



Modelo Preditivo para o IBOVESPA

Solução desenvolvida para o Tech Challenge - Fase 2: um modelo de machine learning capaz de prever se o IBOVESPA subirá ou descerá no dia seguinte, com acurácia mínima de 75%.

Introdução ao Desafio

O Desafio

Desenvolver um modelo de machine learning capaz de prever se o IBOVESPA iria subir ou descer no dia seguinte, com acurácia mínima de 75%.

O IBOVESPA

Principal indicador do mercado de ações brasileiro, composto pelas ações mais negociadas da B3, servindo como termômetro do comportamento do mercado.



Aquisição e Exploração dos Dados

Base de dados composta por registros históricos do IBOVESPA, abrangendo um período de 10 anos (2015-2025).

Variáveis utilizadas:

- Preços de abertura e fechamento
- Máxima e mínima diárias
- Volume negociado
- Variação percentual diária



Dados convertidos, normalizados e ordenados cronologicamente, respeitando a natureza sequencial da série temporal.

Tratamento da Série Temporal

Ordenação Cronológica

Dataset ordenado do mais antigo para o mais recente, garantindo que o modelo fosse treinado apenas com informações anteriores aos dados de teste.

Criação de Features

Desenvolvimento de features derivadas baseadas em janelas temporais, como LAGs, SMA, EMA, Momentum e RSI.

Visualização de Dados

Análise para entender o funcionamento e a importância de cada variável, a correlação entre elas e o impacto no modelo preditivo.

Engenharia de Atributos



LAG (defasagens)

Considera o comportamento do ativo nos dias anteriores, capturando efeitos de continuidade ou reversão de tendência.



Médias Móveis

SMA suaviza oscilações de curto prazo. EMA atribui maior peso aos dados recentes, sendo mais responsiva a mudanças rápidas.



RSI

Índice de Força Relativa mede a velocidade e mudança dos movimentos de preços, detectando condições de sobrecompra ou sobrevenda.

Estas features combinam informações de curto e médio prazo, aumentando a capacidade do modelo de identificar padrões preditivos com mais precisão.

Modelos Testados

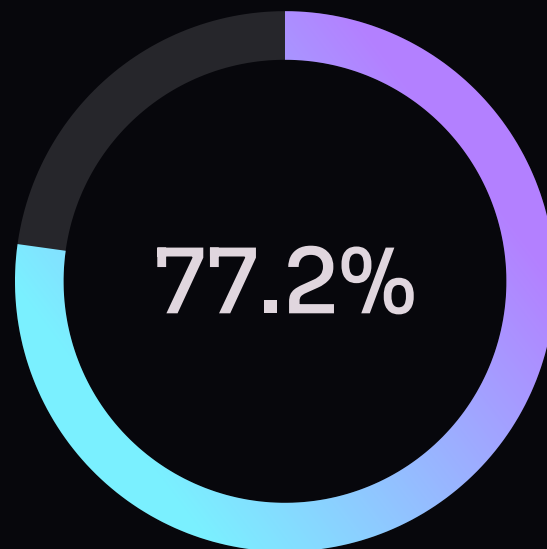
Regressão Logística

- Modelo estatístico simples
- Fácil implementação e rápida execução
- Ideal para problemas de classificação binária
- Base de comparação para modelos mais complexos

Random Forest

- Modelo de ensemble baseado em múltiplas árvores de decisão
- Robusto contra overfitting
- Capaz de lidar com relações não lineares
- Menos sensível a outliers
- Permite avaliar a importância das variáveis

Resultados Alcançados



Acurácia Final

Superando a meta estabelecida de 75%

Regressão Logística foi o modelo final escolhido, obtendo melhor desempenho ao lidar com os indicadores técnicos derivados. Após ajustes e ampliação da base de dados para 10 anos de histórico, superamos a meta estabelecida.



Conclusão

Desafio Superado

Construção de um modelo de machine learning capaz de lidar com dados financeiros reais e históricos, aplicando conceitos fundamentais de ciência de dados.

Aprendizado

O projeto proporcionou um aprendizado sólido sobre modelagem preditiva, design de features e o impacto de diferentes abordagens sobre a performance.

Importância

A experiência reforça a necessidade de respeitar o contexto dos dados, realizar escolhas embasadas e equilibrar desempenho com generalização.