

Análise do Mercado de Ações

Predição de Valores

Cristiano Fuschilo^{1,2}, Prof Sergio Serra¹, Jorge Zavaleta¹

¹Instituto de Computação – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brazil

²Coordenação de Sistemas de Informação – Centro Federal de Educação Tecnológica
Celso Suckow da Fonseca (Cefet/RJ) Rio de Janeiro – RJ – Brazil

cristiano.fuschilo@cefet-rj.br, {serra,zavaleta}@pet-si.ufrj.br

Abstract. *The stock market is one of the most popular areas within the financial market. Nowadays, through the popularization of the internet, the media and the democratization of information, it has become one of the most common ways to obtain alternative income. Stock price prediction has always been widely researched, but due to its dynamic and volatile nature, it is considered one of the most difficult tasks in the field of mathematics and computer science. This study seeks to combine intelligent systems to predict the final prices of a given stock among the most consolidated and traded on the market. Concepts about financial markets, machine learning and data visualization will be studied to support the project. Models were proposed to analyze and find patterns in stock prices, as well as to indicate long-term trends, and finally the implementation of the project for general use. For the analyses, data from yfinance over a period of one year, available in the historical quotes section directly from the Nasdaq website, were used. The study was developed using neural networks and a support vector machine using data from the closing price of stocks traded on the stock exchange.*

Resumo. *O mercado de ações é uma das áreas mais populares dentro do mercado financeiro, hoje em dia, através da popularização da internet, da mídia e da democratização da informação, tornou-se uma das formas mais comuns de obtenção de renda alternativa. A previsão do preço das ações sempre foi muito pesquisada, mas devido à sua natureza dinâmica e volátil, é considerada uma das tarefas mais difíceis no campo da matemática e da ciência da computação. Esse estudo busca combinar sistemas inteligentes para prever os preços finais de uma determinada ação das mais consolidadas e negociadas no mercado. Conceitos sobre mercado financeiro, aprendizado de máquina e visualização de dados serão estudados para dar suporte ao projeto. Foram propostos modelos para analisar e encontrar padrões nos preços das ações, bem como indicar tendências de longo prazo, e por fim a implementação do projeto para uso geral. Para as análises foram utilizados os dados do yfinance no período de um ano, disponíveis na seção de cotações históricas diretamente do site Nasdaq. O estudo foi desenvolvido utilizando redes neurais e uma máquina de vetores de suporte utilizando dados do preço de fechamento das ações negociadas na bolsa de valores.*

1. Introdução

O mercado de ações é afetado por muitos macro fatores econômicos, como cita [Tan et al. 2007], como eventos políticos, políticas da empresa, condições econômicas gerais, expectativas dos investidores, decisões de investidores institucionais, movimento de outros mercados de ações, psicologia dos investidores e etc. Para além dessas características intrínsecas ao mercado financeiro, [Viana 2022] afirma que o número de pessoas físicas, que investem na bolsa como alternativa de renda, vem crescendo. Por conseguinte, o mercado de investimentos cresceu proporcionalmente impulsionado pela democratização da educação financeira, plataformas de investimento, cursos, mídias sociais, entre outros meios de comunicação.

Mesmo com acesso à informação em ascensão, o conhecimento necessário para se analisar investimentos é grande, em razão da complexidade e quantidade dos dados. Isso motiva o uso da Ciência de Dados (aprendizado de máquina / predição) como abordagem para sanar tal problema. Uma análise para tomar decisões é uma atividade complexa, requer tempo e estudo para compreender o problema, desenvolver e treinar um modelo que antecipe satisfatoriamente o mercado.

Neste estudo, utilizaremos os dados do mercado de ações norte americano a fim de prever o valor de uma determinada ação utilizando modelos de predição, a fim de comparar a assertividade de cada um deles. Visto o crescimento do mercado financeiro, este estudo tem como cerne contribuir no desenvolvimento e confrontar os modelos de aprendizado de máquina para prever preços de ações.

1.1. Problema Observado

As aplicações financeiras tem tido um crescimento significativo, segundo a [B3 2024], tendo uma alta de 2% no ano. Desde 2020, a base de investidores já cresceu mais de 80% acima do crescimento observado durante a pandemia da COVID 19 onde o número de pessoas que investem em bolsa saltou de 1 milhão em maio de 2019 para quase 3,2 milhões em outubro de 2020.

Problemas como incerteza do mercado de ações e consumo de informações complexas para tomada de decisões fundam a possibilidade de pesquisas e estudos com o intuito de solucionar ou dar ferramentas que minimizem esses problemas.

A previsão do valor de ações é considerada um problema complexo dada sua natureza altamente dinâmica, segundo [Abu-Mostafa and Atiya 1996] é considerada uma tarefa muito difícil a previsão do mercado de ações em previsão de séries temporais financeiras.

1.2. Objetivo

Analisar o mercado de ações coletando, raspando e armazenando dados das ações ao longo do tempo. Processar esses dados para garantir que sejam padronizados, confiáveis e correlacionados. Analisar tecnicamente as tendências do mercado e avaliar empresas e suas perspectivas utilizando ferramentas de visualização. Desenvolver um modelo para prever consistentemente o preço final diário das ações com precisão satisfatória. Verificar hipóteses e estimativas para impactar positivamente investidores iniciantes.

2. Fundamentação

Em um mundo globalizado, a forte interdependência entre a economia, as sociedades e os mercados de capitais, estimulam a elaboração de ferramentas que possibilitem o entendimento e gerência dessa complexidade. A base desse estudo tem conceitos que devem ser explorados.

2.1. Bolsa de Valores

Bolsa de valores, é mercado organizado para negociação de ações e títulos. Esses mercados foram originalmente abertos a todos, mas atualmente apenas os membros da associação proprietária podem comprar e vender diretamente. Os associados, ou corretores da bolsa, compram e vendem para si ou para terceiros, cobrando comissões por seus serviços. Uma ação pode ser comprada ou vendida apenas se estiver listada em uma bolsa e não pode ser listada a menos que atenda a certos requisitos estabelecidos pelo conselho de administração da bolsa, como visto em [Teweles and Bradley 1998].

2.2. NASDAQ

A NASDAQ (National Association of Securities Dealers Automated Quotations) foi lançada em 1971 para prestar serviços eletrônicos e de listagem para as empresas que vinham sendo negociadas no mercado de balcão, e não em uma bolsa de valores. Caracteriza-se por não contar com um pregão físico, pois as transações são integralmente fechadas por meio de sistemas eletrônicos de negociação. Tornou-se o local de listagem para muitas das novas empresas da era da alta tecnologia, como, por exemplo, a Intel, a Microsoft e a Apple, que lá captaram recursos por meio do lançamento de suas ações e permanecem nesta bolsa até hoje. O seu principal slogan é: “Bem vindo à NASDAQ, onde as ideias de amanhã encontram capital hoje!”

Esta negociação eletrônica, cujo modelo foi revolucionário há 40 anos, agora é o padrão para os mercados em todo o mundo. Com tecnologia de vanguarda, ênfase na transparência e um conjunto de ferramentas avançadas, oferece valor exclusivo que atrai novas empresas ao seu mercado. Com a proposta de valor único, é o local de listagem escolhido pelas mais variadas empresas do mundo.

2.3. Predição

Cada vez mais, economistas lidam com problemas complexos e que exigem ferramentas poderosas para serem resolvidos. Visto que os métodos tradicionais abordados na econometria, como o mínimos quadrados ordinários (MQO), apresentam desempenho satisfatório ao realizar inferências causais, mas não exibem a mesma assertividade ao lidar com problemas de previsão [Varian 2014], é importante a tentativa de investigar a eficiência de outros métodos. A bibliografia aponta limitações do MQO ao realizar inferências preditivas. Dessa forma, estudos tratando de situações complexas e próximas à realidade são comprometidos.

Uma das sugestões da literatura para lidar com problemas preditivos é a utilização de algoritmos de machine learning (aprendizado de máquina). Mesmo apresentando limitações que não garantam sua total eficácia em qualquer problema preditivo, em se tratando de casos complexos, essa abordagem aparenta ser a mais satisfatória

[Fernández-Delgado et al. 2014]. Dessa forma, questionam-se quais algoritmos, exatamente, são os melhores para problemas de previsão e em quais casos podem ser aplicados a fim de gerarem bons resultados

2.4. Machine Learning

O aprendizado de máquina é uma das principais subáreas da inteligência artificial, e é composto por uma coleção de métodos criados a partir de modelos matemáticos baseados na teoria estatística que permitem aos computadores automatizar tarefas com base na descoberta sistemática de padrões nos conjuntos de dados disponíveis ou em experiências passadas [Bhavsar et al. 2017].

No aprendizado supervisionado, o modelo recebe um conjunto de entradas com suas respectivas saídas e busca encontrar uma função que estabeleça uma relação aproximada entre elas. Mais formalmente, o modelo baseado no aprendizado supervisionado busca encontrar uma função $h(x_i)$, denominada hipótese, que se aproxime da função $f(x_i)$, onde $f(x_i)$ é a saída da i -ésima entrada de x [Norvig and Intelligence 2002].

2.4.1. LSTM

LSTM (Long short-term memory) são redes neurais recorrentes que podem aprender a dependência de ordem em problemas de previsão de sequência. A LSTM é usada em tradução de máquina, reconhecimento de fala, etc., devido a seus recursos favoráveis para solução de problemas complexos. LSTM pode armazenar informação prévia também. Isso contribui neste estudo, dado que preços anteriores de ações são fundamentais para a previsão de preços futuros.

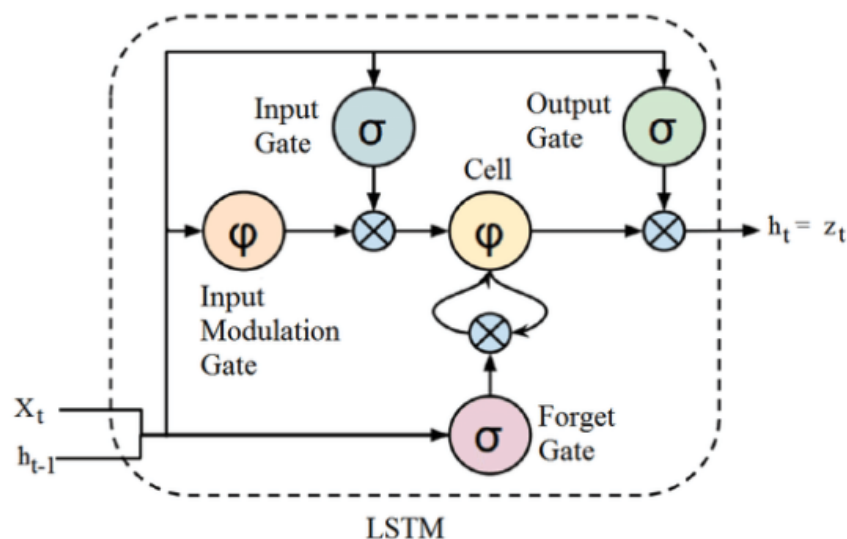


Figure 1. Estrutura LSTM

Fonte: Retirado de deeplearningbook "Capítulo 51– Arquitetura de Redes Neurais Long Short Term Memory (LSTM)"

A estrutura de uma LSTM - como pode ser observado na figura 1 consiste na célula, portão de input, portão de output, e portão de esquecimento. A arquitetura da LSTM é um tipo de rede neural recorrente (RNN) utilizada na área de deep learning. Em contraste, para redes neurais feed-forward convencionais, LSTM tem conexões de feed-back. Uma típica unidade de LSTM compreende quatro componentes: a célula, portão de input, portão de output, e portão de esquecimento.

3. Metodologia

Como já foi dito anteriormente, a ferramenta utilizada para a análise dos dados é o Machine Learning. Para isso será utilizada a linguagem Python e as bibliotecas fundamentais para a análise de dados como: Scikit-Learn, Jupyter Notebook, Numpy, Matplotlib e Pandas

Um modelo que se utiliza de Machine Learning passa pelo treino e teste de dados a fim de verificar se é um bom preditor. Os dados, que estão ordenados em ordem crescente cronologicamente, são subdivididos em dados de treino e dados de teste. O ideal é que a maior parte dos dados seja destinada ao treino e o restante ao teste.

3.1. Coleta e Tratamento

Para desenvolver esse estudo, inicialmente realizou-se a coleta dos dados, usando o yfinance que permite a extração de dados financeiros de várias fontes, como Yahoo Finance. Para posteriormente analisar e selecionar os campos relevantes para assim transformá-los em indicadores técnicos. Os dados públicos utilizados são cotações de diversas ações listadas na bolsa de valores, para o escopo deste estudo foram escolhidas as quatro (4) principais empresas de tecnologia, sendo elas: "APPLE", "GOOGLE", "MICROSOFT" e "AMAZON".

Foi definido a coleta dos dados relativos a um ano menos, a partir da data que o algoritmo é rodado ($n - 1$ ano). Com esses dados armazenados é necessário agora saber quais são eles e que tipo de dado ele tem. Para isso é utilizada a função `.info()` a fim de analisar os campos e tipos de dados de cada campo. Para cada ação temos seis campos, sendo:

Item	Nome	Não Nulos	Tipo	Descrição
0	Adj Close	251 non-null	float64	Preço de fechamento ajustado
1	Close	251 non-null	float64	Preço de fechamento do dia
2	High	251 non-null	float64	Preço máximo do dia
3	Low	251 non-null	float64	Preço mínimo do dia
4	Open	251 non-null	float64	Preço de abertura do dia
5	Volume	251 non-null	float64	Volume negociado no dia

Table 1. Dados Lidos yfinance

A partir daí é possível determinar quais são os dados pertinentes a serem usados nesse estudo; tendo como a principal característica o fato de os dados serem válidos somente em dias úteis, por isso é imprescindível ignorar os dias onde a consulta de dados retorna valores vazios.

3.2. Desenvolvimento

Tendo em vista que o objetivo do presente trabalho é verificar a predição do mercado acionário americano é preciso que as ações escolhidas para a análise sejam ações de maior visibilidade. Para isso, foram escolhidas quatro grandes empresas de tecnologia, quem possuem o índice NASDAQ.

Os dados foram extraídos no período de um ano, do site Yahoo Finance e possuem frequência diária. Alguns gráficos foram necessários a fim de responder perguntas desse estudo, sendo essas, a visão histórica do valor de fechamento, o volume total de ações negociadas a cada dia, a média móvel das várias ações, o retorno diário médio da ação, correlação entre as ações, qual o risco assumido x retorno esperado ao investir em uma determinada ação e por fim o mais importante e o alvo desse estudo "Previsão do preço de fechamento das ações da APPLE inc" onde aplicaremos as técnicas de aprendizagem de máquina e predição para chegar a essa resposta.

4. Resultados

Para determinarmos a previsão do preço de fechamento de uma determinada ação é preciso além de obter a cotação de ações, criar um novo dataframe com apenas a coluna preço de fechamento ('Close') e converter o dataframe em uma matriz numpy para obter o número de linhas para treinar os modelos de machine learning.

A fase de treinamento é composta por seis passos sendo eles: Escalar os dados, criar o conjunto de dados de treinamento, criar o conjunto de dados de treinamento escalado, dividir os dados em conjuntos de dados x_train e y_train, converter o x_train e y_train em matrizes numpy e remodelar os dados

4.1. Para LSTM

A princípio para avaliar os modelos, foram treinados com o conjunto de dados do yfinance. Sendo separados em 3 conjuntos para a LSTM: 70 por cento para treino, 20 por cento para validação e 10 por cento para teste. A seguir estão as figuras e quadros com os resultados dos modelos para os conjuntos citados anteriormente. Gráficos obtidos na análise exploratória sobre os dados para entender como eles se comportam ao longo do tempo.

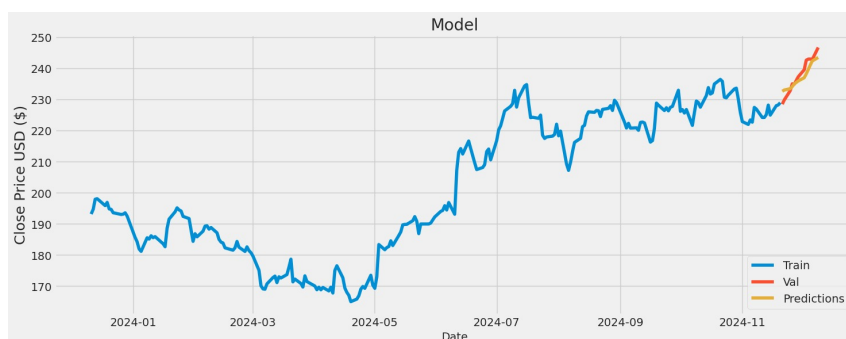


Figure 2. Resultado LSTM

Como pode-se analisar mediante o comportamento do gráfico demonstrado na figura 2 e na tabela 2 o preço previsto de fechamento ('Predictions') tem os valores bastante próximos ao valor de fechamento ('Close').

Data	Close	Predictions
2024-11-21	228.520004	232.652954
2024-11-22	229.869995	233.056152
2024-11-25	232.869995	233.514053
2024-11-26	235.059998	234.171646
2024-11-27	234.929993	235.044373
2024-11-29	237.330002	235.970764
2024-12-02	239.589996	237.028137
2024-12-03	242.649994	238.230423
2024-12-04	243.009995	239.631256
2024-12-05	243.039993	241.058975
2024-12-06	242.839996	242.394363
2024-12-09	246.750000	243.553650

Table 2. Recorte Dados LSTM

Assim podemos afirmar que os resultados apresentados foram satisfatórios, dada a dificuldade da tarefa. Os modelos nos ajudam a entender padrões em grandes volumes de dados e auxiliam na tomada de decisão. Neste projeto foram analisados dados de fechamento diário, aumentando a granularidade dos dados poderíamos obter preços e negociações na ordem de minutos ou segundos, observando tendências ao longo do dia.

References

- Abu-Mostafa, Y. S. and Atiya, A. F. (1996). Introduction to financial forecasting. *Applied intelligence*, 6:205–213.
- B3 (2024). Perfil pessoas físicas. 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/.
- Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., and Dera, D. (2017). Machine learning in transportation data analytics. In *Data analytics for intelligent transportation systems*, pages 283–307. Elsevier.
- Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The journal of machine learning research*, 15(1):3133–3181.
- Norvig, P. R. and Intelligence, S. A. (2002). A modern approach. *Prentice Hall Upper Saddle River, NJ, USA: Rani, M., Nayak, R., & Vyas, OP (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems*, 90:33–48.
- Tan, T. Z., Quek, C., and Ng, G. S. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction 1. *Computational intelligence*, 23(2):236–261.
- Teweles, R. J. and Bradley, E. S. (1998). *The stock market*, volume 64. John Wiley & Sons.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of economic perspectives*, 28(2):3–28.

Viana, J. S. (2022). Mercado financeiro brasileiro: uma análise da percepção dos egressos do curso de secretariado executivo da universidade federal do ceará acerca de uma atuação nesse campo profissional.