

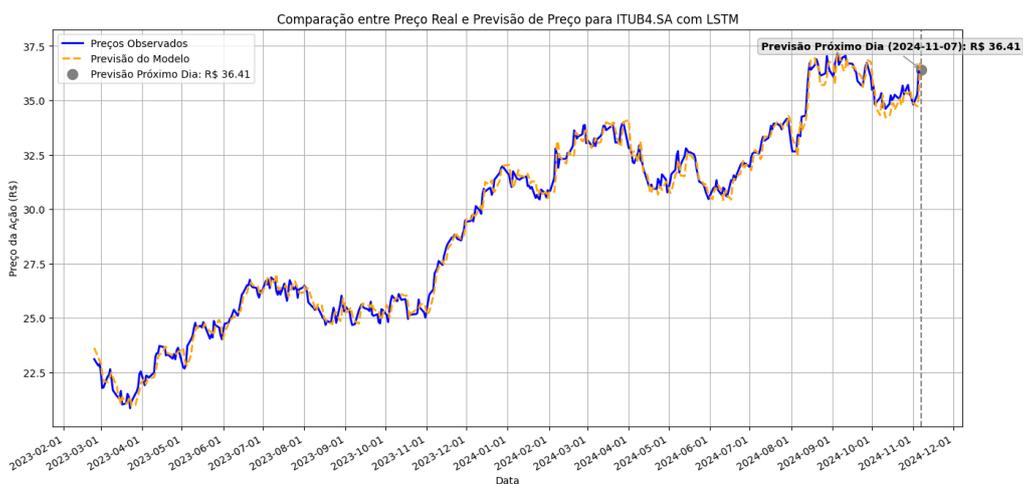
Para o ativo VALE3 (Vale S.A.), analisado no período de 10/11/2014 a 06/11/2024, o modelo apresentou desempenho satisfatório, com RMSE de 1,3175 e MAE de 1,0259, indicando erros médios moderados. O MAPE de 1,6758% sugere boa proximidade entre os valores previstos e reais, embora o  $R^2$  de 0,8564 revele que o modelo capturou aproximadamente 85,64% da variabilidade dos preços, evidenciando alguma dificuldade em prever movimentos mais voláteis. Conforme o gráfico da Figura 9, 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 07/11/2024 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.



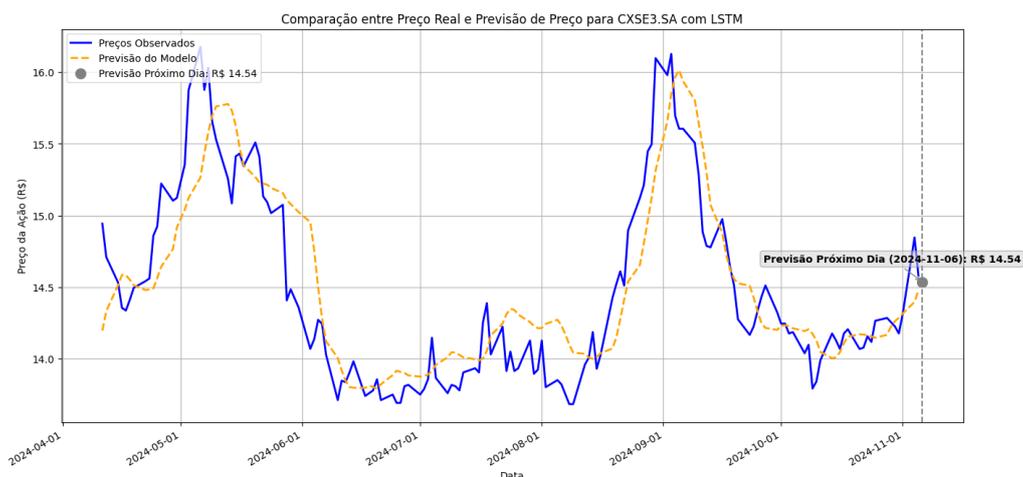
**Figura 9. Comparação entre os valores reais e preditos das ações da Vale (VALE3), no período de 10/11/2014 a 06/11/2024, com previsão para o dia 07/11/2024**

Para o ativo ITUB4 (Itaú Unibanco S.A.), analisado no período de 10/11/2014 a 06/11/2024, 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 10, 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 07/11/2024 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.



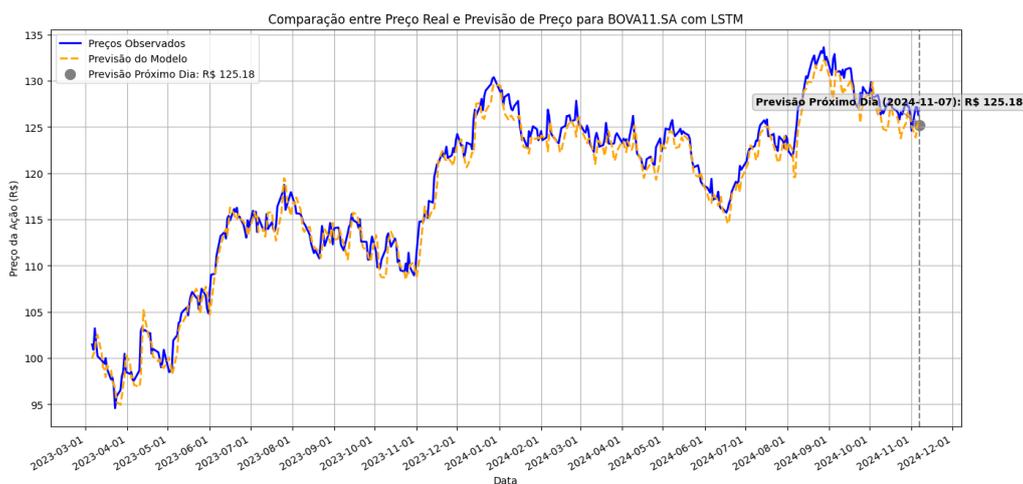
**Figura 10. Comparação entre os valores reais e preditos das ações do Itaú (ITUB4), no período de 10/11/2014 a 06/11/2024, com previsão para o dia 07/11/2024**

Para o ativo CXSE3 (Caixa Seguridade S.A.), analisado no período de 30/04/2021 a 05/11/2024, 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 seu IPO em 2021. O MAPE de 1,6045% ainda reflete precisão razoável, mas, conforme o gráfico na Figura 11, 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 06/11/2024 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.



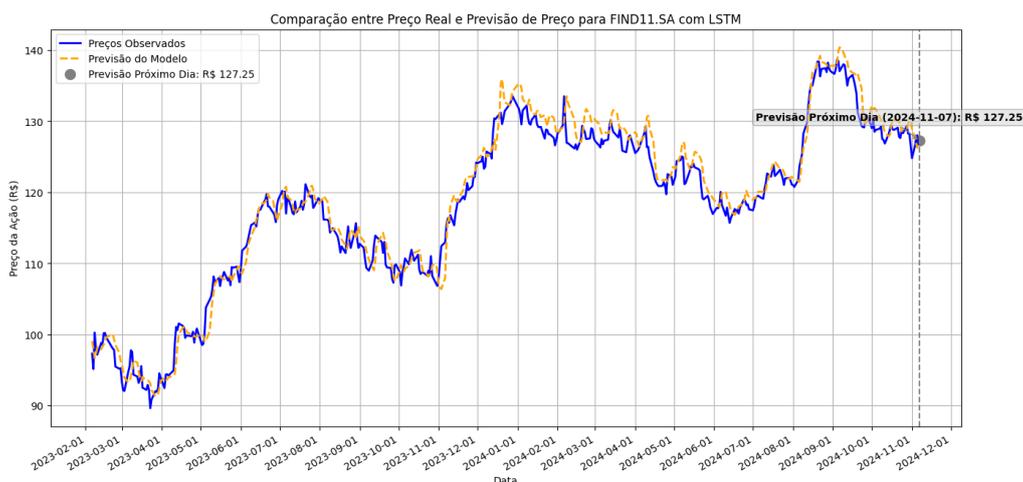
**Figura 11. Comparação entre os valores reais e preditos das ações da Caixa Seguridade (CXSE3), no período de 30/04/2021 a 05/11/2024, com previsão para o dia 06/11/2024**

Para o ETF BOVA11, analisado no período de 10/11/2014 a 06/11/2024, 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 12, 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 07/11/2024 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, mostrando uma pequena diferença em relação ao valor estimado, o que reforça a importância de considerar fatores externos não modelados.



**Figura 12. Comparação entre os valores reais e preditos do ETF BOVA11, no período de 10/11/2014 a 06/11/2024, com previsão para o dia 07/11/2024**

Para o ETF FIND11 que replica o IFNC, cujo acompanha o desempenho médio das empresas com maior representatividade no setor financeiro brasileiro, como bancos, fintechs e seguradoras, analisado no período de 10/11/2014 a 06/11/2024. 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 13, 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 apartir da execução é 07/11/2024 com valor previsto de R\$127,25, reflete 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 13. Comparação entre os valores reais e preditos do ETF FIND11, no período de 10/11/2014 a 06/11/2024, com previsão para o dia 07/11/2024**

#### 4.2. Comparação dos resultados com trabalhos similares

Visando demonstrar a eficácia deste estudo, uma comparação com o modelo LSTM de [Silveira 2021] para o ativo PETR4, observa-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.

Além do resultado numérico, a escolha do trabalho de [Silveira 2021] para comparação justifica-se pela similaridade de contexto e abordagem. Assim como neste estudo, [Silveira 2021] focou na B3, empregando dados históricos de ativos e validando o modelo sobre o ativo PETR4. Ademais, ambas as pesquisas adotaram métricas semelhantes, como RMSE e MSE, e utilizaram arquiteturas baseadas em redes neurais recorrentes 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.

#### 5. Conclusão

Neste estudo, implementou-se um modelo LSTM para prever os preços de ações e ETFs 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 mostrou-se 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.

Para trabalhos futuros, visando a 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

- Alves, P. C. P. and Prado, S. C. C. (2022). Estudo comparativo entre algoritmos de machine learning aplicados à previsão de séries temporais do mercado financeiro. *Congresso de Tecnologia-Fatec Mococa*. Vol. 6. No. 2.
- Ayyildiz, N. and Iskenderoglu, O. (2024). How effective is machine learning in stock market predictions? *Heliyon*, 10(2):e24123.
- B3 (2024). Perfil de investidores pessoa física. [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/](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, C. M. (2011). *Pattern Recognition and Machine Learning*. Springer.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics)*. Wiley.
- dos Santos, J. C. (2022). Abordagens de aprendizado de máquina para séries temporais financeiras. *USP*.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. The MIT Press.
- Google (2024). *Google Colaboratory*. Acesso em: 15 out. 2024.
- Gülmez, B. (2023). Stock price prediction with optimized deep lstm network with artificial rabbits optimization algorithm. *Expert Systems with Applications*, 227:120346.
- Hyndman, R. J. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. Otexts.

- Jiang, F., Ma, T., and Zhu, F. (2024). Fundamental characteristics, machine learning, and stock price crash risk. *Journal of Financial Markets*, 69:100908.
- Li, K. Y. (2024). Machine learning for stock price prediction. <https://neptune.ai/blog/predicting-stock-prices-using-machine-learning>. Acesso em: 16 out. 2024.
- Lin, Yaohu, e. a. (2021). Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme. *IEEE*.
- Matplotlib (2024). *Matplotlib*. Acesso em: 16 out. 2024.
- Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., and Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. *Procedia Computer Science*, 216:96–102. 7th International Conference on Computer Science and Computational Intelligence 2022.
- Najem, R., Amr, M. F., Bahasse, A., and Talea, M. (2024). Advancements in artificial intelligence and machine learning for stock market prediction: A comprehensive analysis of techniques and case studies. *Procedia Computer Science*, 231:198–204. 14th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 13th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (EUSPN/ICTH 2023).
- Nascimento, O., Santos, F., and Ferreira, K. (2022). Previsão de preços de ações utilizando inteligência artificial. In *Anais do I Brazilian Workshop on Artificial Intelligence in Finance*, pages 37–47, Porto Alegre, RS, Brasil. SBC.
- Neptune.ai (2024). *Neptune Documentation*. Acesso em: 16 out. 2024.
- Pandas Development Team (2024). *Pandas Documentation*. Acesso em: 16 out. 2024.
- Python Software Foundation (2024). Python. Acesso em: 15 out. 2024.
- Ren, T., Li, S., and Zhang, S. (2024). Stock market extreme risk prediction based on machine learning: Evidence from the american market. *The North American Journal of Economics and Finance*, 74:102241.
- Santana, D. (2022). Ajustando o preço das ações yahoo finance. <https://medium.com/@sant1/ajustando-o-preo-das-ames-yahoo-finance-9bd5520bf941>. Acesso em: 16 out. 2024.
- Scikit-learn (2024). *Scikit-learn Documentation*. Acesso em: 16 out. 2024.
- Silveira, R. B. d. (2021). Avaliação de modelos preditivos para o mercado de ações utilizando machine e deep learning. *UCS*.
- TensorFlow (2024). *TensorFlow*. Acesso em: 16 out. 2024.
- Turing, A. M. (1950). I.—computing machinery and intelligence. *Mind*, LIX(236):433–460.
- Yfinance (2024). *Yfinance Library Documentation*. Acesso em: 16 out. 2024.