

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

## Roteiro de apresentação (falas completas) — duração sugerida: 35–40 minutos

**Como usar:** leia as falas em voz alta para ensaiar; adapte termos para o seu jeito de falar. Entre colchetes, há **didascálias** de tempo, ênfase e transições. Onde aparecer (**slide: ...**) é o título aproximado do slide.

---

### 0) Abertura e alinhamento (2 min)

**Fala:** “Boa tarde, sou o Rafael da Silva Melo, aluno de Ciências da Computação na UFG/Regional Catalão. Hoje apresento meu PFC1, intitulado *Predição Automática de Indicativos Financeiros para Bolsa de Valores Considerando o Aspecto Temporal*, orientado pelo Prof. Dr. Márcio de Souza Dias. [olhar banca]

O objetivo desta apresentação é contextualizar o problema, justificar a proposta, detalhar **escopo, método e cronograma** do PFC2 e, por fim, discutir as contribuições esperadas.”

**Transição:** “Começo pelo pano de fundo do mercado e os desafios intradiários na B3.”

---

### 1) Sumário e objetivo da sessão (1 min) — (slide: Sumário)

**Fala:** “Seguirei esta ordem: Contextualização → Motivação → Pergunta central → Fundamentos → Modelos e posição do trabalho → Objetivos e delimitações → Dados e preparação → Arquitetura → Métricas → Testes → Cronograma.”

**Dica:** aponte rapidamente cada tópico no slide, sem ler.

---

### 2) Contextualização (4 min) — (slide: Contextualização / Desafio intradiário)

**Fala:** “Bolsas são infraestruturas que asseguram **formação de preços, liquidez e alocação eficiente**. No Brasil, a B3 integra negociação, registro, compensação e liquidação sob marcos legais específicos. [gesto]

No **intraday**, surgem efeitos de **microestrutura**: periodicidades ao longo do pregão, **espaçamentos irregulares entre negócios e impacto de ordens**. Isso torna os sinais mais ruidosos, especialmente pela presença do **spread bid-ask** e baixa profundidade em certos momentos. [pausa]

Esse cenário pede modelos **sensíveis ao tempo**, que considerem **volatilidade e liquidez** intradiárias, e **avaliem probabilidades**, não apenas acertos binários.”

**Transição:** “Com esse pano de fundo, por que este projeto é necessário agora?”

---

### 3) Motivação (2–3 min) — *(slide: Motivação)*

**Fala:** “Na B3, a **volatilidade intradiária** é elevada. **Modelos estáticos** degradam com o tempo; eles não se adaptam facilmente a **mudanças de regime**. Então proponho avaliar um **modelo híbrido temporal (CNN+LSTM)** para gerar **indicativos financeiros mais robustos** – com potencial de apoiar decisão em cenários voláteis e de contribuir para **gestão de risco**.”

**Ênfase:** “Aqui, ‘indicativos’ significa **probabilidades direcional +1 barra** (15 min) para movimentos de alta/baixa, não sinal de trading garantido.”

---

### 4) Pergunta central do estudo (2 min) — *(slide: Pergunta central)*

**Fala:** “A pergunta norteadora é: **Em que condições** — janelas, regimes de volatilidade e custo — um **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), **com avaliação walk-forward** e **métricas pós-custos**?”

**Nota de ênfase:** “O foco é **condicionar o desempenho** a **contextos de mercado** e **custo de transação**, e **medir qualidade probabilística**.”

---

### 5) Fundamentos de séries temporais (4–5 min) — *(slide: Séries temporais)*

**Fala:** “Séries financeiras combinam **tendência, sazonalidade, ciclos e ruído**. Em horizontes intradiários, a **estacionaridade** é tratada via **retornos** e **janelas rolantes**; já a **heterocedasticidade** requer normalizar apenas com dados de treino em cada janela, evitando vazamento. [mostrar bullets]

No design do modelo, **CNNs** capturam **padrões locais** (motivos de curto prazo), enquanto **LSTMs** lidam melhor com **dependências de maior alcance**. Por isso, **híbrido**.”

**Transição:** “Antes do proposto, reviso rapidamente os modelos de referência.”

---

### 6) Modelos clássicos e avanços (5 min) — *(slides: O que é um modelo?; Modelos clássicos; Avanços)*

**Fala:** “Um **modelo de previsão** aprende padrões históricos para projetar o futuro. Clássicos como **ARIMA** têm **simplicidade** e **interpretabilidade**, mas assumem linearidade/estacionaridade. Métodos como **Random Forest/XGBoost** funcionam bem com atributos, mas **ignoram sequência**.”

Na prática intradiária da B3, tais abordagens tendem a **falhar em reversões súbitas** e **mudanças de regime**. Daí a motivação por arquiteturas **temporais profundas**.”

**Transição:** “E onde este trabalho se posiciona?”

---

## 7) Posição do trabalho (2 min) — (slide: Onde este trabalho se posiciona)

**Fala:** “Posiciono o estudo em **modelos temporais híbridos (CNN+LSTM)**, indo além de estatística e ML tabular. Busco mitigar: **falta de memória, validação incorreta e baixa robustez**. Contribuo com **pipeline transparente e reprodutível**.”

---

## 8) Objetivo geral e específicos (3–4 min) — (slide: Objetivos / Objetivos específicos)

**Fala:** “**Objetivo geral:** avaliar a viabilidade e o desempenho do híbrido **CNN+LSTM** na **predição direcional +1 barra (15m)** em **ações líquidas da B3**, sob **walk-forward + embargo**, comparado a **baselines**.”

**Específicos:** [apontar] 1) Construir pipeline **dados → features → treino → avaliação**, sem vazamento; 2) Comparar com **Naive/Drift, ARIMA, Prophet, LSTM puro**; 3) Usar **métricas probabilísticas** (Brier, Log-Loss, **calibração ECE**) e de decisão (Balanced Accuracy, F1, MCC; apoio com AUC-PR); 4) **Backtests pós-custos:** regra simples, custos/slippage, **retorno, Sharpe, max drawdown, turnover**; 5) **Robustez:** **Diebold–Mariano**, bootstrap em blocos; sensibilidade a **janelas, horizontes** e **custos** sob **regimes de volatilidade**; 6) **Reprodutibilidade:** seeds, manifesto de dependências e scripts únicos; **ameaças à validade** documentadas.”

**Observação:** Reforce que o estudo **não** promete lucro nem *timing* perfeito; o foco é **predição probabilística e avaliação justa**.

---

## 9) Delimitações e fora do escopo (2 min) — (slide: Delimitações / Fora do escopo)

**Fala:** “Amostra: **PETR4, VALE3, ITUB4** como exemplos de alta liquidez; período **jan/2020–jul/2025**; frequência **15 minutos** (sensibilidades em **5/30m**); horizonte **+1 barra**.”

Validação **walk-forward** com **embargo temporal**. Fora do escopo: **HFT** em milissegundos; **cripto/mercados internacionais**; **preço absoluto** ou qualquer promessa de **lucro garantido**.”

---

## 10) Dados e preparação (4–5 min) — (slide: Dados e preparação)

**Fala:** “Fontes: **B3 e fornecedores**. Cuidados: **grupamentos, horários de pregão**, detecção de **barras anômalas** (ex.: leilões, *outliers*, quedas de feed), e **ajustes** (proventos, desdobramentos quando aplicável). ”

**Transformações:** retorno logarítmico; **normalização Min-Max por janela** (apenas treino) para evitar vazamento; **features técnicas** como **MME 9/21/50, Bandas de Bollinger, RSI**; além de **sazonalidades intra-dia/semanais**.

**Nota prática:** manter um **checklist de integridade temporal** e *drop* de sessões incompletas/feriados.”

**Transição:** “Com os dados prontos, como organizo os modelos?”

---

## 11) Arquitetura dos modelos (4–5 min) — (slide: Arquitetura dos modelos)

**Fala:** “Estabeleço uma **escada de complexidade**: [apontar] **Naive/Drift** → **ARIMA/Prophet** → **LSTM puro** → **CNN+LSTM** (proposto).

**Intuição:** CNN extrai **motivos locais**; LSTM modela **dependências longas**. A **camada final** usa **sigmoide** para a **probabilidade direcional**. Em paralelo, adiciono **sazonalidades** simples (diária/semanal) como *features*.”

**Comentário técnico:** mencionar **regularização**, **early stopping**, e **otimização bayesiana de hiperparâmetros** por bloco.

---

## 12) Métricas (3–4 min) — (slide: Métricas)

**Fala:** “Avalio três eixos: (i) **Direcional** (Balanced Accuracy/F1/MCC), (ii) **Probabilístico** (Brier, Log-Loss, calibração/ECE), e (iii) **Trading** (retorno líquido, Sharpe, máx. drawdown), sempre pós-custos.

**Fórmulas úteis:**

- **Brier score:**  $\text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2$
- **Log-loss:**  $-\frac{1}{N} \sum [y_i \log p_i + (1 - y_i) \log (1 - p_i)]$
- **Sharpe** (diário):  $\frac{\mathbb{E}[R - R_f]}{\sigma(R - R_f)}$
- **Balanced Accuracy:**  $\frac{1}{2}(\text{TPR} + \text{TNR})$ .

**Racional:** probabilidades bem calibradas **valem mais** do que *hit rate* isolado.”

---

## 13) Testes e desenho experimental (4–5 min) — (slide: Testes)

**Fala:** “Uso **walk-forward** por blocos temporais com **embargo** para evitar contaminação por adjacência. Para comparação, aplico **Diebold–Mariano**.

Para *trading*, **backtests long-only** e **long/short**, com **custos de transação** e **slippage**, reportando **retorno**, **Sharpe** e **drawdown**. Faço **análises de sensibilidade** a custos e **regimes de volatilidade** (calmaria vs. choque).”

**Transição:** “Por fim, mostro o **cronograma** do PFC2.”

---

## 14) Cronograma PFC2 (2–3 min) — (slide: Cronograma PFC2)

**Fala:** “De **set/2025 a jan/2026**: [apontar marcos] (1) Auditoria e preparação dos dados; (2) Particionamento e protocolo; (3) Baselines; (4) LSTM e CNN+LSTM; (5) Avaliação e calibração; (6) Backtests; (7) Robustez e nota técnica; (8) Reprodutibilidade e organização do repositório; (9) Redação de Resultados/Discussão.”

**Fecho:** “O objetivo é chegar ao PFC2 com **scripts reprodutíveis**, **tabelas consolidadas** e **limitações** claramente descritas.”

---

## 15) Considerações finais e convite a perguntas (2 min)

**Fala:** “Recapitulando: o trabalho define **escopo, hipótese e protocolo sensível ao tempo**, com foco em **probabilidades calibradas e utilidade prática**. Os próximos passos visam **comparações justas e robustez** sob custos e volatilidade.

Agradeço à banca e fico à disposição para perguntas.”

**Pausa:** (Respire, sorria e convide a banca a começar.)

---

## Notas do apresentador (cola rápida por slide)

- **Contexto:** B3 = integração de negociação/clearing; intraday tem microestrutura (spread, periodicidade, impacto de ordens).
  - **Motivação:** degradação de modelos estáticos; regimes de volatilidade; necessidade de calibração.
  - **Pergunta:** quando e por que CNN+LSTM supera baselines sob custos e volatilidade.
  - **Fundamentos:** tendência/sazonalidade/ciclos/ruído; estacionaridade via retornos; heterocedasticidade; CNN (local) + LSTM (global).
  - **Modelos:** ARIMA/Prophet vs. LSTM/CNN; limitações de métodos tabulares.
  - **Objetivos:** métricas probabilísticas, backtests pós-custos, robustez e reprodutibilidade.
  - **Delimitações:** PETR4/VALE3/ITUB4; 15m (+5/30m); 2020–2025; +1 barra; walk-forward + embargo.
  - **Dados:** integridade temporal; normalização por janela; indicadores técnicos; sazonalidades intradiárias.
  - **Arquitetura:** escada de complexidade; sigmoide na saída.
  - **Métricas:** Brier/Log-Loss/ECE; Balanced Acc., F1, MCC, AUC-PR; Sharpe/drawdown.
  - **Testes:** DM test; sensibilidade a custos/volatilidade; backtests long-only/long-short.
  - **Cronograma:** bloqueado por entregáveis; scripts funcionais.
- 

## Perguntas frequentes da banca (com respostas-curtas)

1) **Por que não usar só Transformer?** — *Transformers brilham em dependências longas, mas em intraday de alta frequência, CNN+LSTM pode ser mais eficiente e estável com janelas menores; além disso, comparo com variantes atencionais como parte dos baselines para testar essa hipótese.* 2) **Como evita vazamento de informação?** — *Split temporal estrito; normalização e seleção de hiperparâmetros apenas no treino; embargo temporal entre janelas para reduzir contaminação.* 3) **Por que Balanced Accuracy, MCC e ECE?** — *Para lidar com desequilíbrio de classes e avaliar calibração. Acerto bruto pode ser enganoso; ECE e Brier capturam qualidade probabilística.* 4) **E se custos aumentarem?** — *Análise de sensibilidade em custos/slippage; reporto métricas de trading pós-custos, com turnover e stress tests.\** 5) **Generaliza para outros ativos?** — *\*O protocolo é transferível; os resultados, porém, são condicionais à liquidez e microestrutura de cada papel.\** 6) **Qual a utilidade prática?** — *\*Melhor gestão de risco e priorização de cenários via probabilidades calibradas; potenciais ganhos em decisão tática.\** 7) **Limitações?** — *\*Choques extremos, drift de distribuição e latência de execução; mitigação via re-treino periódico, ensembling e monitoring de calibração.\** 8) **Como garantir reprodutibilidade?** — *\*Seeds fixas, manifesto de dependências, scripts únicos, logs de hiperparâmetros e README executável\*.*

---

## PALAVRAS-CHAVE E DEFINIÇÕES (para consulta rápida durante a fala)

- **Microestrutura:** efeitos do **livro de ofertas** (profundidade, **spread bid-ask**), periodicidades do pregão e impacto de ordens.
- **Walk-forward:** esquema de treino→teste que **avança no tempo**; cada bloco usa janelas anteriores como treino e testa no futuro; evita viés.
- **Embargo temporal:** intervalo entre janelas de treino e teste para **reduzir contaminação** por dependência temporal.
- **Brier Score:** erro quadrático médio das **probabilidades** previstas.
- **Log-Loss:** penaliza previsões **mal calibradas**; usa log-probabilidades.
- **ECE (Expected Calibration Error):** distância média entre **probabilidade prevista** e **frequência observada** por *bins*.
- **Balanced Accuracy:** média de TPR e TNR; robusto a **classes desbalanceadas**.
- **MCC:** correlação preditiva; útil quando há desequilíbrio.
- **Sharpe:** retorno ajustado ao risco (desvio-padrão).
- **Slippage:** diferença entre preço teórico e **execução**; aumenta com **volatilidade** e **baixa liquidez**.
- **Turnover:** rotatividade de posições; proxy de custo.
- **Liquidez intradiária:** volume por minuto, spread, profundidade; varia ao longo do pregão.
- **Heterocedasticidade:** variância dos retornos **não constante**; comum em intraday.
- **Naive/Drift:** usa a última direção/preço como *baseline*.
- **ARIMA/Prophet:** modelos **decompositivos/autorregressivos**; boa linha de base, porém limitados sob não linearidades.
- **LSTM:** RNN com memória de longo prazo; boa em dependências temporais.
- **CNN:** extrai **padrões locais** em janelas temporais com **convoluções 1D**.

---

## Gatilhos de fala (expressões curtas para usar ao vivo)

- “O ponto aqui é **probabilidade calibrada**, não palpite.”
- “Eu **condiciono** o desempenho por **regime de volatilidade** e **custo**.”
- “**Sem vazamento**: normalizo e *tune* só no treino.”
- “Penso na arquitetura como **CNN (local) + LSTM (longo alcance)**.”
- “Backtests **pós-custos**; resumo por **retorno, Sharpe, drawdown**.”
- “**Diebold-Mariano** testa se diferenças de erro são estatisticamente significativas.”

---

## Checklist de coerência dos slides (correções sugeridas)

1) **Ortografia:** corrigir “**Balanced Accurace**” → **Balanced Accuracy**; “**Acurácial Direcional**” → **Acurácia Direcional**; padronizar **ARIMA** (evitar “Arima”). 2) **Consistência:** onde diz “Naive / Drift Arima”, separar **Naive/Drift** de **ARIMA**; “Sazonalidades diárias/semanais” fica melhor no slide de **Dados/Features** (ou duplicado em ambos com objetivo distinto: *feature engineering* vs. ajuste de arquitetura). 3) **Escopo:** no slide de **Dados e preparação**, acrescentar nota sobre **feriados/sessões parciais, leilões de abertura/fechamento** e **tratamento de outliers** (*feeds, bad ticks*). 4) **Metodologia:** no slide de **Métricas**, incluir **MCC** e **AUC-PR** (se já citados entre objetivos) para coerência. 5) **Reprodutibilidade:** garantir um *slide* curto com **seeds, manifesto, scripts** (pode ser no final, antes do cronograma), ou inserir bullets no

slide de **Testes**. 6) **Aviso ético**: incluir **disclaimer** “Estudo acadêmico, não recomendação de investimento”.

---

## Planejamento de tempo detalhado (35–40 min)

- Abertura (2) + Sumário (1) = **3 min**
  - Contexto (4) + Motivação (3) = **7 min** (acum. 10)
  - Pergunta (2) + Fundamentos (5) = **7 min** (acum. 17)
  - Modelos/Posição (5) = **5 min** (acum. 22)
  - Objetivos/Delimitações (4) = **4 min** (acum. 26)
  - Dados/Preparação (5) = **5 min** (acum. 31)
  - Arquitetura (4) + Métricas (4) = **8 min** (acum. 39)
  - Testes (4) + Cronograma (3) = **7 min** (acum. 46)
  - Perguntas ( $\geq 4$ ) → ajuste natural para 35–40 min conforme ritmo.
- 

## Referências (para citar oralmente durante a apresentação)

- Box & Jenkins (modelagem ARIMA; decomposição).
  - Hyndman & Athanasopoulos (fundamentos de previsão).
  - Taylor & Letham (Prophet).
  - Diebold & Mariano (teste de acurácia preditiva).
  - Sharpe (índice de Sharpe).
  - Andersen & Bollerslev; Engle & Russell; Hasbrouck; Amihud (microestrutura: volatilidade, duração, profundidade, iliquidez).
- 

## references.bib (anexe ao seu trabalho)

```
@book{BoxJenkins2015,  
  author = {Box, George E. P. and Jenkins, Gwilym M. and Reinsel, Gregory C.  
and Ljung, Greta M.},  
  title = {Time Series Analysis: Forecasting and Control},  
  edition = {5},  
  publisher = {Wiley},  
  year = {2015},  
  isbn = {9781118675021}  
}  
  
@book{HyndmanAthanasopoulos2018,  
  author = {Hyndman, Rob J. and Athanasopoulos, George},  
  title = {Forecasting: Principles and Practice},  
  edition = {2},  
  publisher = {OTexts},  
  year = {2018},  
  url = {https://otexts.com/fpp2/}  
}
```

```

@article{TaylorLetham2018,
  author = {Taylor, Sean J. and Letham, Benjamin},
  title = {Forecasting at Scale},
  journal = {The American Statistician},
  year = {2018},
  volume = {72},
  number = {1},
  pages = {37--45},
  doi = {10.1080/00031305.2017.1380080}
}

@article{DieboldMariano1995,
  author = {Diebold, Francis X. and Mariano, Roberto S.},
  title = {Comparing Predictive Accuracy},
  journal = {Journal of Business
    \& Economic Statistics},
  year = {1995},
  volume = {13},
  number = {3},
  pages = {253--263}
}

@article{Sharpe1994,
  author = {Sharpe, William F.},
  title = {The Sharpe Ratio},
  journal = {Journal of Portfolio Management},
  year = {1994},
  volume = {21},
  number = {1},
  pages = {49--58}
}

@article{AndersenBollerslev1997,
  author = {Andersen, Torben G. and Bollerslev, Tim},
  title = {Intraday periodicity and volatility persistence in financial
    markets},
  journal = {Journal of Empirical Finance},
  year = {1997},
  volume = {4},
  number = {2-3},
  pages = {115--158}
}

@article{EngleRussell1998,
  author = {Engle, Robert F. and Russell, Jeffrey R.},
  title = {Autoregressive Conditional Duration: A New Model for Irregularly
    Spaced Transaction Data},
  journal = {Econometrica},
  year = {1998},
  volume = {66},
  number = {5},

```



```

    pages = {1127--1162}
}

@book{Hasbrouck2007,
  author = {Hasbrouck, Joel},
  title = {Empirical Market Microstructure: The Institutions, Economics, and
Econometrics of Securities Trading},
  publisher = {Oxford University Press},
  year = {2007}
}

@article{Amihud2002,
  author = {Amihud, Yakov},
  title = {Illiquidity and stock returns: cross-section and time-series
effects},
  journal = {Journal of Financial Markets},
  year = {2002},
  volume = {5},
  number = {1},
  pages = {31--56}
}

```

## Apêndice — versões curtas para improviso

**Elevator pitch (30–40s):** “Estudo preditivo intradiário na B3: avalio um **híbrido CNN+LSTM** para prever **probabilidades direcionais +1 barra (15m)** em **ações líquidas**, usando **walk-forward com embargo**, **métricas probabilísticas** e **backtests pós-custos**. Comparo com **ARIMA, Prophet e LSTM** e testo **robustez a custos e regimes de volatilidade**. Contribuo com **pipeline reprodutível** para o PFC2.”

**Fecho (20s):** “O ganho está em **probabilidades confiáveis** e um **protocolo justo** para ambientes voláteis. O PFC2 entrega evidências robustas e reprodutíveis.”