

Construindo Modelo de Machine Learning

Grupo 150

FIAP + alura

POSTECH

Sumário:

1. Aquisição de dados e tratamento (Willian)
2. Estratégia de engenharia de atributos (Arthur)
3. Análise Exploratória (Jackson)
4. Definição do Modelo (Luis)
5. Justificativa e métricas de desempenho (Sofia)

Objetivo do Projeto

- Desenvolver um modelo preditivo para a tendência diária do IBOVESPA (alta/baixa).
- Meta de acurácia mínima: 75% no conjunto de teste (últimos 30 dias).

Aquisição dos Dados

- Fonte: Dados históricos do IBOVESPA, disponíveis publicamente no br.investing.com.
- Período: 18 de janeiro de 2008 a 18 de junho de 2025 (aproximadamente 17 anos de dados).
- Formato Inicial: Arquivo CSV, lido com Pandas.
- Estruturação: Data definida como índice do DataFrame para análise temporal

Tratamento dos Dados

- **Propósito:** Transformar dados brutos em formato adequado para Machine Learning.
- **Renomeação de Colunas:** Padronização para termos de mercado: 'close', 'open', 'high', 'low', 'volume', 'daily_return'.
- **Verificação de Duplicatas:** Nenhuma duplicata encontrada.
- **Tratamento de Nulos.**
- Ajuste de Tipos de Dados: 'daily_return' (variação percentual) convertida para float

Estratégia de engenharia de atributos

Indicadores de Mercado

- **RSI:** Mede o momentum da ação
- **Bandas de Bollinger:** Identifica sobrecompra e sobrevenda com base nos valores da ação
- **MACD:** Mudanças no momentum e tendência da ação
- **ADX:** Força da tendência (positiva ou negativa)
- **Z-Score:** Distância do valor médio da ação

Lags e janelas

- **Retorno da ação:** avaliar ritmo das variações (1, 2, 3 e 5 dias)
- **Consistência do momentum:** olhando se a variação dos últimos 3 dias é positiva
- **Mudanças na volatilidade:** comparando volatilidade recente com a histórica
- **Variação absoluta** (delta) com relação ao dia anterior
- **Média móvel exponencial** de 10 dias
- **Defasagem:** (lag) no valor de fechamento e *target*

Resultado Final

- Após avaliação das features utilizando **análise exploratória** e **testes estatísticos** chegamos a 21 features que foram utilizadas no modelo

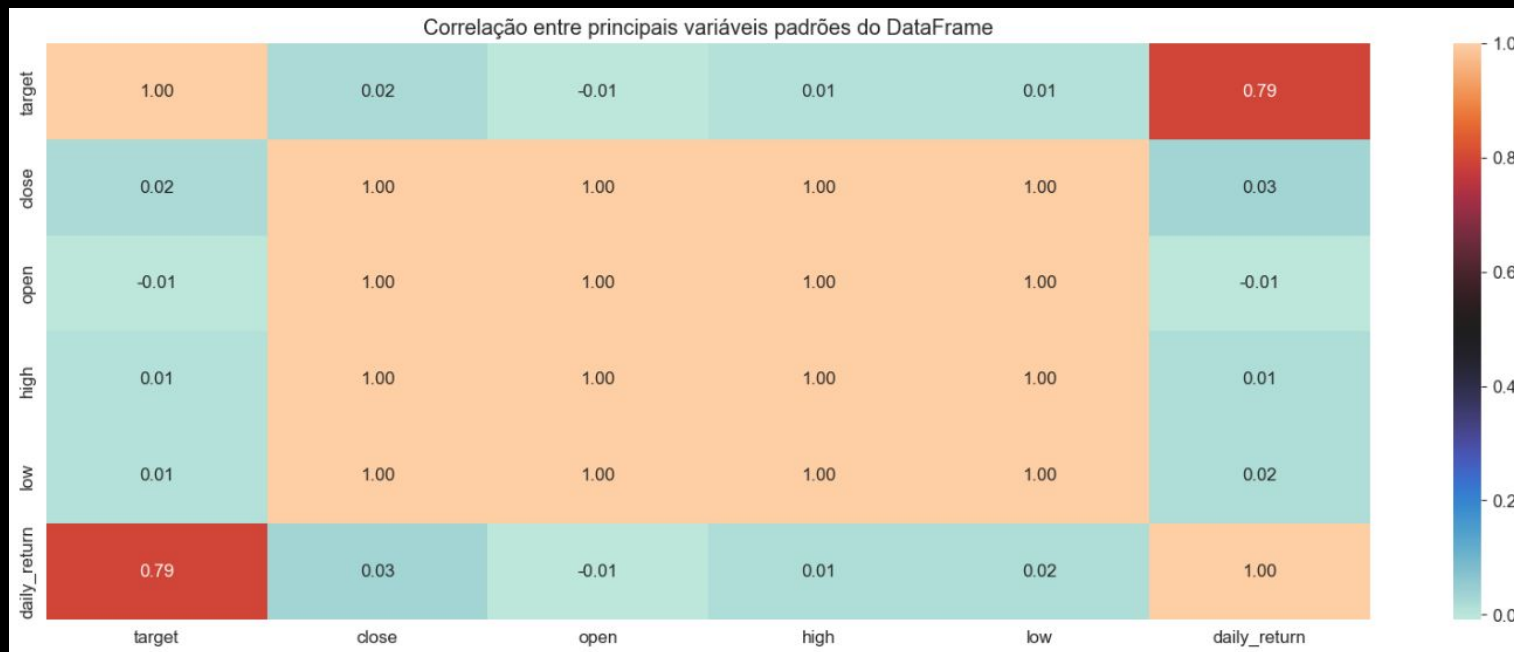
Análises exploratória iniciais

| | target | close | open | high |
|-------|---------------|-----------------|-----------------|-----------------|
| count | 2.593.000.000 | 2.593.000.000 | 2.593.000.000 | 2.593.000.000 |
| mean | 0.521404 | 93.888.234.092 | 93.854.124.180 | 94.691.504.049 |
| std | 0.499638 | 27.587.726.441 | 27.587.881.807 | 27.700.706.970 |
| min | 0.000000 | 37.497.000.000 | 37.501.000.000 | 38.031.000.000 |
| 25% | 0.000000 | 68.355.000.000 | 68.344.000.000 | 68.846.000.000 |
| 50% | 1.000.000 | 101.031.000.000 | 101.017.000.000 | 102.100.000.000 |
| 75% | 1.000.000 | 116.677.000.000 | 116.667.000.000 | 117.701.000.000 |
| max | 1.000.000 | 140.110.000.000 | 140.109.000.000 | 140.382.000.000 |

Análise histórica do índice IBOVESPA dos últimos 10 anos

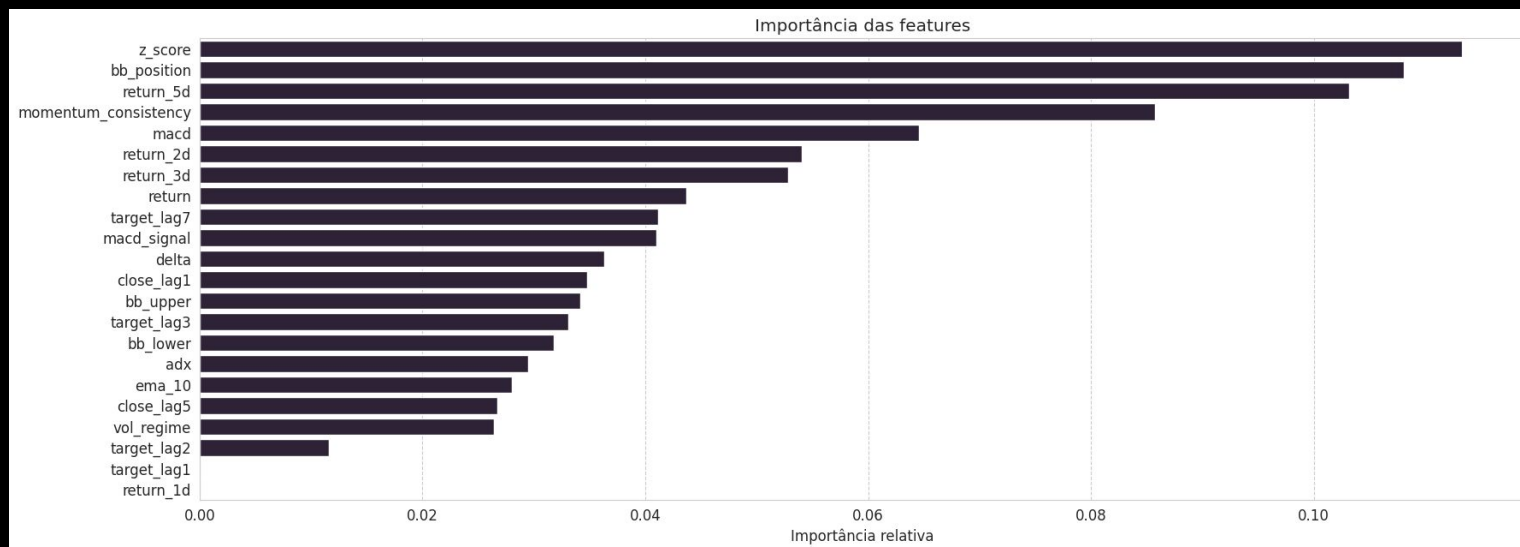


Análise de Correção da Features



Escolha do modelo

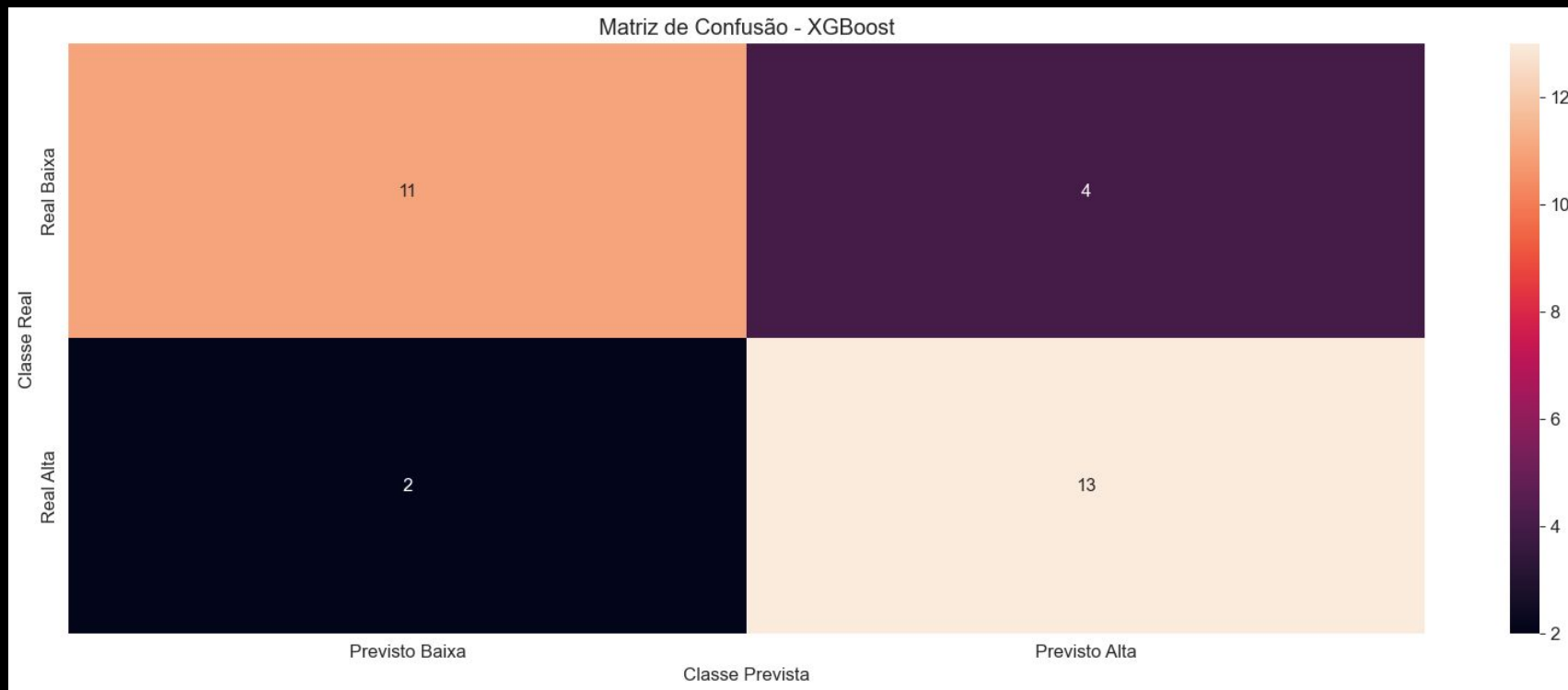
- Regressão x Classificação
- XGBoost, Random Forest
- Principais Features: Tendências de preço, Volatilidade e Momentum.



Avaliação do modelo

| <i>Modelo</i> | <i>Acurácia</i> | <i>Precisão</i> | <i>Recall</i> | <i>F1-Score</i> |
|----------------------------|-----------------|-----------------|---------------|-----------------|
| <i>XGBoost</i> | 80.00 % | 80.54 % | 80.00 % | 79.91 % |
| <i>Random Forest</i> | 80.00 % | 80.54 % | 80.00 % | 79.91 % |
| <i>Logistic Regression</i> | 76.67 % | 77.78 % | 76.67 % | 76.43 % |
| <i>SVM</i> | 70.00 % | 70.09 % | 70.00 % | 69.97 % |
| <i>Decision Tree</i> | 56.67 % | 56.70 % | 56.67 % | 56.62 % |

Avaliação do modelo



Melhor acurácia na validação cruzada do XGBoost: **77.71%**

Relatório de classificação do XGBoost:

| | Precisão | Recall | F1-Score | Suporte |
|-----------------|----------|--------|-------------|---------|
| 0.0 (Baixa) | 0.86 | 0.80 | 0.83 | 15 |
| 1.0 (Alta) | 0.81 | 0.87 | 0.84 | 15 |
| acurácia | | | 0.83 | 30 |
| média macro | 0.83 | 0.83 | 0.83 | 30 |
| média ponderada | 0.83 | 0.83 | 0.83 | 30 |

OBRIGADO