

Projeto: Previsão Direcional do IBOVESPA (D+1)

- Objetivo: prever se o fechamento do IBOVESPA no próximo pregão (D+1) será maior (\uparrow) ou menor (\downarrow) que o fechamento de hoje.
- Meta: acurácia direcional $\geq 75\%$ nos últimos 30 pregões (janela de teste definida pelo desafio).

Dados de Entrada

- Fonte: histórico diário do IBOVESPA (Investing.com).
- Colunas: Data, Abertura, Máxima, Mínima, Último (fechamento), Var% (variação diária), Vol. (volume).
- Período considerado: 2022-10-11 até 2025-10-23.
- 959 linhas, 0 valores nulos nas colunas originais.

Limpeza e Pré-Processamento

- Conversão de Data (string dd.mm.yyyy → datetime ordenável).
- Conversão de Vol.: '3,5M' → 3.500.000; '1,2B' → 1.200.000.000.
- Conversão de Var%: '1,23%' → 0.0123.
- Criação de colunas de calendário: ano, mês, dia da semana, AnoMes.
- Observação: volume apresenta desvio padrão muito alto → indício de liquidez irregular.

Engenharia de Atributos

- Retornos de preço: ret_1 (1 dia), ret_5d, ret_10d.
- Tendência / sobre-extensão: médias móveis e EMAs (5,10,20,50,...), spreads (preço - média), relação preço/MM.
- Volatilidade: largura das Bandas de Bollinger (20 dias).
- Liquidez: médias móveis de volume em múltiplas janelas.
- Momentum/força: RSI(14).
- Target = fechamento de amanhã (shift -1); dados sem NaN após limpeza final.

Preparação para Modelagem

- Série ordenada temporalmente (sem embaralhar).
- Split imposto pelo desafio: treino = histórico até antes dos últimos 30 pregões; teste = últimos 30 pregões.
- Cada linha usa apenas informação disponível naquele dia (sem vazamento de futuro).
- Esse teste final simula operação real: prever $D+1$ usando apenas D .

Seleção de Features

- Pipeline:
 - • StandardScaler
 - • SelectKBest(f_regression)
 - • Ridge
 - • Validação TimeSeriesSplit (5 divisões no tempo)
- GridSearchCV escolhe k (nº de features) e alpha (Ridge).
- Features finais selecionadas:
 - ['Abertura','Máxima','Mínima','mm_5','ema_5','mm_10','ema_10','mm_20','ema_20'].

Modelo SARIMAX

- SARIMAX = ARIMA com variáveis exógenas.
- Parâmetros p, d, q via ACF/PACF e estacionariedade ($d=1$).
- Sazonalidade via `auto_arima` ($m=5$ ~ padrão semanal de pregão).
- Exógenas = features selecionadas (tendência/nível de preço).
- Desempenho no teste (~30 dias):
 - • MAE ≈ 0.7391
 - • MSE ≈ 0.8189
 - • MAPE ≈ 0.51 %

Modelo Prophet

- Prophet com sazonalidade diária, semanal e anual.
- Mesmas regressoras externas adicionadas como regressoras.
- Tuning via cross_validation + performance_metrics:
 - • changepoint_prior_scale
 - • seasonality_prior_scale
 - • seasonality_mode
 - • changepoint_range
- Desempenho no teste após tuning:
 - • MAE \approx 0.4384
 - • MSE \approx 0.4583
 - • MAPE \approx 0.30 %

Métrica de Direção (Sobe/Cai)

- Convertendo previsão numérica em classe \uparrow/\downarrow :
- • Sobe se $\text{Preço_previsto}(D+1) > \text{Preço_atual}(D)$.
- SARIMAX (janela de teste):
- • Accuracy $\approx 93.1\%$
- • Balanced Accuracy $\approx 93.8\%$
- Prophet simples (janela de teste):
- • Accuracy $\approx 86.2\%$
- • Balanced Accuracy $\approx 86.8\%$
- Meta do desafio ($\geq 75\%$) foi atendida.

Conclusões

- Pipeline completo: coleta → limpeza → features → seleção → modelagem → validação temporal.
- SARIMAX apresentou melhor acurácia direcional e boas métricas de erro.
- Prophet mostrou erro absoluto menor (MAE, MAPE) ao prever preço.
- Próximos passos:
 - Simular walk-forward diário (refit contínuo).
 - Analisar robustez em períodos de alta volatilidade.
 - Evoluir para um classificador direto de direção (↑/↓).