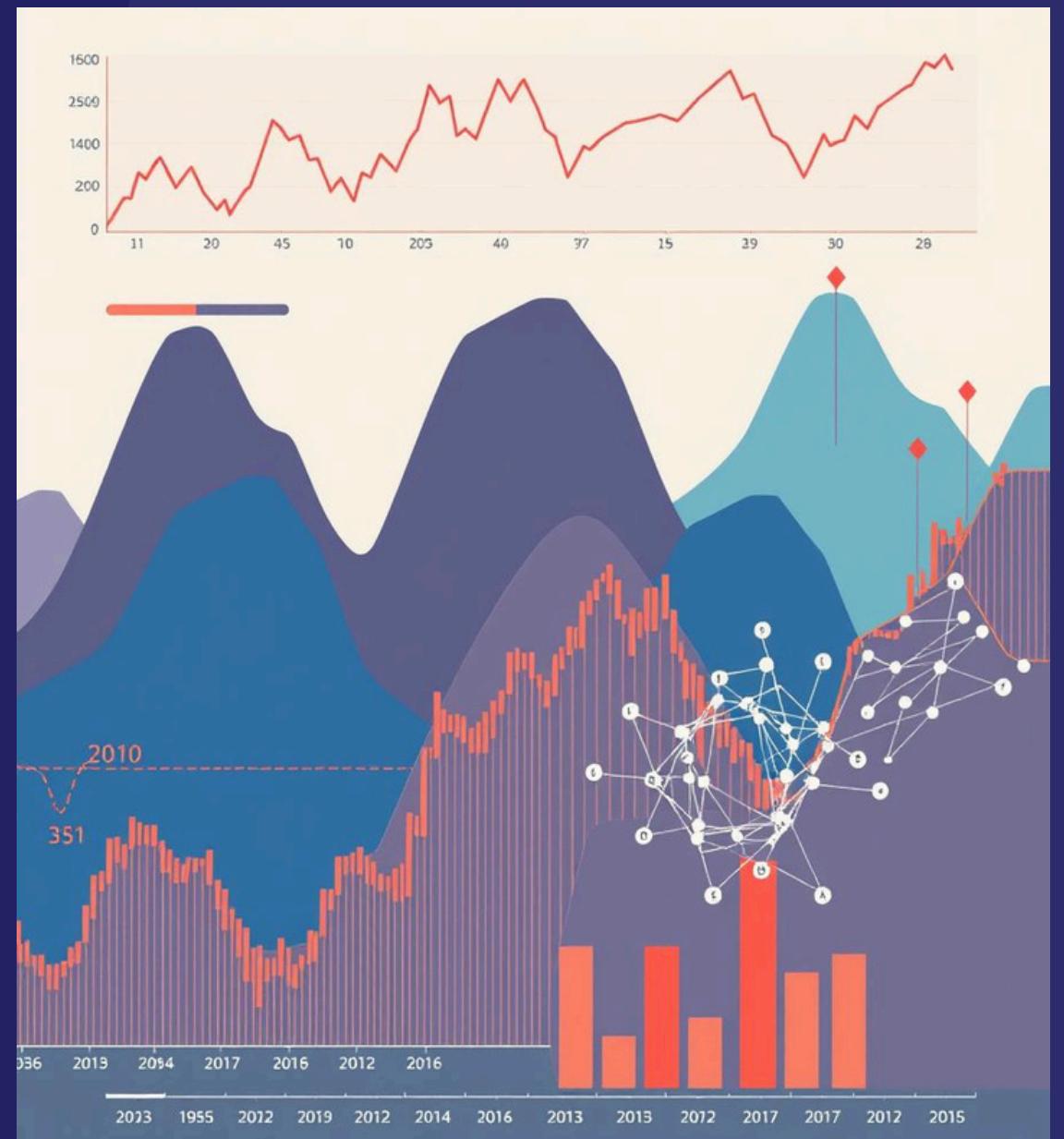
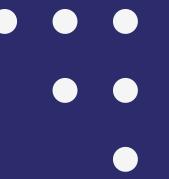


# Predição Automática de Indicativos Financeiros para **Bolsa de Valores** Considerando o Aspecto Temporal



Projeto Final de Curso 1 - Universidade Federal de Catalão  
Rafael da Silva Melo  
Orientador: Prof. Dr. Márcio de Souza Dias



# Sumário

- Contextualização
- Motivação
- Séries temporais
- O que é um modelo de previsão
- Modelos clássicos de previsão financeira
- Onde esse trabalho se encontra
- Objetivos gerais e Objetivos específicos
- Delimitações e fora do escopo
- Dados e preparação
- Arquitetura dos modelos
- Métricas e Testes
- Cronograma PFC2
- Referências
- Dúvidas



# Contextualização

Com os avanços da tecnologia, a complexidade e a volatilidade dos ativos negociados em bolsa de valores aumentaram significativamente.



## Bolsas de valores

Infraestruturas de mercado onde compradores e vendedores negociam valores mobiliários sob regras claras e supervisão, promovendo formação de preços, liquidez e alocação eficiente de recursos.



## B3 - Brasil, Bolsa, Balcão

Integra ambientes de negociação, registro, compensação e liquidação, operando sob marcos legais como a Lei do Mercado de Valores Mobiliários e a Lei das Sociedades por Ações.



# Contextualização

Embora o mercado de ações não seja teoricamente previsível, muitos investidores identificam padrões que motivam a busca por estratégias de previsão.



## Desafio Intradiário

No horizonte intradiário, emergem problemas de microestrutura — periodicidades sistemáticas, espaçamento irregular entre negócios e impacto de ordens — que condicionam a dinâmica de preços.

- Poucas ordens
- BID e ASK

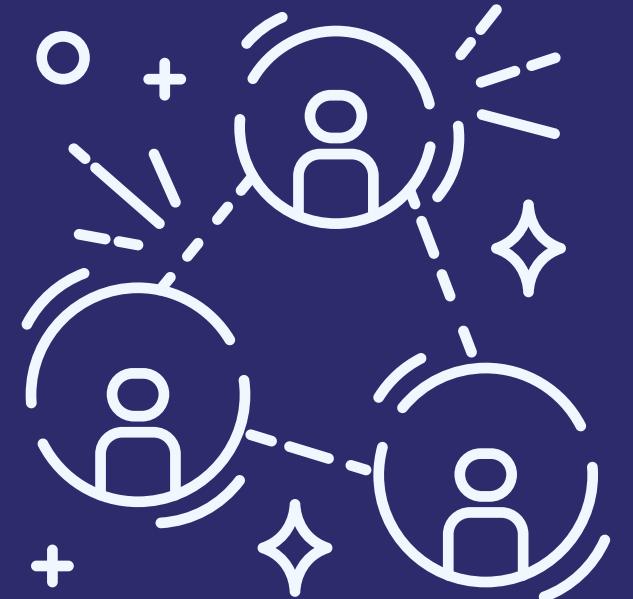
# Trabalhos correlatos

Vaswani et al. (2017)  
Wu et al. (2020)  
Lim, Zohren & Roberts (2021)  
Transformers

Fjellström (2022)  
Kabir et al. (2025)  
Redes neurais profundas



- Rotela Jr. et al. (2014) — ARIMA aplicado ao Ibovespa
- Box et al. (2015) — ARIMA/Box-Jenkins (Time Series Analysis: Forecasting and Control)
- Vaswani et al. (2017) — Transformer base (Attention Is All You Need)
- Wu et al. (2020) — Transformer aplicado a séries temporais
- Ding & Qin (2020) — LSTM para múltiplos preços (abertura, máxima, mínima)
- Lim, Zohren & Roberts (2021) — TFT (Temporal Fusion Transformer) para séries multivariadas
- Souza et al. (2022) — híbrido (LSTM + SVM + RF)
- Fjellström (2022) — ensembles de LSTM (maior estabilidade)
- Nascimento, Santos & Ferreira (2023) — LSTM vs. ARIMA/Prophet (LSTM superior)
- Kabir et al. (2025) — híbrido (LSTM + Transformer + MLP)

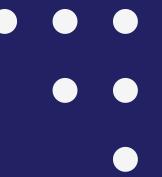


# Motivação



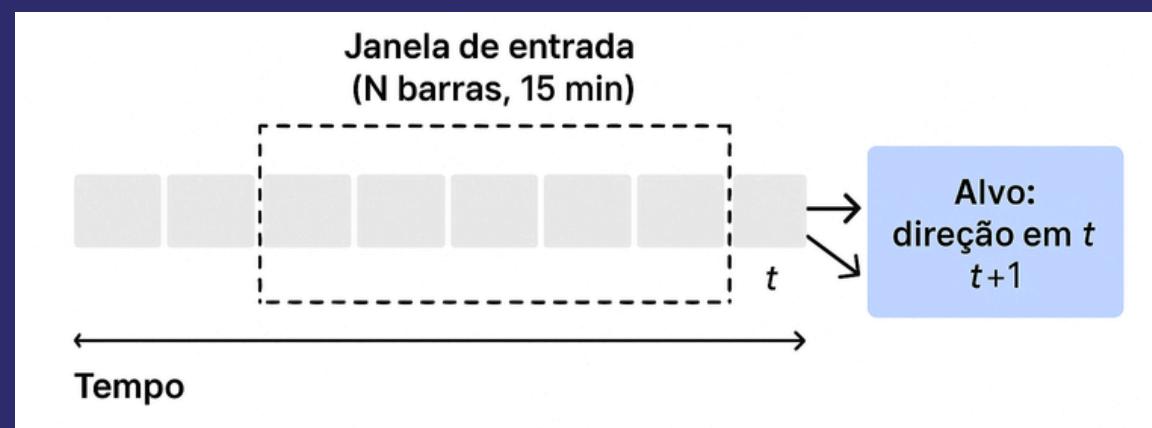
- Contexto: B3 apresenta alta volatilidade intradiária e efeitos de microestrutura que aumentam a incerteza preditiva.
- Lacuna: Modelos estáticos degradam com o tempo e não se adaptam a novos regimes.
- Proposta: Avaliar um modelo híbrido temporal para gerar indicativos financeiros mais robustos.
- Impacto esperado: Apoio à tomada de decisão em cenários voláteis e gestão de risco.





# Pergunta central do estudo

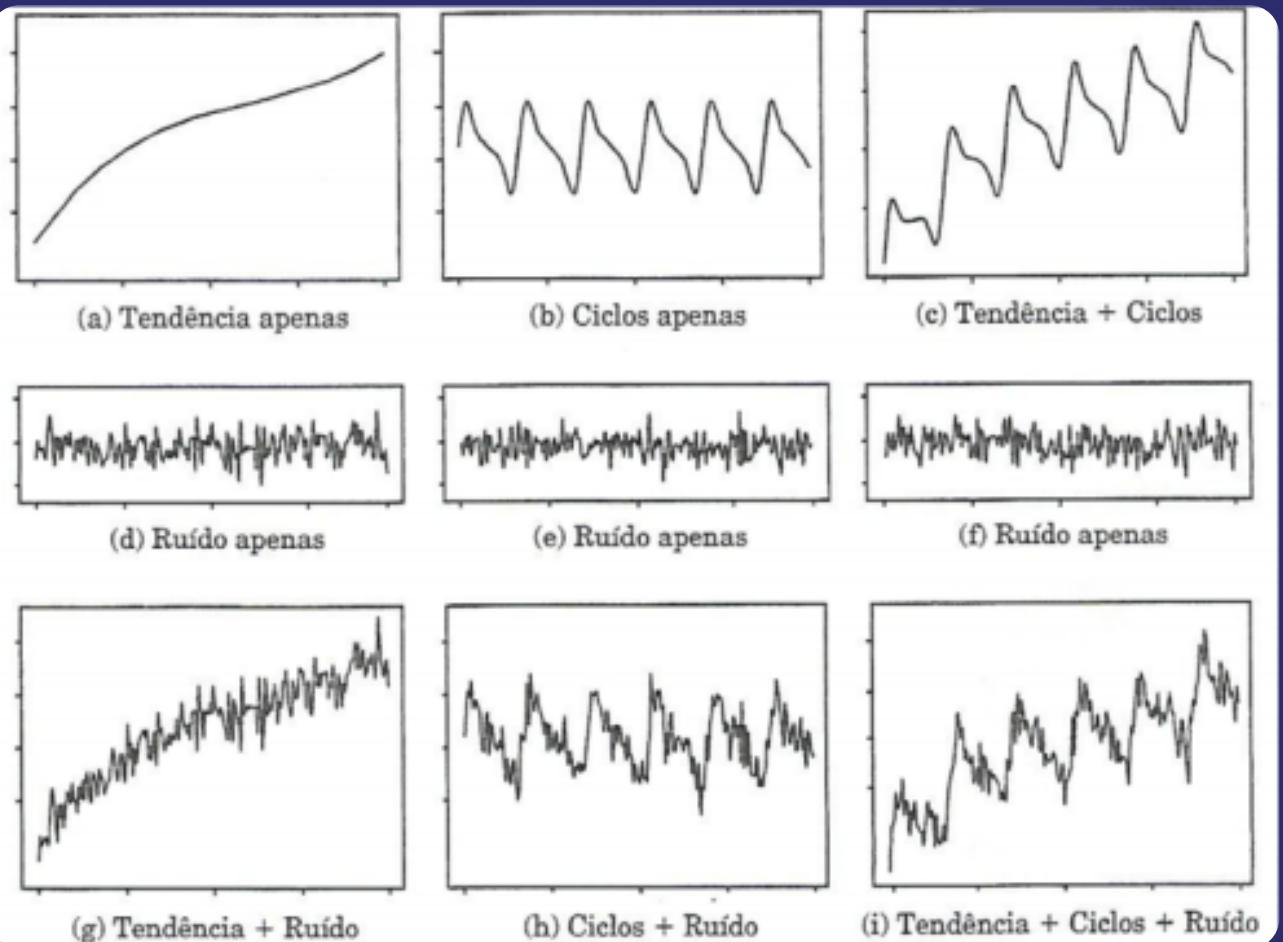
Em que condições (janelas, regimes de volatilidade e custo) um modelo CNN+LSTM supera comparadores clássicos (Naive, ARIMA, Prophet) e variantes com atenção na tarefa de previsão direcional intradiária (+1 barra/15 min) em ações líquidas da B3 (2020-2025), sob avaliação walk-forward e métricas pós-custos?



# Séries temporais

Segundo Box e Jenkins e amplamente discutido em Hyndman e Athanasopoulos, as séries temporais financeiras apresentam componentes como:

- Estacionaridade → retornos e janelas rolantes
  - Não exige
- Heterocedasticidade → normalizar apenas no treino de cada janela
  - CNN períodos menores
  - LSTM períodos maiores
- Microestrutura → ruídos que geram falsas reversões
  - Poucas ordens
  - BID e ASK



- Tendência
- Sazonalidade
- Ciclos
- Ruído aleatórios

# O que é um modelo de previsão?

- Conjunto de regras matemáticas/estatísticas que aprendem padrões em dados históricos.
- Entrada: séries temporais (preços, volumes, indicadores técnicos).
- Saída: projeção futura (preço, tendência, sinal de compra/venda).
- Objetivo: reduzir incerteza e apoiar a tomada de decisão.

- ARIMA
- Árvores de Decisão / Random Forest / XGBoost
- RNN → LSTM



# Modelos Clássicos de Previsão Financeira

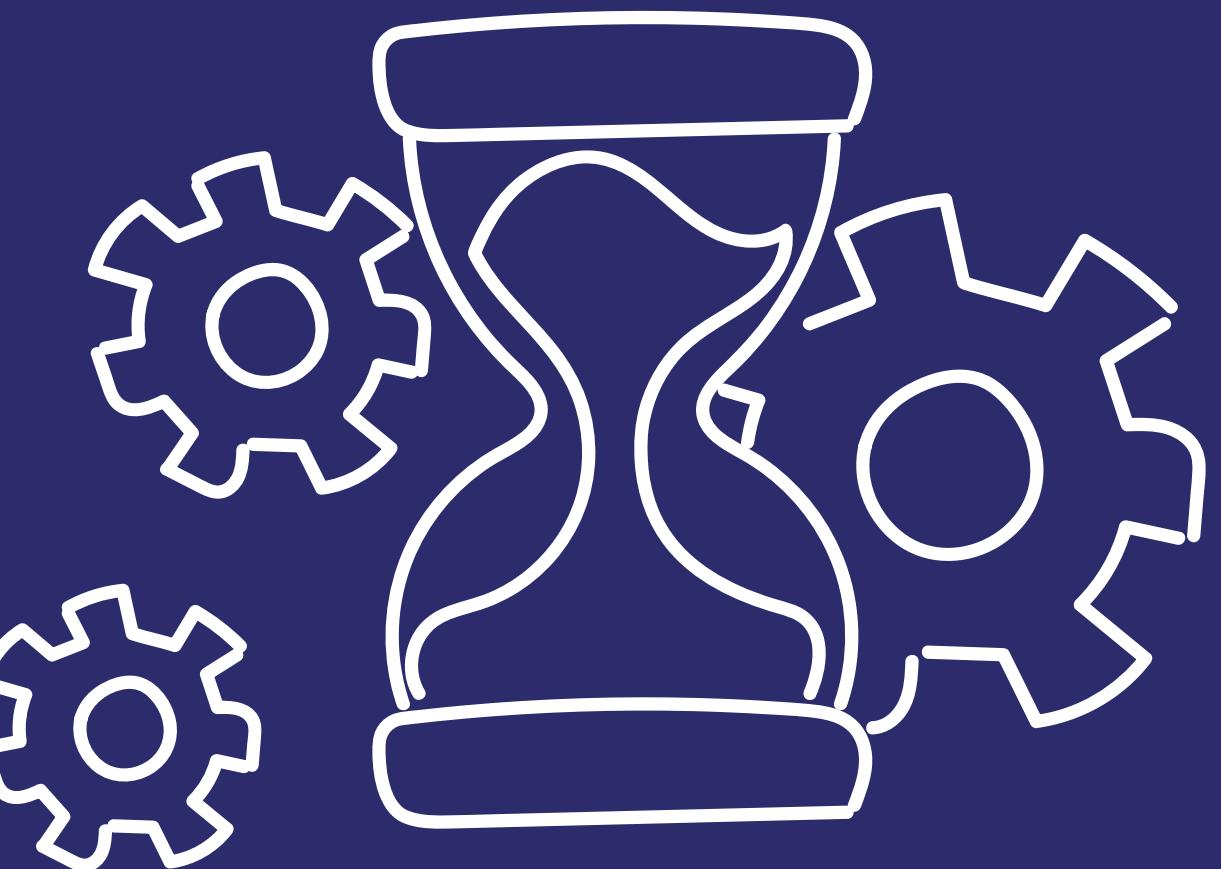
- ARIMA e modelos autorregressivos: capturam padrões lineares de curto prazo.
- Vantagem: simplicidade e boa interpretabilidade.
- Limitação: assumem linearidade e estacionariedade.
- Exemplo aplicado: em estudos da B3, podem capturar uma tendência curta, mas falham em prever reversões súbitas.



# Avanços sobre os modelos tradicionais

- Redes neurais feedforward (MLPs): exploraram não linearidades, mas sem memória temporal.
- Métodos baseados em árvore (Random Forest, XGBoost): boa performance em atributos, mas ignoram sequência.

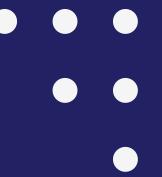
Esse modelo foi muito importante para os primeiros estudos da bolsa, porque eram simples e funcionavam em séries mais estáveis. Mas, quando aplicados em dados da B3 intradiária, eles se mostram limitados, pois não conseguem lidar bem com mudanças bruscas de comportamento.



# Onde este trabalho se posiciona

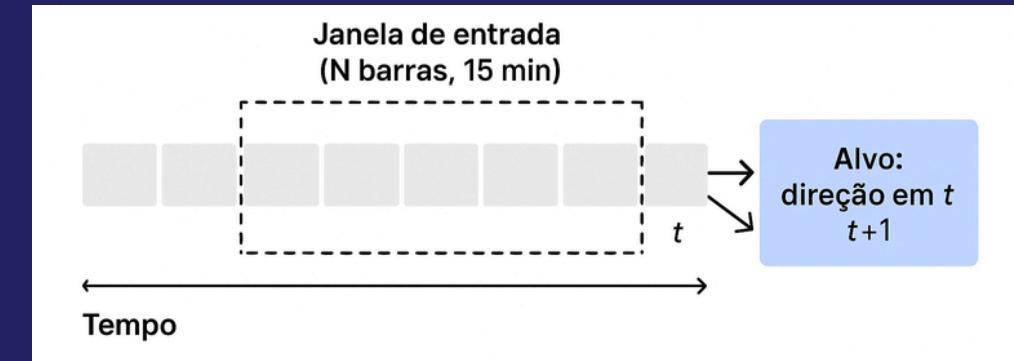
- Explora modelos temporais híbridos (CNN+LSTM) para prever indicativos financeiros.
- Vai além de modelos estatísticos e ML simples.
- Busca superar: falta de memória, validação incorreta, pouca robustez.
- Contribui com pipeline transparente e reproduzível.





# Objetivos

Avaliar a viabilidade e o desempenho de um híbrido CNN+LSTM para prever a direção +1 barra (15m) em ações líquidas da B3, sob walk-forward + embargo, comparado a baselines.



## Objetivos específicos

- Pipeline: dados → features → treino → avaliação, sem vazamento;
- Comparação justa: Naive/Drift, ARIMA, Prophet, LSTM puro.
- Métricas probabilísticas e de decisão: Brier/Log-Loss, calibração (ECE). Balanced Accuracy , F1-score, MCC de auxílio.
- Backtests pós-custos: regra simples de trading, custos/slippage, reportando retorno, Sharpe, max drawdown, turnover.
- Robustez e significância: Diebold–Mariano, bootstrap em blocos; sensibilidade a janelas, horizontes e custos (e regimes de volatilidade).
- Reprodutibilidade: seeds fixas, manifesto de dependências, scripts únicos de execução e registro de ameaças à validade.

# Delimitações



## Ativos

Ações líquidas da B3.

Exemplos: PETR4, VALE3 e ITUB4



## Frequência

15 minutos.

Com análises de sensibilidade em 5 e  
30 minutos.

## Horizonte

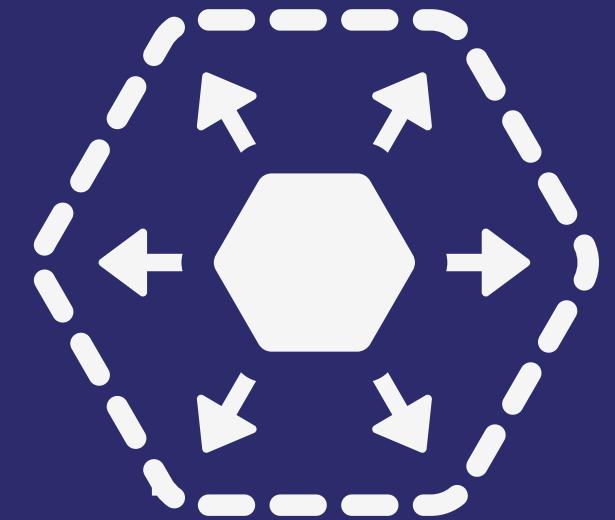
+1 barra (próxima vela)

## Validação

Walk-forward + embargo temporal

## Período

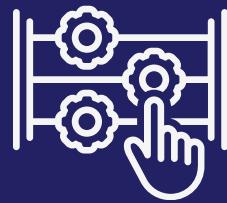
Jan/2020 - Jul/2025



# Fora do escopo

- Previsões de alta frequência em milissegundos
- Criptomoedas ou mercados internacionais.
- Predição de preço absoluto ou lucro garantido.

# Dados e preparação



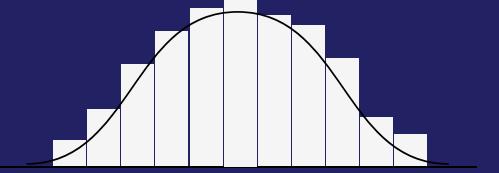
## Ajustes

Grupamentos  
Horários do pregão  
Barras com registros anômalos



## Dados

B3 e seus fornecedores



## Normalização

Min-Max para cada etapa do walk-forward  
Retorno logarítmico  
Indicadores técnicos MME 9/21/50,  
Bandas de Bollinger  
RSI



# Arquitetura dos modelos

Ordem crescente de complexidade, viabilizando comparações consistentes



## Naive / Drift

Repetir a direção da última barra



## ARIMA

Média Móvel Autorregressiva Integrada



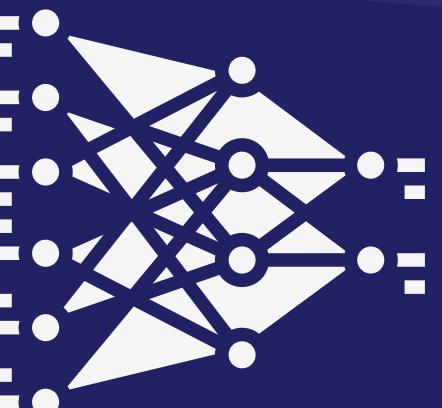
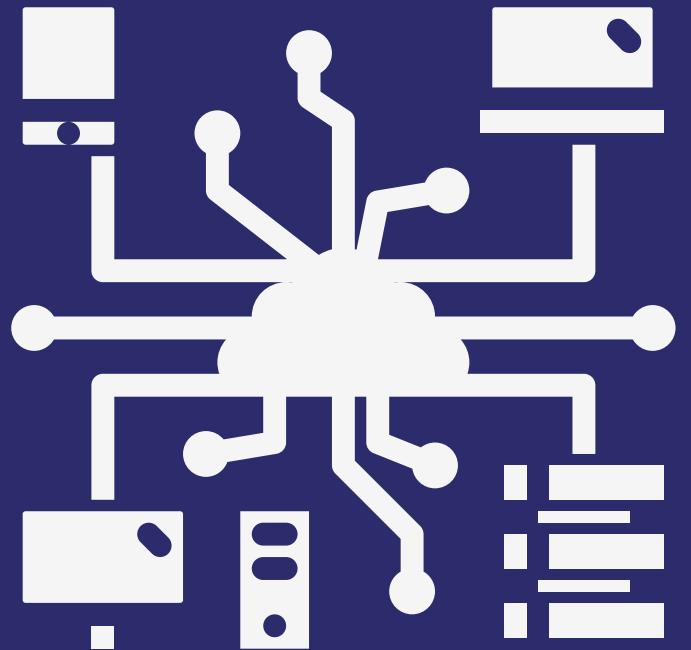
## Prophet

Sazonalidades diárias/semanais.



## LSTM puro

Dependências de longo prazo



## Proposto

CNN + LSTM

CNN → padrões locais na janela temporal

LSTM → dependências de maior alcance

Saída → Síntese de probabilidade, tentando equilibrar movimentos bruscos e persistentes

# Métricas

- Acurácia Direcional → Comprar ou vender
- Qualidade probabilística → Garante que seja confiável, não só classificações
- Métricas de trading:
  - Retorno líquido
  - Sharpe
  - Max Drawdown



# Testes

- Backtest
  - Long e short
  - Custos de transação
- Robustez
  - Diebold-Mariano entre os baselines
- Volatilidade (calmaria vs choque)

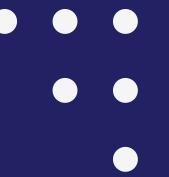


# Cronograma PFC2

| Etapas  | Set | Out | Nov | Dez | Jan | Entregáveis / Marcos  |
|---|-----|-----|-----|-----|-----|---|
| Aquisição e auditoria de dados; ajustes integridade temporal  |     |     |     |     |     | Dataset auditado; checklist de integridade; universo de ativos definido.    |
| Pré-processamento e engenharia de atributos   |     |     |     |     |     |   |
| Particionamento e desenho experimental  |     |     |     |     |     | Folds definidos; protocolo de avaliação documentado; ambiente reproduzível. |
| Baselines: Naive/Drift, ARIMA, Prophet (execução nos 1-2)   |     |     |     |     |     | Relatório curto com desempenho por bloco.                                   |
| Modelo proposto (CNN-LSTM) e LSTM puro.   |     |     |     |     |     | train.py / evaluate.py funcionais;  |
| Avaliação preditiva: Brier, Log-Loss, Balanced  |     |     |     |     |     | Tabela consolidada por modelo/fold; curvas de confiabilidade.               |
| Backtests (long-only e long/short) com custos   |     |     |     |     |     | backtest.py; relatório operacional (retorno, Sharpe, drawdown).             |
| Robustez e significância: Diebold-Mariano; regimes de volatilidade; sensibilidade (5/30m)             |     |     |     |     |     | Nota técnica.   |
| Reprodutibilidade: manifesto de dependências; organização de repositório/artefatos; README executável |     |     |     |     |     | Repositório final organizado; scripts e experimentos replicáveis.           |
| Redação de Resultados, Discussão  |     |     |     |     |     | Capítulos redigidos com figuras e tabelas; limitações descritas.            |



# Dúvidas



# Referências

- <https://www.suno.com.br/artigos/bolsa-de-valores-/>
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M.  
Time Series Analysis: Forecasting and Control. 5. ed. [S.l.]:  
Wiley, 2015. ISBN 9781118675021. Citado nas páginas 6, 9,  
25, 31 e 37.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. Forecasting:  
Principles and Practice. 2. ed. OTexts, 2018. 2nd ed.  
Disponível em: . Citado nas páginas 6, 9, 19, 31 e 37.