

Previsão de Tendências no Mercado: Classificação de Ações via Análise de Order Book com scikit-learn

Victor F Rodrigues , Ana Clara Senne , Gustavo Lazotti

Faculdade de Computação e Informática (FCI)

Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

{10403390,104,10}@mackenzista.com

Abstract: This project aims to disseminate knowledge about the application of modern artificial intelligence techniques - in particular, LSTM neural network models - for stock price forecasting, contributing to more informed decision-making by investors. Using historical data obtained via Yahoo Finance, the study focuses on Apple's stock (AAPL) from 2010 to 2020, carrying out exploratory analysis, pre-processing and predictive modeling. In addition to offering a potential tool for mitigating risks in the variable income market, the project serves as a learning tool and develops skills in machine learning for students.

Resumo: Este projeto tem como objetivo disseminar o conhecimento acerca da aplicação de técnicas modernas de inteligência artificial – em especial, modelos de redes neurais LSTM – para a previsão de preços de ações, contribuindo para a tomada de decisão mais consciente por parte de investidores. Utilizando dados históricos obtidos via Yahoo Finance, o estudo foca na ação da Apple (AAPL) no período de 2010 a 2020, realizando uma análise exploratória, pré-processamento e modelagem preditiva. Além de oferecer uma ferramenta potencial para mitigar riscos no mercado de renda variável, o projeto serve como instrumento de aprendizado e desenvolvimento de habilidades em machine learning para os discentes.

1. Introdução

O mercado financeiro contemporâneo é amplamente reconhecido por sua alta volatilidade e incertezas, características que tornam a tomada de decisão um desafio constante para investidores. Nesse cenário dinâmico, a interpretação eficaz de dados históricos e a identificação de tendências emergem como competências essenciais para a mitigação de riscos e a maximização de retornos (Fama, 1970; Malkiel, 2003). Com o avanço da inteligência artificial (IA) e das técnicas de deep learning, novas possibilidades de modelagem preditiva surgiram, tornando-se uma alternativa poderosa para o suporte analítico no processo decisório (Goodfellow, Bengio e Courville, 2016).

Modelos de redes neurais recorrentes, em especial as Long Short-Term Memory networks (LSTM), destacam-se por sua capacidade de capturar padrões temporais complexos em séries financeiras, superando abordagens tradicionais baseadas apenas em análise estatística clássica (Brownlee, 2017; Fischer & Krauss, 2018). A utilização dessas metodologias permite maior precisão na previsão de preços de ativos e oferece suporte estratégico na identificação de pontos de reversão de tendência e momentos críticos de mercado.

A democratização do acesso a métodos modernos de análise financeira, por meio da disseminação de ferramentas baseadas em IA, tem o potencial de reduzir desigualdades socioeconômicas, capacitando pequenos investidores a competir em condições mais equilibradas com grandes players institucionais (Faceli et al., 2021). Ao ampliar o acesso a tecnologias sofisticadas, promove-se uma maior inclusão financeira e um ambiente de investimento mais justo e transparente.

Além disso, o ensino e a aplicação prática de técnicas de IA, como a modelagem com LSTM, constituem um diferencial formativo para estudantes de ciências exatas e econômicas, proporcionando o desenvolvimento de competências técnicas altamente demandadas pelo mercado de trabalho contemporâneo (Géron, 2019; Chollet, 2021). O contato direto com dados reais e com metodologias modernas de machine learning permite a consolidação de habilidades críticas em ciência de dados, análise preditiva e finanças quantitativas.

Assim, este projeto surgiu pela combinação de dois fatores principais: (1) a importância social e econômica de oferecer ferramentas preditivas acessíveis, que possibilitem uma melhor gestão de investimentos, e (2) a oportunidade de proporcionar a todos interessados uma formação prática e aplicada em inteligência artificial, utilizando dados históricos e metodologias de ponta para a solução de problemas reais no mercado financeiro.

2. Descrição do Problema

O principal desafio enfrentado por investidores no mercado de renda variável é a alta volatilidade e a imprevisibilidade dos preços das ações. Muitos investidores dependem de análises tradicionais e indicadores básicos, os quais podem não capturar de forma eficaz as complexas dinâmicas do mercado. Dessa forma, a ausência de ferramentas preditivas avançadas limita a capacidade de antecipar mudanças bruscas e gerenciar riscos de forma eficaz. O projeto propõe o uso de modelos LSTM, que são capazes de capturar padrões temporais complexos, para melhorar a acurácia na previsão de preços e fornecer insights valiosos para a tomada de decisão.

3. Aspectos Éticos do uso da IA no Mercado Financeiro

A aplicação de técnicas de inteligência artificial no setor financeiro envolve importantes questões éticas e de responsabilidade que não podem ser negligenciadas. Em primeiro lugar, a **transparência** dos sistemas automatizados é essencial. Modelos preditivos, especialmente aqueles baseados em redes neurais profundas, como as LSTM, tendem a ser percebidos como "caixas-pretas" devido à sua complexidade interna (Goodfellow, Bengio e Courville, 2016). Para mitigar esse desafio, torna-se imprescindível o desenvolvimento de mecanismos de explicabilidade (XAI - Explainable Artificial Intelligence), possibilitando que investidores e usuários compreendam as limitações e o funcionamento dos algoritmos utilizados (Ribeiro, Singh e Guestrin, 2016).

A **equidade** também é uma preocupação central. Algoritmos de machine learning podem inadvertidamente reproduzir ou amplificar vieses existentes nos dados históricos (Barocas, Hardt e Narayanan, 2019). No contexto financeiro, isso poderia gerar práticas discriminatórias que favoreçam determinados grupos de investidores em detrimento de outros, acentuando desigualdades já presentes. Assim, é fundamental que as etapas de desenvolvimento e treinamento dos modelos incluam estratégias de detecção e mitigação de vieses algorítmicos.

Em termos de **privacidade e segurança**, mesmo quando se utilizam dados publicamente disponíveis, como os de mercados financeiros, o tratamento adequado dessas informações é imprescindível para evitar exposição indevida ou usos não autorizados (Shokri et al., 2017). A adoção de boas práticas em segurança da informação, como a implementação de protocolos de anonimização e controle de acesso, é uma exigência ética para garantir a proteção dos dados dos investidores e a integridade do ambiente de negociação.

A **responsabilidade** social e institucional na criação e aplicação de sistemas baseados em IA é outro aspecto crucial. Desenvolvedores e pesquisadores devem estar cientes de que previsões financeiras podem induzir comportamentos de mercado e gerar impactos em larga escala (Russell e Norvig, 2022). Dessa forma, recomenda-se a implementação de auditorias periódicas e o estabelecimento de salvaguardas para prevenir o uso indevido dos modelos preditivos, evitando práticas como manipulação de mercado ou especulação irresponsável.

Por fim, a construção de sistemas éticos e transparentes em finanças, alinhados com os princípios da responsabilidade social corporativa e com os Objetivos de Desenvolvimento Sustentável (ODS) da ONU, é fundamental para que as soluções de inteligência artificial possam gerar benefícios sociais, econômicos e ambientais de maneira equitativa (Floridi et al., 2018).

4. Dataset

A obtenção dos dados históricos da ação da Apple (AAPL) para este projeto foi realizada por meio da biblioteca `yfinance`, uma ferramenta popular em Python para acesso a dados financeiros públicos da plataforma Yahoo Finance. A `yfinance` fornece uma interface simplificada para download de séries temporais de preços de ativos, permitindo a extração de informações como preços de abertura (Open), fechamento (Close), máximos (High), mínimos (Low), volume de negociação (Volume), além de eventos corporativos como pagamento de dividendos (Dividends) e desdobramentos de ações (Stock Splits) (Ran, 2019).

Entre as principais vantagens da `yfinance` destacam-se sua facilidade de uso, atualizações frequentes e a capacidade de obter dados ajustados automaticamente para splits e dividendos, o que é crucial para análises financeiras realistas. A biblioteca também suporta consultas de múltiplos ativos simultaneamente, personalização de intervalos de tempo (diário, semanal, mensal) e integração fluida com bibliotecas populares de ciência de dados, como `pandas` e `NumPy`. Esse conjunto de funcionalidades torna a `yfinance` uma escolha robusta para projetos acadêmicos e aplicações práticas que envolvem análise de séries temporais financeiras (VanderPlas, 2017).

No presente trabalho, os dados extraídos via `yfinance` abarcaram o período de 1º de janeiro de 2010 a 31 de dezembro de 2020, fornecendo uma base sólida e consistente para a construção de modelos preditivos. A qualidade e a diversidade dos dados obtidos permitiram a realização de uma análise exploratória detalhada, a aplicação de técnicas de pré-processamento adequadas, e a criação de sequências temporais que alimentaram a rede LSTM utilizada na modelagem de preços futuros.

5. Metodologias e Resultados Esperados

A metodologia adotada neste projeto envolve a aplicação de técnicas de ciência de dados e aprendizado profundo para a previsão de preços da ação da Apple (AAPL). A primeira etapa consistiu na **coleta de dados históricos** utilizando a biblioteca `yfinance`, que permitiu a obtenção de informações diárias abrangendo preços de abertura, fechamento, máximos, mínimos, volume de negociação, dividendos e eventos de desdobramento de ações. Este conjunto de dados, cobrindo o período de 2010 a 2020, serviu como base para a construção e treinamento do modelo preditivo.

Em seguida, foi realizada uma **análise exploratória** detalhada com o objetivo de entender a estrutura dos dados e identificar padrões relevantes. Foram aplicadas técnicas estatísticas descritivas, bem como visualizações gráficas, como histogramas, boxplots e heatmaps de correlação, para investigar a distribuição das variáveis e suas inter-relações. Esta etapa foi fundamental para detectar eventuais anomalias ou outliers que pudessem comprometer a modelagem subsequente.

O **pré-processamento** dos dados incluiu a normalização das variáveis numéricas utilizando a técnica `MinMaxScaler`, garantindo que todas as entradas estivessem na mesma faixa de valores. Além disso, foram geradas **sequências temporais** com uma janela de 60 dias, estratégia comum em problemas de previsão de séries temporais, de modo a capturar as dependências temporais necessárias para alimentar o modelo de rede neural LSTM (Hochreiter & Schmidhuber, 1997).

Para a **modelagem**, foi implementada uma rede neural do tipo Long Short-Term Memory (LSTM), reconhecida por sua eficácia na modelagem de dados sequenciais e séries temporais financeiras. O modelo foi treinado com os dados históricos e avaliado por meio de sua divisão em conjuntos de treino (80%) e teste (20%), de maneira a garantir a robustez na validação dos resultados. A performance da rede foi mensurada principalmente pela métrica do **Erro Quadrático Médio (MSE)**, sendo esta apropriada para problemas de regressão contínua como a previsão de preços (Géron, 2019).

Entre os **resultados esperados**, destaca-se a obtenção de **previsões acuradas** dos preços de fechamento das ações, permitindo a identificação de tendências de mercado e possíveis pontos de reversão. Além disso, o projeto busca contribuir para a **redução de riscos** nas decisões de investimento, fornecendo uma ferramenta que auxilie investidores a gerenciar suas carteiras de forma mais estratégica. Por fim, o projeto tem um forte caráter educacional, proporcionando aos discentes uma oportunidade prática de aplicar conceitos de **machine learning** e **análise de séries temporais** em um problema real do mercado financeiro, servindo como referência para iniciativas acadêmicas e profissionais futuras.

6. Referências

- Fama, E. F. (1970). *Efficient Capital Markets: A Review of Theory and Empirical Work*. Journal of Finance, 25(2), 383–417. <https://doi.org/10.2307/2325486>
- Malkiel, B. G. (2003). *The Efficient Market Hypothesis and Its Critics*. Journal of Economic Perspectives, 17(1), 59–82. <https://doi.org/10.1257/089533003321164958>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Brownlee, J. (2017). *Deep Learning for Time Series Forecasting*. Machine Learning Mastery.
- Fischer, T., & Krauss, C. (2018). *Deep learning with long short-term memory networks for financial market predictions*. European Journal of Operational Research, 270(2), 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>
- Faceli, K., Lorena, A. C., Gama, J., & Carvalho, A. C. P. L. F. (2021). *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina* (2ª ed.). LTC.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
- Chollet, F. (2021). *Deep Learning with Python* (2nd ed.). Manning Publications.
- Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. <http://fairmlbook.org>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Minds and Machines, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. <https://doi.org/10.1145/2939672.2939778>
- Russell, S., & Norvig, P. (2022). *Inteligência Artificial: Uma Abordagem Moderna* (4ª ed.). LTC.
- Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). *Membership Inference Attacks Against Machine Learning Models*. Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP). <https://doi.org/10.1109/SP.2017.41>
- Ran, D. (2019). *yfinance: Download Market Data from Yahoo! Finance API with Python*. Disponível em: <https://github.com/ranaroussi/yfinance>. Acesso em: 6 abr. 2025.

- VanderPlas, J. (2017). *Python Data Science Handbook*. Sebastopol: O'Reilly Media.
 - Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
 - Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>