

Universidade Federal de Goiás
Regional Catalão
Unidade Acadêmica Especial de Biotecnologia
Curso de Bacharelado em Ciências da Computação

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

Rafael da Silva Melo

Catalão – GO
2025

Rafael da Silva Melo

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

Monografia apresentada ao Curso de Bacharelado em Ciências da Computação da Universidade Federal de Goiás – Regional Catalão, como parte dos requisitos para obtenção do grau de Bacharel em Ciências da Computação.
EXEMPLAR DE DEFESA I

Orientador: Prof. Dr. Márcio de Souza Dias

Catalão – GO
2025

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Melo, Rafael da Silva

Predição Automática de Indicativos Financeiros para Bolsa de Valores Considerando o Aspecto Temporal [manuscrito] / Rafael da Silva Melo. – 2025.

83 p.: il.

Orientador: Prof. Dr. Márcio de Souza Dias

Monografia (Graduação) – Universidade Federal de Goiás, Unidade Acadêmica Especial de Biotecnologia, Ciências da Computação, 2025.

Bibliografia.

1. predição.
 2. indicadores financeiros.
 3. séries temporais.
 4. bolsa de valores.
 5. aprendizado profundo.
- I. Dias, Márcio de Souza, orient. II. Título.

CDU 004

Rafael da Silva Melo

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

Monografia apresentada ao curso de Bacharelado em Ciências da Computação da Universidade Federal de Goiás – Regional Catalão.

Trabalho aprovado em 01 de Agosto de 2025.

Márcio de Souza Dias
Orientador

Marcio Antonio Duarte
Universidade Federal de Catalão

Thiago Jabur Bittar
Universidade Federal de Catalão

Catalão – GO
2025

RESUMO

MELO, R. S.. **Predição Automática de Indicativos Financeiros para Bolsa de Valores Considerando o Aspecto Temporal.** 2025. 83 p. Monografia (Graduação) – Unidade Acadêmica Especial de Biotecnologia, Universidade Federal de Goiás – Regional Catalão, Catalão – GO.

O trabalho introduz um modelo híbrido de aprendizado profundo, utilizando uma combinação de redes neurais LSTM + CNN para prever indicadores financeiros da bolsa brasileira B3 — Brasil, Bolsa, Balcão — com base em dados de séries temporais intradiárias.

A análise incorpora dados de liquidez, volatilidade e volume negociado, juntamente com engenharia de características e uma metodologia de validação cruzada temporalmente consciente para melhorar a robustez e a repetibilidade dos resultados.

Sua arquitetura foi construída para inferir cenários de estresse e choques de mercado, algo fora do alcance de métodos clássicos como ARIMA ou Prophet. Um importante marco é que produz valores muito precisos, o que importa muito mais para negociantes da bolsa e gerentes de risco.

Um pipeline baseado em Python será construído para ser usado na reprodução e geração de dados, além da agregação de dados e análise subsequente na qual este estudo se baseia. Essas contribuições implicam em uma maior estabilidade em ambientes voláteis e insights que podem ser transferidos para vários ativos financeiros.

Palavras-chave: predição, indicadores financeiros, séries temporais, bolsa de valores, aprendizado profundo.

ABSTRACT

MELO, R. S.. **Predição Automática de Indicativos Financeiros para Bolsa de Valores Considerando o Aspecto Temporal.** 2025. 83 p. Monografia (Graduação) – Unidade Acadêmica Especial de Biotecnologia, Universidade Federal de Goiás – Regional Catalão, Catalão – GO.

The paper introduces a hybrid deep learning model that combines LSTM and CNN neural networks to predict financial indicators of the Brazilian exchange B3 — Brasil, Bolsa, Balcão — based on intraday time-series data.

The analysis incorporates liquidity, volatility, and traded volume data, together with feature engineering and a time-aware cross-validation methodology to improve the robustness and replicability of the results.

Its architecture was designed to infer stress scenarios and market shocks, something beyond the reach of classical methods such as ARIMA or Prophet. A key advantage is that it produces highly accurate values, which matters much more to stock traders and risk managers.

A Python-based pipeline will be built to support reproducibility and data generation, as well as data aggregation and the subsequent analysis on which this study relies. These contributions imply greater stability in volatile environments and insights that can be transferred to multiple financial assets.

Keywords: prediction, financial indicators, time series, stock exchange, deep learning.

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Contextualização	13
1.2	Limitações dos modelos clássicos	16
1.3	Problema de Pesquisa	17
1.3.1	<i>Justificativa</i>	19
1.4	Objetivos	19
1.4.1	<i>Objetivo geral</i>	19
1.4.2	<i>Objetivos específicos</i>	20
1.4.3	<i>Contribuição Esperada</i>	20
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Mercado financeiro em geral	23
2.1.1	<i>Regulação e Estrutura da B3 como Infraestrutura de Mercado</i>	24
2.1.2	<i>Conectividade e Ambientes de Negociação</i>	24
2.1.3	<i>Ativos</i>	24
2.1.4	<i>Principais conceitos utilizados no mercado financeiro</i>	25
2.1.5	<i>Indicadores técnicos</i>	26
2.1.6	<i>Preço real, o que acontece</i>	27
2.1.6.1	<i>Representação dos preços em Candlesticks</i>	28
2.2	Estacionariedade e transformações usuais	29
2.2.1	<i>Volatilidade e Liquidez Intradiária</i>	30
2.2.2	<i>Componente Temporal em Predição de Séries Financeiras</i>	31
2.2.3	<i>Séries Temporais e a Aplicação de Deep Learning em Finanças</i>	32
2.2.3.1	<i>Redes Recorrentes Long Short-Term Memory (LSTM)</i>	32
2.2.3.2	<i>Arquitetura Transformer e o Mecanismo de Atenção</i>	34
3	TRABALHOS RELACIONADOS	37
3.1	Modelos Clássicos e Estatísticos	37
3.2	Aprendizado de Máquina e Modelos Híbridos	38
3.3	Redes Neurais Profundas	40
3.4	Transformers para Séries Temporais	41
3.5	LLMs atuais aplicados a Séries Temporais Financeiras	42
3.6	Síntese Crítica	44

4	METODOLOGIA	47
4.1	<i>Aquisição e descrição dos dados</i>	48
4.1.1	<i>Fonte e universo</i>	48
4.1.2	<i>Recorte temporal e granularidade</i>	49
4.1.3	<i>Variáveis</i>	49
4.1.4	<i>Tratamento de qualidade</i>	49
4.1.5	<i>Estatística descritiva do dataset</i>	49
4.2	<i>Pré-processamento e engenharia de atributos</i>	49
4.2.1	<i>Normalização e prevenção de data leakage</i>	50
4.2.2	<i>Atributos calculados</i>	50
4.2.3	<i>Janela temporal (lookback) e trade-off</i>	51
4.3	<i>Modelos - baselines e proposto</i>	52
4.3.1	<i>Baselines e justificativa.</i>	52
4.3.2	<i>Modelo proposto: CNN–LSTM.</i>	53
4.3.3	<i>Hiperparâmetros e tuning.</i>	53
4.4	<i>Desenho experimental e treinamento</i>	54
4.4.1	<i>Protocolo temporal (walk-forward) e particionamento</i>	54
4.4.2	<i>Treinamento: épocas, batch, early stopping e sementes</i>	55
4.4.3	<i>Tuning de hiperparâmetros: por bloco, não aninhado</i>	55
4.5	<i>Métricas de Avaliação</i>	56
4.5.1	<i>Métrica principal e justificativa.</i>	56
4.5.2	<i>Métricas probabilísticas e calibração.</i>	57
4.5.3	<i>Agregação dos resultados entre folds.</i>	57
4.6	<i>Backtests e custos</i>	58
4.6.1	<i>Regras do backtest.</i>	58
4.6.2	<i>Custos e slippage.</i>	58
4.6.3	<i>Métricas do backtest.</i>	58
4.7	<i>Robustez e significância</i>	59
4.7.1	<i>Teste de Diebold–Mariano.</i>	59
4.7.2	<i>Análise por regimes e por granularidade.</i>	59
4.8	<i>Reprodutibilidade</i>	59
4.8.1	<i>Organização do repositório e versões.</i>	59
4.8.2	<i>Reprodução do zero.</i>	60
5	RESULTADOS E DISCUSSÃO	61
5.1	<i>Visão geral dos resultados e organização dos artefatos</i>	61
5.2	<i>Comparação geral: CNN–LSTM vs. baselines</i>	62
5.3	<i>Estabilidade no tempo: desempenho por fold do walk-forward</i>	63
5.4	<i>Top configurações do Optuna e sensibilidade do lookback</i>	63
5.5	<i>Significância estatística: teste de Diebold–Mariano</i>	64

5.6	Backtests: impacto prático após custos	66
5.7	Robustez: regimes de volatilidade	66
5.8	Síntese	67
6	CONSIDERAÇÕES FINAIS	69
	REFERÊNCIAS	71
7	GLOSSÁRIO	77
8	SIGLAS	83



INTRODUÇÃO

1.1 Contextualização

As bolsas de valores são 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 (MISHKIN; EAKINS, 2018; FABOZZI; MODIGLIANI, 2009). No Brasil, esse papel é desempenhado pela B3 — Brasil, Bolsa, Balcão —, que 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 (B3, 2025b; BRASIL, 1976a; BRASIL, 1976b). Em mercados acionários, os preços refletem expectativas sobre fluxos de caixa futuros e risco, mas também exibem forte volatilidade e variação de liquidez (BODIE; KANE; MARCUS, 2021).

No horizonte intradiário, emergem problemas clássicos de microestrutura — periodicidades sistemáticas ao longo do dia, espaçamento irregular entre negócios e impacto de ordens (inclusive via *spread*¹ e custo de imediatismo) — que condicionam a dinâmica de preços e a interpretação de sinais de curto prazo (ANDERSEN; BOLLERSLEV, 1997; ENGLE; RUSSELL, 1998; HASBROUCK, 2007; AMIHUD, 2002; DUFOUR; ENGLE, 2000). Além disso, choques macroeconômicos e eventos exógenos podem deslocar rapidamente as cotações, elevando a incerteza (BANK, 2020). Esse contexto motiva a investigação de métodos preditivos sensíveis ao tempo e às nuances operacionais do ambiente de negociação brasileiro.

Com os avanços da tecnologia, a complexidade e a volatilidade dos ativos negociados em bolsa de valores aumentaram de forma significativa. Embora o mercado de ações não seja teoricamente previsível, muitos investidores identificam padrões que motivam a busca por estratégias de previsão(LO, 2017). No mundo, ativos fortemente correlacionados com mercados internacionais e fluxo de notícias, bem como a crescente participação de investidores individuais

¹ Diferença entre o preço de compra (bid) e o de venda (ask) de um ativo.

e institucionais, reforçam a necessidade de ferramentas que esclareçam sinais de curto, médio e longo prazo.

Neste estudo, considera-se explicitamente um modelo híbrido baseado em redes recorrentes do tipo LSTM combinadas a convoluções unidimensionais (CNN+LSTM), visando capturar simultaneamente dependências de longo prazo e padrões locais em séries intradiárias da B3. Essa escolha se apoia em evidências de que LSTMs representam melhor relações não lineares e dependências temporais prolongadas, enquanto camadas convolucionais extraem motivos de curto prazo a partir de janelas deslizantes; em conjunto, busca-se maior robustez frente a ruído e a mudanças de regime de mercado. Para assegurar comparabilidade, os resultados serão confrontados com baselines estatísticos (ARIMA, Prophet) e com variantes baseadas em Transformers reportadas nas literaturas ([DING; QIN, 2020](#); [NASCIMENTO; SANTOS; FERREIRA, 2023](#); [ZENG et al., 2022](#); [WEN et al., 2022](#)).

No entanto, como Zvi Bodie and Alex Kane and Alan J. Marcus ([BODIE; KANE; MARCUS, 2021](#)) dizem em seu livro, o mercado de ações é impulsionado por variáveis econômicas (taxas de juros, taxas de câmbio, indicadores de inflação), fatores corporativos (demonstrações financeiras da empresa e mudanças na governança) e também eventos externos (crises políticas, pandemias, choques de oferta) como diz ([BANK, 2020](#)). Estes quesitos permitem que, em cada série temporal, sejam produzidas componentes como tendência, sazonalidade e ciclos, conforme a decomposição tradicional proposta em Box e Jenkins ([BOX et al., 2015](#)) e amplamente discutida em Hyndman e Athanasopoulos ([HYNDMAN; ATHANASOPOULOS, 2018](#)):

- Tendência: define se os preços atuais estão subindo ou caindo a curto prazo.
- Sazonalidade: é qualquer tipo de padrão repetitivo entre ciclos diários a anuais.
- Ciclos: um movimento de oscilação de médio a longo prazo, que não segue um calendário fixo, mas decorre de condições econômicas, corporativas e externas.
- Ruído aleatório: variações imprevisíveis, que não podem ser explicadas por tendência, sazonalidade ou ciclos, geralmente decorrentes de choques ou flutuações aleatórias.

A figura 1 apresenta diferentes combinações dos componentes básicos de uma série temporal: tendência, ciclos e ruído. Cada gráfico mostra isoladamente ou em conjunto esses elementos, evidenciando como eles moldam o comportamento dos dados ao longo do tempo. A tendência representa o movimento de longo prazo (aumento ou queda contínua), os ciclos mostram padrões repetitivos de médio prazo ligados a fatores econômicos ou sazonais, e o ruído corresponde a flutuações aleatórias sem padrão definido. Assim, ao combinar esses elementos, observamos séries mais complexas e realistas, aproximando-se do que ocorre em dados empíricos.

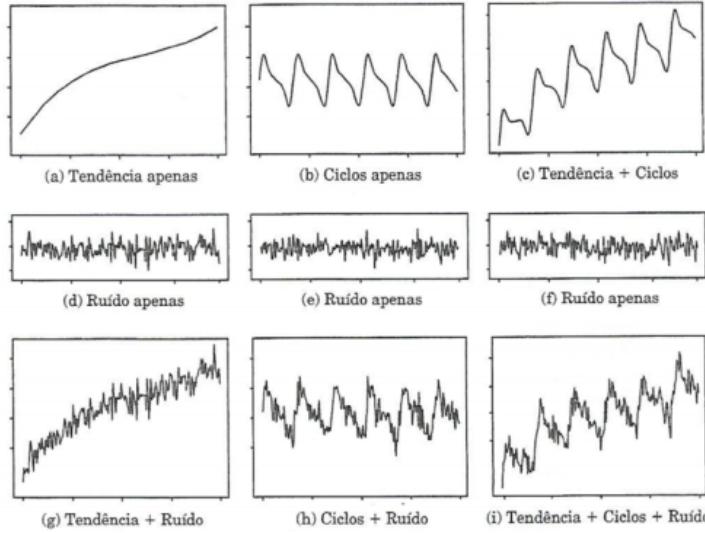


Figura 1 – Explicação visual sobre cada série temporal. Fonte: ([LEONUNESBIO, 2014](#)).

É aqui que avanços arquitetônicos recentes entram em cena — como o *Temporal Fusion Transformer* (TFT) ([LIM; ZOHREN; ROBERTS, 2021](#)), variantes Transformers voltadas à não-estacionaridade ([LIU et al., 2022](#)), mecanismos de atenção por “patches” (PatchTST) ([NIE et al., 2023](#)) e redes temporais dedicadas como o *TimesNet* ([WU et al., 2023](#)). Além disso, modelos fundacionais e abordagens baseadas em **Grandes Modelos de Linguagem (LLMs)** vêm sendo aplicados a séries temporais, como o *TimesFM* ([GOEL et al., 2025](#)) e o *TimeGPT* ([LATHUILÈRE et al., 2024](#)), juntamente com métodos que alinham LLMs à tarefa de previsão (LLM4TS; Time-LLM) ([CHANG et al., 2023; JIN et al., 2023](#)) e aplicações no domínio financeiro ([ALONSO; FRANKLIN, 2024](#)). Há também um debate quanto à efetividade prática dos Transformers em séries temporais ([ZENG et al., 2022; WEN et al., 2022](#)). Para cenários com dados multimodais, destacam-se propostas como o *Hidformer* ([LIANG et al., 2025](#)). Apesar desses avanços, contextos intradiários permanecem desafiadores — periodicidade intradiária e efeitos de microestrutura ([ANDERSEN; BOLLERSLEV, 1997; HASBROUCK, 2007](#)), irregularidade temporal/duração entre negócios ([ENGLE; RUSSELL, 1998](#)) e custo computacional — tema que será retomado na seção de Trabalhos Relacionados.

A própria natureza do comércio, marcada por negociações frequentes e pelo uso de estratégias algorítmicas — isto é, sistemas automatizados que executam ordens de compra e venda com base em regras predefinidas, muitas vezes em alta frequência —, enfatiza a importância de modelos capazes de capturar rapidamente padrões transitórios sem perder de vista o horizonte de médio prazo de maneira rigorosa ([SUNKI et al., 2024](#)).

Outras pesquisas já propuseram modelos híbridos para previsão de preços de ações — por exemplo, o modelo proposto por ([JUNIOR, 2013](#)), que combina redes neurais, técnicas estatísticas e indicadores técnicos via comitê de aprendizado. Entretanto, tais propostas, em geral, não exploram esquemas de agregação mais sofisticados (como metaclassificação em dois estágios) nem avaliam de forma sistemática a robustez a mudanças súbitas de regime ([SOUZA](#)

et al., 2022; TSAY, 2010).

Coerente com o título desta monografia, explicitamos os **alvos preditivos (indicativos financeiros)** a serem modelados: direção do return na próxima barra, medida de volatilidade intradiária (p.ex., amplitude/variância realizada) e sinais de liquidez, como volume negociado e spread efetivo (TSAY, 2010; DACOROGNA *et al.*, 2001; WOOD; MCINISH; ORD, 1985; MISHKIN; EAKINS, 2018; JR; OUTROS, 2014; NASCIMENTO; SANTOS; FERREIRA, 2023).

Portanto, este trabalho avança a linha dos híbridos ao empregar um modelo CNN+LSTM com camadas convolucionais para padrões locais, validação temporal walk-forward e uma avaliação realista pós-custos no contexto da B3, com foco em estabilidade e reproduzibilidade dos resultados.

1.2 Limitações dos modelos clássicos

Métodos estatísticos tradicionais, como ARIMA e seus desdobramentos sazonais, oferecem boa interpretabilidade e desempenho em séries estacionárias ou de baixa não linearidade. Todavia, como demonstram estudos recentes, esses modelos tendem a perder eficácia quando confrontados com mercados “barulhentos”, em que pequenos choques(notícias inesperadas, decisões judiciais, compras ou vendas grandes em pouco tempo) podem alterar radicalmente as projeções (SUNKI *et al.*, 2024).

Ding e Qin (DING; QIN, 2020) demonstraram o modelo associado de rede baseada em Long Short-Term Memory - LSTM² é capaz de predizer simultaneamente preço de abertura, mínima e máxima, alcançando acurácia superior a 95% em dados históricos de bolsa. (NASCIMENTO; SANTOS; FERREIRA, 2023) compararam modelos ARIMA, Prophet e LSTM para séries financeiras, mostraram que arquiteturas baseadas em redes recorrentes capturam melhor dependências de tempo mais longas. Já propostas de *ensembles* híbridos (por exemplo, um comitê que combina múltiplas LSTMs paralelas) sugerem ganhos de robustez e precisão. Diante do cenário dos modelos híbridos e *ensembles*, que combinam redes profundas, métodos estatísticos e indicadores técnicos, a fim de aproveitar as vantagens de cada abordagem e mitigar suas fragilidades, tem-se o exemplo promissor que é o uso de metaclassificadores, em que vários algoritmos (e.g., redes LSTM, SVM, random forest) produzem previsões independentes que são então combinadas por um modelo adicional, responsável por aprender como integrar essas saídas e gerar uma decisão final mais consistente (SOUZA *et al.*, 2022).

Ao mesmo tempo, a inclusão de indicadores técnicos, como médias móveis, índice de força relativa (RSI) ou bandas de Bollinger, tem se mostrado útil para modelos de ML no enriquecimento de informações de tendência e volatilidade marcadas por janelas de preços históricos. Isso melhora a capacidade das redes neurais de discriminar entre sinais reais e ruído

² rede recorrente com memória de longo prazo

(DING; QIN, 2020). Se a volatilidade for suficientemente alta, esses comitês têm se mostrado mais estáveis e confiáveis na previsão do que qualquer um desses métodos usados isoladamente.

Embora vários estudos tenham afirmado precisões acima de 90% no caso de previsões diárias, análises subsequentes mostraram um viés de "falsa segurança": muitos resultados não são muito úteis em cenários reais de mercado porque ignoram a dependência temporal real entre treinamento e teste.

1.3 Problema de Pesquisa

Resultados de previsão intradiária frequentemente não se sustentam em uso real porque violam a ordem temporal, ignoram regimes de liquidez/volatilidade e desconsideram custos de transação. Nesse cenário, permanece incerto se modelos de aprendizado profundo — em especial arquiteturas híbridas CNN+LSTM — geram sinais direcionais realmente utilizáveis (horizonte +1 barra/15 min) para ações líquidas da B3 entre 2020–2025, quando avaliados sob um protocolo rigoroso de validação walk-forward e comparados a baselines clássicos e arquiteturas com atenção.

Delimitação do escopo:

Este estudo situa-se no universo de bolsas de valores e mercados organizados, em que a negociação se dá sob regras de listagem, leilões, formação de preço e procedimentos operacionais definidos (no caso brasileiro, pela B3) (B3, 2025). Em janelas intradiárias, o comportamento do preço é influenciado por aspectos de *microestrutura* (ex.: liquidez, *spread*, intensidade de ordens, periodicidades intradiárias), o que afeta diretamente a previsibilidade e a avaliação de modelos (HASBROUCK, 2007; ANDERSEN; BOLLERSLEV, 1997; DUFOUR; ENGLE, 2000; ENGLE; RUSSELL, 1998). À luz dessa dinâmica, explicitamos a seguir as delimitações adotadas.

- **Ativos:** Consideramos ações líquidas negociadas na B3 (e.g., PETR4, VALE3, ITUB4) e, quando pertinente, a carteira do Ibovespa como referência de mercado. A escolha por alta liquidez visa mitigar efeitos de microestrutura adversos (como slippage e spread amplos) que distorcem sinais intradiários (HASBROUCK, 2007).
- **Período e granularidade:** Os experimentos cobrem jan/2020–jul/2025 em barras de 15 minutos, com análises de sensibilidade em 5 e 30 minutos. Essa granularidade equilibra ruído e sinal em alta frequência e captura periodicidades intradiárias conhecidas (picos de volatilidade na abertura/fechamento, padrões de almoço, etc.) (ANDERSEN; BOLLERSLEV, 1997; HYNDMAN; ATHANASOPOULOS, 2018).
- **Tarefa:** Prever a **direção do movimento de preço** na próxima vela (horizonte +1 barra), formulada como classificação probabilística. Tal formulação permite avaliar calibração e

qualidade probabilística, não apenas acerto bruto (BRIER, 1950; GUO *et al.*, 2017) — ver também a interpretação visual da tarefa na Figura 2.

- **Validação e treinamento:** Para evitar viés temporal (ook-ahead) e superestimação de desempenho, adotamos validação walk-forward com janelas sequenciais, seleção de hiperparâmetros por otimização Bayesiana e controle de sementes/versionamento para reproduzibilidade. Esse arranjo segue boas práticas de validação em séries temporais e backtesting (BERGMEIR; BENÍTEZ, 2012; PRADO, 2018).

- **Métricas:** Aferimos:

- I. *acurácia direcional com banda morta*, para robustez a microvariações;
- II. Brier Score e Log-Loss, para qualidade das probabilidades e calibração (BRIER, 1950; GUO *et al.*, 2017);
- III. métricas pós-custos de trading (retorno líquido, drawdown e índice de Sharpe), refletindo utilidade prática sob custos de transação (SHARPE, 1994)

Para comparar previsores quando necessário, usamos testes de acurácia preditiva como Diebold–Mariano (DIEBOLD; MARIANO, 1995).

- **Comparadores (baselines):** Contrastamos o modelo proposto com referências clássicas e modernas:

- I. ARIMA/Box–Jenkins como baseline estatístico canônico (BOX *et al.*, 2015; HYNDMAN; ATHANASOPOULOS, 2018; JR; OUTROS, 2014);
- II. Prophet, representando um decompositor aditivo com sazonalidades/mudanças de nível (TAYLOR; LETHAM, 2018);
- III. arquiteturas modernas para séries temporais, incluindo variantes baseadas em Transformers, dada sua capacidade de capturar dependências de longo alcance (WEN *et al.*, 2022; ZENG *et al.*, 2022).

A ilustração 2 sintetiza a tarefa: em uma série de preços intradiária (barras de 15 min), cada ponto no tempo gera um vetor de atributos (preço, volume, indicadores técnicos e variáveis de contexto) e o modelo estima a probabilidade de a próxima barra encerrar acima (ou abaixo) da atual. O horizonte é de +1 barra (curto prazo), e as janelas de treino/teste avançam sequencialmente no tempo (esquema walk-forward), refletindo o fluxo real de mercado e evitando contaminação temporal.

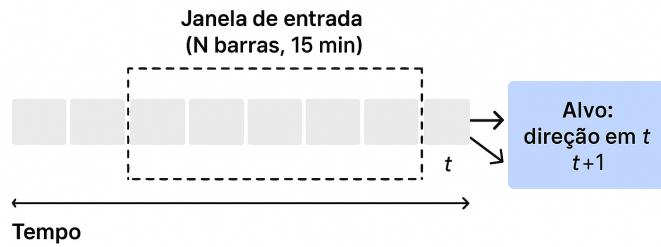


Figura 2 – Delimitação da tarefa: previsão direcional intradiária (+1 barra, 15 min). Fonte: autor.

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?

1.3.1 Justificativa

A previsão intradiária de movimentos de preços na B3 apresenta relevância prática para apoio à decisão em ambientes voláteis e com custos de transação não desprezíveis. Do ponto de vista acadêmico, persiste uma lacuna entre resultados reportados em ambientes controlados e o desempenho sob *walk-forward*³, com avaliação pós-custos e controle de deriva de conceito. Este estudo busca reduzir essa lacuna ao (i) delimitar um cenário intradiário realista, (ii) empregar validação temporal adequada e (iii) comparar um modelo híbrido (CNN+LSTM) a *baselines* consolidados. O potencial impacto inclui melhores critérios de avaliação para tarefas direcionais, evidências sobre robustez em diferentes regimes de volatilidade e um *pipeline* reproduzível que pode ser replicado e estendido por outros pesquisadores.

1.4 Objetivos

1.4.1 Objetivo geral

Avaliar de forma quantitativa e reproduzível, a capacidade de modelos de aprendizado de máquina sensíveis ao tempo em prever o comportamento direcional de indicadores financeiros

³ Método em que o treino e o teste avançam em janelas sequenciais, mantendo a ordem temporal.

intradiários, comparando seu desempenho a baselines estatísticos e neurais sob um protocolo experimental realista (validação walk-forward e backtext⁴) pós-custos.

1.4.2 Objetivos específicos

1. Construir e documentar o conjunto de dados intradiários (OHLCV e indicadores técnicos) para o período definido (2020–2025) e frequência fixa (15 minutos), especificando fontes, regras de limpeza, critérios de inclusão de ativos e prevenção de vazamento temporal.
2. Definir o desenho experimental com validação walk-forward (janelas deslizantes, embargo temporal e padronização dentro de cada janela), estratégia de seleção de hiperparâmetros e controle de variáveis (sementes, versões e ambiente).
3. Treinar e comparar *baselines* representativos com o(s) modelo(s) proposto(s) sensíveis à ordem temporal, mantendo critérios comparáveis de otimização.
4. Mensurar a qualidade preditiva por múltiplas métricas: acurácia direcional com banda morta, F1/recall para a classe positiva, Brier score e AUC-PR, incluindo análise de calibração das probabilidades.
5. Avaliar desempenho operacional convertendo sinais em regras simples de negociação e realizando backtests com custos de transação e slippage, reportando retorno, Sharpe, max drawdown e turnover, além de cortes por regimes de volatilidade.
6. Testar significância e robustez dos resultados por meio do teste de Diebold-Mariano, bootstrap em blocos e análises de sensibilidade (janelas temporais, subconjuntos de ativos, horizonte de previsão e feature sets).
7. Garantir reproduzibilidade com repositório versionado, manifesto de dependências, *scripts* de preparo/treino/avaliação, diários de experimento e registro explícito de ameaças à validade interna e externa.

1.4.3 Contribuição Esperada

Este trabalho busca contribuir em quatro frentes principais:

Primeiro, com o desenvolvimento de um modelo híbrido (CNN+LSTM) capaz de combinar padrões locais e dependências de longo prazo em séries temporais financeiras, adequando-se às particularidades do mercado intradiário da B3.

Em segundo lugar, ao realizar uma avaliação empírica comparativa com *baselines* estabelecidos (ARIMA, Prophet, LSTM puro e variantes Transformer), sob um protocolo realista de validação walk-forward e backtests pós-custos.

⁴ Simulação histórica de uma estratégia sob regras e dados passados.

A terceira contribuição está na análise da robustez do modelo em diferentes condições de mercado, explorando regimes de volatilidade e choques pontuais.

Finalmente, planeja-se lançar um pipeline reproduzível com a fonte organizada, bem como documentação detalhada, não apenas para apoiar a futura replicação de experimentos, mas também para facilitar a adaptação e extensão por outros pesquisadores interessados nessas abordagens.

Este trabalho promoverá e eliminará barreiras técnicas à comunidade científica e, desencadeará a co-construção do conhecimento. Além disso, a natureza aberta do pipeline permite que novas metodologias sejam testadas em igualdade de condições, tornando os resultados mais comparáveis e, portanto, contribuindo para conclusões mais robustas.



FUNDAMENTAÇÃO TEÓRICA

2.1 Mercado financeiro em geral

Em primeiro lugar, é preciso compreender que o mercado financeiro é um espaço institucionalizado no qual diversos agentes econômicos negociam diferentes tipos de ativos, como títulos de dívida, *commodities*, derivativos e moedas (MISHKIN; EAKINS, 2018).

As ações também fazem parte desses ativos, frações adequadas do capital social de empresas de capital aberto cuja negociação é realizada regularmente por meio das bolsas de valores (BRASIL, 1976b). Investidores que compram uma ação tornam-se acionistas, ou proprietários parciais das corporações que emitiram tais ações, e, consequentemente, têm certos direitos legais (o direito de receber sua parte proporcional dos dividendos distribuídos pela corporação e, muitas das vezes, o direito de votar em decisões corporativas indiretamente). Isso significa que as ações não são apenas sobre como a empresa ganha dinheiro na realidade, mas também sobre o que os investidores acham que ela gerará no futuro — portanto, a negociação de ações nada mais é do que uma percepção entre as pessoas e as dinâmicas econômicas gerais.

A B3 opera em um ambiente regulado que garante segurança jurídica e transparência às operações realizadas pelos participantes, permitindo que investidores forneçam recursos às empresas ou ao governo em troca de remuneração futura, geralmente sob forma de juros, dividendos ou ganhos de capital advindos da valorização dos ativos adquiridos de acordo com a (BRASIL, 1976a).

Existem muitas bolsas de valores espalhadas por todo o mundo, cada uma operando sob regulamentações específicas, negociando ativos próprios e refletindo as particularidades econômicas e culturais das regiões onde estas estão inseridas. Exemplos notórios incluem a Bolsa de Nova York (New York Stock Exchange, s.d.), a Nasdaq (NASDAQ, s.d.) nos Estados Unidos, a Bolsa de Londres no Reino Unido (London Stock Exchange, s.d.), e a Bolsa de Valores de São Paulo no Brasil (B3, s.d.).

2.1.1 Regulação e Estrutura da B3 como Infraestrutura de Mercado

Considerando o funcionamento do mercado financeiro brasileiro, é essencial destacar o papel central da B3 (B3, 2025b). A B3 é reconhecida como uma das principais empresas de infraestrutura de mercado financeiro no mundo, atuando nos ambientes de bolsa e de balcão organizado (B3, 2025b).

Neste contexto, a B3 atua como contraparte central, integrando funções de registro, custódia, liquidação e supervisão em uma cadeia verticalmente estruturada. Essa estrutura assegura a formação adequada de preços e a ampla divulgação das ofertas no livro central, com o objetivo de proteger os investidores e garantir eficiência ao mercado (B3, 2025a; B3, 2025b).

Logo, os participantes do mercado — corretoras, formadores de mercado, provedores de tecnologia e emissores — devem operar conforme os procedimentos estabelecidos pela B3, incluindo padrões de conexão como DMA, co-location e PNP, além da manutenção de sistemas de auditoria e controles internos descritos no Manual de Procedimentos Operacionais de Negociação (B3, 2025).

2.1.2 Conectividade e Ambientes de Negociação

Para que os modelos de predição sejam robustos, é essencial compreender os diferentes tipos de conexão ao ambiente de negociação da B3. Segundo o Manual de Procedimentos Operacionais de Negociação da B3, as conexões ao ambiente de negociação são classificadas conforme sua origem e hospedagem — incluindo acesso via mesa de operações, assessores e conexão direta patrocinada (DMA) (B3, 2025). Essa última, conhecida como DMA (Direct Market Access), permite que investidores ou provedores enviem ordens diretamente ao sistema de negociação, sem intermediação manual, reduzindo latência e aumentando a eficiência na execução de trades (B3, 2025).

Além do mais, a B3 oferece o serviço de *co-location*, onde servidores de corretoras ou investidores são instalados fisicamente em data centers da bolsa (como COLO0/COLO1), diminuindo ainda mais a latência em relação ao DMA tradicional fora do centro de dados da B3 (B3, 2025). Essa configuração é especialmente importante para estratégias de alta frequência (HFT), nas quais milissegundos influenciam diretamente os retornos.

Portanto, ao lidar com dados de alta frequência em seu modelo de predição é fundamental considerar que os atrasos (latência) e o tipo de conexão (DMA, co-location ou assessor) impactam a qualidade e a reproduzibilidade dos resultados no ambiente real de negociação.

2.1.3 Ativos

Os ativos financeiros são instrumentos ou contratos negociados no mercado financeiro, com características e funções específicas, permitindo aos investidores diversificar seus investi-

mentos de acordo com diferentes perfis de risco e expectativas de retorno (MISHKIN; EAKINS, 2018; FABOZZI; MODIGLIANI, 2009). Conforme introduzido na seção 2.1, esses ativos incluem **ações, títulos públicos e privados, moedas, derivativos e commodities** (FABOZZI; MODIGLIANI, 2009). Cada categoria apresenta particularidades quanto ao comportamento dos preços, liquidez e risco associado, tornando imprescindível uma análise detalhada antes de qualquer decisão de investimento.

Os preços desses ativos são influenciados por fatores variados, desde condições climáticas e ambientais até eventos geopolíticos e econômicos globais (BANK, 2020), o que faz deles ativos voláteis, cujos preços podem oscilar significativamente em curtos períodos. Além disso, muitos desses ativos são interligados, e variações no preço de um deles podem influenciar diretamente os demais, aumentando a complexidade das análises financeiras e exigindo dos investidores estratégias consistentes e informações atualizadas para tomadas de decisão mais assertivas (MISHKIN; EAKINS, 2018).

2.1.4 Principais conceitos utilizados no mercado financeiro

Descrever a direção predominante do movimento dos preços ao longo do tempo caracteriza o que chamamos de **tendência**. Quando um ativo apresenta uma sequência consistente de preços ascendentes, diz-se que há uma **tendência de alta**; quando ocorre o inverso, com preços descendentes de forma persistente, trata-se de uma **tendência de baixa**. Se os preços não manifestam uma direção clara, oscilando em faixas relativamente estáveis, caracteriza-se como ausência de tendência ou **mercado lateralizado** (BODIE; KANE; MARCUS, 2021).

Além do conceito de tendência, outros termos importantes são frequentemente utilizados para descrever os movimentos e comportamentos observados na bolsa de valores. Um desses termos é a **volatilidade**, que representa o grau de oscilação dos preços dos ativos ao longo de um determinado período, refletindo a instabilidade ou a incerteza do mercado naquele momento (MISHKIN; EAKINS, 2018). Quando se fala em **liquidez**, refere-se à facilidade com que um ativo pode ser convertido em dinheiro sem causar grande impacto no seu preço. Um ativo com alta liquidez pode ser negociado rapidamente, enquanto um ativo com baixa liquidez possui menos compradores e vendedores, dificultando sua negociação imediata (FABOZZI; MODIGLIANI, 2009).

Outro termo amplamente utilizado é o *dividendo*, que corresponde a uma parcela dos lucros que as empresas distribuem periodicamente aos acionistas como forma de remuneração pelo investimento realizado (BRASIL, 1976b; BODIE; KANE; MARCUS, 2021).

É comum ouvir falar em **bear market** e **bull market**. O primeiro descreve períodos prolongados de queda nos preços das ações e pessimismo generalizado no mercado. Por outro lado, um **bull market** refere-se a fases caracterizadas por aumentos contínuos e generalizados dos preços dos ativos financeiros, acompanhados por otimismo dos investidores em relação ao

futuro econômico ([BODIE; KANE; MARCUS, 2021](#)).

Por fim, **Índice de mercado** diz respeito a indicadores que medem o desempenho médio de um conjunto de ativos negociados em bolsa ([DAYCOVAL, 2024](#)), como o Ibovespa no Brasil ou o S&P 500 nos Estados Unidos, servindo como referência para avaliar o comportamento geral do mercado financeiro.

2.1.5 Indicadores técnicos

Além dos índices de mercado, existem também os chamados indicadores técnicos. Eles são muito usados na análise gráfica, justamente para tentar entender como os preços se comportam e, a partir disso, decidir o melhor momento de comprar ou vender um ativo. A diferença é que, enquanto os índices de mercado mostram o desempenho geral de um grupo de ações ou ativos, os indicadores técnicos olham para cada ativo separadamente. Eles funcionam com base em cálculos matemáticos feitos sobre o histórico de preços e sobre o volume de negociações, tentando transformar esses números em sinais que ajudem o investidor a agir com mais segurança.

Além dos índices de mercado tradicionais, existem outros tipos de indicadores, denominados indicadores técnicos, utilizados principalmente na análise gráfica para avaliar o comportamento dos preços e tomar decisões relacionadas à compra e venda de ativos. Diferentemente dos índices de mercado, que avaliam o desempenho geral de um conjunto específico de ativos, os indicadores técnicos são aplicados individualmente em cada ativo, com base em cálculos matemáticos realizados sobre os preços históricos e volumes de negociação ([MURPHY, 1999](#)).

Entre os indicadores técnicos mais utilizados, destacam-se:

- **Bandas de Bollinger:** indicam a volatilidade e possíveis níveis de suporte e resistência dos preços ([BOLLINGER, 2001](#));
- **Médias Móveis:** suavizam os dados históricos e ajudam a identificar a direção predominante dos preços ([MURPHY, 1999](#));
- **Índice de Força Relativa (RSI):** mede a força e a velocidade dos movimentos dos preços para avaliar possíveis situações de sobrecompra ou sobre venda;
- **MACD (Moving Average Convergence Divergence):** busca identificar mudanças nas tendências por meio das interações entre médias móveis.

A Figura 3 ilustra apenas como exemplo didático a aplicação prática desses indicadores em um gráfico de candles com periodicidade de 15 minutos. As **Médias Móveis de 9, 21 e 50 períodos** suavizam o comportamento dos preços e ajudam a identificar tendências de curto,

médio e longo prazo ([MURPHY, 1999](#)). Em torno das velas estão as **Bandas de Bollinger (20x2)**, que se expandem ou se contraem conforme a volatilidade do mercado, sinalizando possíveis regiões de suporte e resistência ([BOLLINGER, 2001](#)). Na parte inferior do gráfico, as **janelas do Índice de Força Relativa (RSI) com 9, 21 e 50 períodos** medem a força do movimento dos preços, indicando potenciais situações de sobrecompra ou sobrevenda ([MURPHY, 1999](#)).



Figura 3 – Mini-chart de 15 minutos da ação PETR4 (Petrobras), com candles, Médias Móveis de 9, 21 e 50 períodos, Bandas de Bollinger (20x2) e RSI de 9, 21 e 50. Fonte: Yahoo Finance, adaptado pelo autor.

Esses indicadores são ferramentas essenciais para investidores e analistas técnicos, permitindo uma melhor compreensão dos movimentos de preços e auxiliando na definição de estratégias para investimentos em curto, médio ou longo prazo. ([MURPHY, 1999](#))

2.1.6 Preço real, o que acontece

Normalmente, ao acompanharmos o mercado financeiro, prestamos atenção no preço final das ações. Aliás, é exatamente esse valor que vemos estampado nos jornais e sites especializados após o fechamento da bolsa ([MISHKIN; EAKINS, 2018](#)). Só que, enfim, esse número por si só conta apenas parte da história do que realmente aconteceu durante o pregão. A verdade é que, ao longo do dia, acontecem inúmeras transações de compra e venda, fazendo com que o preço oscile diversas vezes—às vezes, até bruscamente—antes de chegar ao valor definitivo no encerramento do dia ([MURPHY, 1999](#)).

Suponha-se que uma ação abra o pregão com preço elevado devido a notícias positivas divulgadas no dia anterior. Isso não garante que ela vá continuar assim durante todo o período. Na verdade, pode ocorrer o oposto: em poucas horas, influenciada por novas informações ou pela movimentação de grandes investidores, o preço pode despencar ([BODIE; KANE; MARCUS, 2021](#)). Por isso, apenas olhar o preço final é insuficiente para entender o comportamento completo da ação. É aí que surgem métricas mais especializadas.

O **preço de abertura**, por exemplo, revela o ponto inicial do dia e pode já sinalizar tendências imediatas devido a transações realizadas antes do pregão oficial—o chamado pre-market. Já o preço máximo alcançado durante o dia nos diz muito sobre a força compradora, sendo um sinal claro do otimismo momentâneo dos investidores (MURPHY, 1999). Se esse preço máximo estiver muito distante do fechamento, talvez seja sinal de alerta, indicando uma possível reversão de tendência. O inverso ocorre com o preço mínimo, que reflete o maior nível de pessimismo momentâneo. Caso esteja próximo ao preço de fechamento, talvez seja sinal de recuperação da confiança dos investidores no decorrer do dia (PRING, 2014).

Além disso, o volume de negociação também é uma parte integrante de toda essa narrativa. Em suma, é o total transacionado em reais durante o dia. Isso pode revelar muito sobre o interesse dos investidores naquela ação.

Uma grande empresa como a Vale ou a Petrobras pode movimentar milhões, enquanto uma pequena movimenta pouco. Por exemplo, um aumento repentino no volume geralmente indica a chegada de algum evento significativo. Por outro lado, uma diminuição no volume pode mostrar o interesse diminuindo e também implicar que a tendência pode estar se revertendo de sua direção atual.

Por fim, é essencial destacar que essas métricas podem e devem ser analisadas em conjunto. Cada uma delas fornece pistas importantes, mas é a combinação de todas que permite uma leitura mais precisa e rica sobre o que realmente aconteceu com os preços durante o dia na bolsa. Na prática, é esse conjunto completo de informações que possibilita aos investidores e analistas técnicos tomarem decisões mais assertivas e também ajuda redes neurais a preverem com mais acurácia os movimentos futuros do mercado (MURPHY, 1999; BODIE; KANE; MARCUS, 2021).

2.1.6.1 *Representação dos preços em Candlesticks*

O gráfico 4 de candlestick é uma das formas mais intuitivas de representar as oscilações de preço de ativos financeiros. Cada vela informa, em um determinado intervalo de tempo, os preços de abertura, fechamento, máxima e mínima, como dito na imagem.

Quando o fechamento é superior à abertura, a vela tende a ser clara, sinalizando valorização; se inferior, escura, indicando desvalorização. Essa configuração visual permite não apenas uma leitura rápida das tendências, mas também auxilia na identificação de possíveis reversões. De acordo com estudo recente de Phankokkruad e Worasuwannarak (2025), padrões como Doji e Hammer são usados como indicativos de mudança de direção nos preços e demonstram aplicabilidade em modelos preditivos baseados em redes neurais (PHANKOKKRUAD; WORASUWANNARAK, 2025).

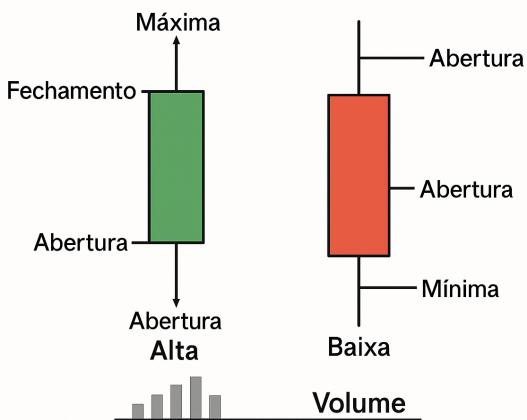


Figura 4 – Candlesticks. Fonte: [Giacomel \(2016\)](#).

2.2 Estacionariedade e transformações usuais

Mesmo quando os preços compartilham uma raiz unitária, embora os níveis de preços sejam comumente referidos como não estacionários em pesquisas e práticas financeiras, as flutuações de curto prazo podem frequentemente ser modeladas adequadamente por processos aproximadamente estacionários. É por isso que é uma tradição já estabelecida na área trabalhar com retornos e não com preços de ações simples (**CAMPBELL; LO; MACKINLAY, 1997; TSAY, 2010**). Neste texto, eles usam retornos logarítmicos definidos por $r_t = \ln(P_t) - \ln(P_{t-1})$, o que torna a série temporal estacionária em relação à escala e permite composições aditivas ao longo do tempo (**CAMPBELL; LO; MACKINLAY, 1997**).

Para avaliar a estacionaridade, podemos usar os seguintes testes complementares: o teste ADF (Augmented Dickey–Fuller), que tem como hipótese nula a existência de uma raiz unitária, e o KPSS, que assume como nula que os dados são estacionários em torno de um nível/tendência. Evita-se conclusões parciais através de uma leitura conjunta (**SAID; DICKEY, 1984**).

Para controle de escala e heterocedasticidade, são executadas transformações z-score móveis ou monotônicas (por exemplo, Box–Cox), ajustadas apenas nos dados de treinamento e depois aplicadas ao conjunto de validação/teste para evitar vazamento de dados (**HYNDMAN; ATHANASOPOULOS, 2021**).

Em algumas séries intradiárias, também se remove a tendência do efeito sazonal da periodicidade diária (abertura/fechamento), seja por meio de dummies intradiários ou subtraindo a média por faixa de tempo, removendo variações sistemáticas que não se generalizam fora do padrão diário (**DACOROGNA *et al.*, 2001**).

2.2.1 Volatilidade e Liquidez Intradíária

Segundo o Manual de Procedimentos Operacionais de Negociação da B3 (2019), mecanismos como a atuação de formadores de mercado, a estruturação das fases de negociação e parâmetros de lote mínimo influenciam significativamente os padrões intradiários de liquidez e volatilidade (B3, 2025). Observa-se, por exemplo, que há picos de volatilidade na abertura e no fechamento da sessão, enquanto o período intermediário costuma apresentar menor atividade.

A *volatilidade intradiária* pode ser mensurada por indicadores como o desvio-padrão de retornos em janelas móveis de 1, 5 ou 15 minutos (ANDERSEN; BOLLERSLEV, 1997). Andersen & Bollerslev (1997) demonstraram que a volatilidade não é constante ao longo do dia, concentrando-se em horários específicos e gerando efeitos de heterocedasticidade e autocorrelação nos retornos (ANDERSEN; BOLLERSLEV, 1997).

Por sua vez, a *liquidez* refere-se à facilidade de negociação de um ativo sem afetar substancialmente seu preço, sendo frequentemente avaliada por medidas como:

- Volume médio por minuto;
- Spread bid–ask, diferença entre o maior preço que alguém está disposto a pagar por um ativo e o menor preço pelo qual alguém está disposto a vendê-lo;
- Índice de iliquidez de Amihud, que relaciona o retorno absoluto ao volume negociado (AMIHUD, 2002);
- Profundidade do livro de ofertas, capturando o total de ordens disponíveis em níveis próximos ao melhor preço (HASBROUCK, 2007).

Estudos empíricos mostram que períodos de baixa liquidez tendem a apresentar maior volatilidade, o que prejudica a acurácia de modelos preditivos baseados em padrões técnicos (DUFOUR; ENGLE, 2000). Além disso, ativos com elevada liquidez exibem maior previsibilidade em horizontes de curtíssimo prazo, justificando a seleção desses papéis como base de dados para este trabalho.

Do ponto de vista metodológico, a variação intradiária de volatilidade e liquidez introduz dois principais desafios ao modelo:

1. **Heterocedasticidade** — volatilidade variável ao longo do dia requer o uso de modelos capazes de capturar regimes de alta e baixa variância (ex.: GARCH intradiário);
2. **Autocorrelação temporal** — séries de duração de ordens e inter-arrival times impactam a independência dos retornos, demandando abordagens como Autoregressive Conditional Duration (ENGLE; RUSSELL, 1998).

Dessa forma, a análise conjunta de volatilidade e liquidez intradiária fornece ao novo modelo, maior robustez e sensibilidade aos diferentes momentos da sessão, permitindo antecipar cenários de risco e otimizar as decisões de predição.

2.2.2 Componente Temporal em Predição de Séries Financeiras

Na predição de séries financeiras, o componente temporal é o elemento estruturante que vai além da simples ordem cronológica dos dados. Primeiro, efeitos sazonais e de calendário — como feriados, janelas de pré-abertura e fechamento — geram variações sistemáticas em volume e volatilidade que precisam ser modeladas explicitamente para evitar vieses sazonais (HYNDMAN; ATHANASOPOULOS, 2018). Em seguida, a escolha do tamanho da janela de análise (*rolling window*) implica em um *trade-off*: janelas muito longas diluem choques pontuais, enquanto janelas muito curtas amplificam ruídos e podem induzir sobreajuste (HYNDMAN; ATHANASOPOULOS, 2018).

No gráfico da Figura 5, esse equilíbrio é ilustrado visualmente: à medida que aumentamos o tamanho da janela de análise, reduzimos o impacto do ruído de curto prazo, mas também perdemos a capacidade de reagir rapidamente a mudanças repentinhas no mercado. Por outro lado, janelas muito curtas capturam com mais agilidade as variações recentes, porém amplificam flutuações aleatórias que podem induzir ao sobreajuste. Dessa forma, a figura evidencia o dilema central entre estabilidade e sensibilidade, que precisa ser considerado na escolha do horizonte temporal para modelagem de séries financeiras.

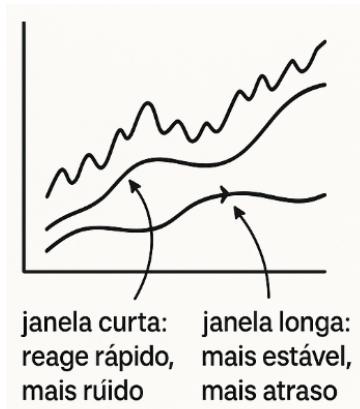


Figura 5 – *Trade-off* do tamanho da janela. Fonte: o autor

Para avaliar corretamente esse *trade-off* e estimar a performance out-of-sample, empregam-se técnicas de validação temporal avançada, tais como *time series cross-validation* com bloqueio por períodos e backtesting walk-forward, que previnem o *look-ahead bias* e fornecem indicadores realistas de acurácia (BERGMEIR; BENÍTEZ, 2012). Adicionalmente, modelos de mudança de regime (*regime-switching*), como o proposto por Hamilton (1989), capturam transições abruptas

entre fases de mercado (e.g., bull vs. bear), ajustando a sensibilidade do algoritmo conforme o regime vigente ([HAMILTON, 1989](#)).

No contexto deste trabalho, integrar todos esses elementos temporais — sazonalidades, janelas de análise otimizadas, validação avançada, detecção de regimes e embeddings de tempo — torna o modelo mais sensível tanto a movimentos rápidos quanto a evoluções estruturais, garantindo previsões mais estáveis e acuradas em diferentes momentos da sessão de negociação.

2.2.3 Séries Temporais e a Aplicação de Deep Learning em Finanças

A modelagem de dados financeiros frequentemente se baseia no conceito de séries temporais, que consistem em uma sequência de observações coletadas em intervalos de tempo regulares. Uma série temporal é mais do que uma simples lista de números. A sua característica fundamental é a dependência temporal, onde o valor de uma observação em um determinado momento pode ser influenciado por valores passados. No contexto da bolsa de valores, uma série temporal pode ser composta pelos preços de fechamento diário de uma ação, pelo volume de negociações a cada minuto ou por qualquer outra métrica que evolua ao longo do tempo.

Para ilustrar, considere a Tabela 1, que apresenta uma série temporal multivariada hipotética para a ação PETR4, com dados intradiários coletados a cada 15 minutos. Cada linha representa um ponto no tempo, e as colunas (variáveis) representam diferentes métricas financeiras, como os preços de abertura, máxima, mínima, fechamento e o volume negociado. A análise dessa série busca identificar padrões, tendências e sazonalidades que possam ser utilizados para prever os movimentos futuros dos preços.

Tabela 1 – Exemplo de Série Temporal Multivariada da Ação PETR4 (intervalo de 15 minutos)

Timestamp	Abertura (R\$)	Máxima (R\$)	Mínima (R\$)	Fechamento (R\$)	Volume
2025-08-18 10:00	30,50	30,65	30,48	30,62	1.250.000
2025-08-18 10:15	30,62	30,70	30,55	30,68	890.000
2025-08-18 10:30	30,68	30,69	30,50	30,52	1.100.000
2025-08-18 10:45	30,52	30,58	30,45	30,55	950.000

A modelagem de séries temporais financeiras com aprendizado profundo avançou significativamente, com o surgimento de arquiteturas especializadas capazes de lidar com dependências complexas e padrões não estacionários. A seguir, detalham-se duas das arquiteturas mais influentes neste domínio: LSTM e Transformer.

2.2.3.1 Redes Recorrentes Long Short-Term Memory (LSTM)

As Redes Neurais Recorrentes (RNNs) foram projetadas para lidar com dados sequenciais, possuindo "loops" que permitem que a informação persista ao longo do tempo. Contudo, as RNNs tradicionais enfrentam o problema do desaparecimento do gradiente (vanishing gradient),

o que as impede de aprender dependências de longo prazo, uma característica essencial para séries temporais financeiras.

Para superar essa limitação, Hochreiter e Schmidhuber (1997) propuseram a arquitetura Long Short-Term Memory (LSTM), uma variante sofisticada de RNN. A principal inovação da LSTM é sua unidade de memória, composta por um estado de célula (cell state) e três portões (gates) que regulam o fluxo de informação. O estado de célula atua como uma esteira rolante, transportando informações relevantes ao longo da sequência com poucas alterações. Os portões são estruturas neurais que, por meio de funções de ativação sigmoide, decidem qual informação deve ser adicionada ou removida do estado de célula.

A Figura 6 ilustra o funcionamento de uma célula LSTM. Seus componentes são:

- **Portão de Esquecimento (Forget Gate):** Decide qual informação do estado de célula anterior (C_{t-1}) deve ser descartada. Ele analisa a entrada atual (X_t) e o estado oculto anterior (h_{t-1}) para gerar um valor entre 0 (esquecer completamente) e 1 (manter completamente).
- **Portão de Entrada (Input Gate):** Determina quais novas informações serão armazenadas no estado de célula. Ele é composto por duas partes: uma camada sigmoide que decide quais valores serão atualizados e uma camada tanh que cria um vetor de novos valores candidatos.
- **Portão de Saída (Output Gate):** Define qual será a saída da célula. A saída é baseada no estado da célula filtrado, que passa por uma camada tanh para normalização e, em seguida, é multiplicado pela saída da camada sigmoide do portão.

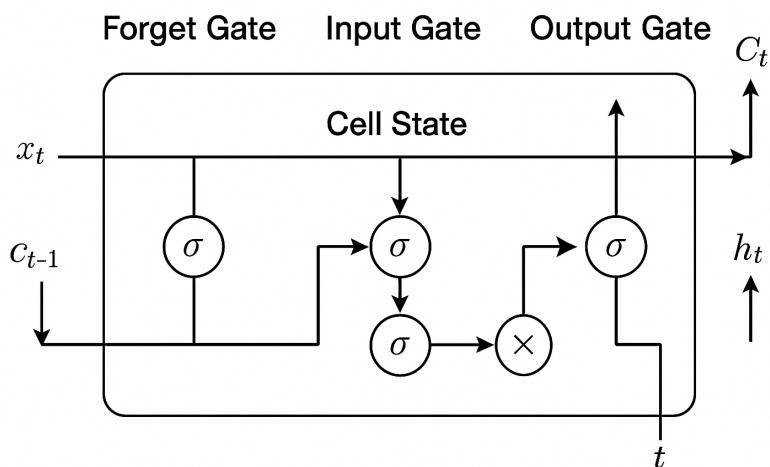


Figura 6 – Arquitetura de uma célula de memória LSTM, ilustrando o estado de célula e os três portões (esquecimento, entrada e saída) que regulam o fluxo de informação. Fonte: Adaptado de (OLAH, 2015).

Essa estrutura de portões permite que a LSTM aprenda a reter informações por longos períodos, tornando-a particularmente eficaz na previsão de cotações, onde eventos passados podem influenciar significativamente os preços futuros.

2.2.3.2 Arquitetura Transformer e o Mecanismo de Atenção

Apesar de seu sucesso, as LSTMs e outras redes recorrentes possuem uma limitação inerente: seu processamento é sequencial. Isso significa que, para processar o décimo ponto de uma série, é preciso antes ter processado os nove anteriores, o que dificulta a paralelização e a captura de dependências entre pontos muito distantes na sequência.

Para resolver essas questões, (VASWANI *et al.*, 2017) introduziram a arquitetura Transformer, que abandona a recorrência e se baseia inteiramente no mecanismo de atenção (attention mechanism). A ideia central da atenção, especificamente da auto-atenção (self-attention), é permitir que o modelo pondere a importância de todas as outras posições da sequência de entrada ao processar uma única posição. Em vez de passar a informação sequencialmente, o Transformer pode "olhar" para toda a série temporal de uma só vez e identificar quais pontos do passado são mais relevantes para prever o futuro.

A arquitetura do Transformer, ilustrada na Figura 7, é composta por dois blocos principais: o Codificador (Encoder) e o Decodificador (Decoder). Seus componentes-chave incluem:

- **Auto-Atenção Multi-Cabeças (Multi-Head Self-Attention):** Em vez de calcular a atenção uma única vez, o modelo o faz múltiplas vezes em paralelo (as "cabeças"), permitindo que ele se concentre em diferentes aspectos da sequência. Por exemplo, uma cabeça pode focar na tendência de curto prazo, enquanto outra foca em padrões sazonais de longo prazo.
- **Codificação Posicional (Positional Encoding):** Como o modelo não processa os dados em ordem, a informação sobre a posição de cada ponto na sequência é adicionada aos dados de entrada. Isso garante que a ordem cronológica da série temporal seja preservada.
- **Camadas Feed-Forward:** Após as camadas de atenção, cada posição passa por uma rede neural feed-forward para processamento adicional.

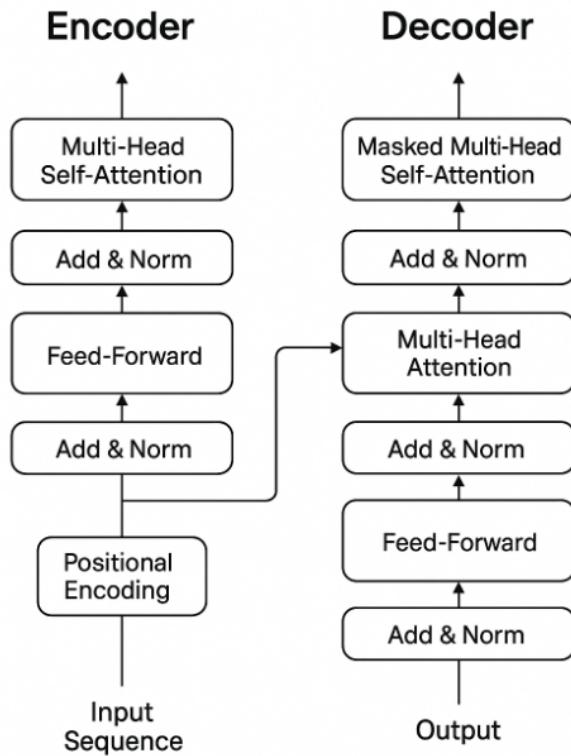


Figura 7 – A arquitetura do modelo Transformer, destacando os blocos de codificador e decodificador, as camadas de atenção multi-cabeças e as redes feed-forward. Fonte: Adaptado de Vaswani et al. (2017) ([VASWANI et al., 2017](#)).

A aplicação de Transformers em finanças é recente, porém bastante promissora. Modelos como o Temporal Fusion Transformer (TFT) proposto por ([LIM; ZOHREN; ROBERTS, 2021](#)), combinam múltiplas fontes de dados e camadas de atenção interpretáveis, adequando-se a dados financeiros complexos. Estudos recentes como esse, mostram que, em ambientes de alta frequência, os Transformers podem superar as LSTMs ([WU et al., 2020](#)), principalmente devido à sua capacidade de modelar dependências de longo prazo de forma mais eficaz e estável .

Assim, a escolha arquitetural deve ser orientada não apenas pela natureza sequencial dos dados, mas também pelas características operacionais do mercado alvo, sendo o Transformer e suas variantes uma tendência crescente na modelagem de séries temporais financeiras.



TRABALHOS RELACIONADOS

A literatura sobre *forecasting* de séries temporais financeiras é vasta, com abordagens que evoluíram dos modelos estatísticos clássicos para arquiteturas híbridas integrando *machine learning*, redes neurais e indicadores técnicos.

3.1 Modelos Clássicos e Estatísticos

Os modelos estatísticos tradicionais, especialmente os da família ARIMA (Autoregressive Integrated Moving Average), baseiam-se na metodologia Box–Jenkins ([BOX et al., 2015](#)), integrando componentes autoregressivo (AR), de média móvel (MA) e de diferenciação (I) para lidar com séries não estacionárias.

[Rotela Jr. et al. \(JR; OUTROS, 2014\)](#) ajustaram o modelo ARIMA para previsão usando uma longa série histórica mensal do Ibovespa (1995–2013), otimizando em relação ao erro MAPE, e obtiveram valores mais baixos, o que indica um bom ajuste nos dados agregados.

Outros estudos comparativos indicam que a precisão do ARIMA se deteriora à medida que o horizonte de previsão aumenta, enquanto outros modelos, como o LSTM, são mais robustos em cenários de janela mais longa e não lineares ([SOUZA et al., 2022](#)).

A principal limitação do ARIMA reside em sua premissa de que todas as informações relevantes estão contidas em valores passados da própria série — o que ignora fontes externas, como indicadores técnicos, choques exógenos ou eventos de mercado súbitos. Além disso, sua sensibilidade a pontos de inflexão subestima impactos abruptos como crises e anúncios de política econômica — situações frequentes no contexto da B3 ([B3, 2025](#)).

Assim sendo, embora o ARIMA forneça boa interpretabilidade e seja eficiente para previsões de curto prazo em séries estacionárias, ele revela fragilidade frente à alta volatilidade, dependências não lineares e dinâmicas multivariadas típicas do mercado acionário brasileiro. Desse modo, para o ambiente da B3 — marcado por ruído, sazonalidade e eventos exógenos

— esses modelos demonstram limitações relevantes que justificam a exploração de arquiteturas mais sofisticadas e híbridas.

3.2 Aprendizado de Máquina e Modelos Híbridos

Métodos de previsão híbridos unem instrumentos estatísticos tradicionais (por exemplo, ARIMA) com algoritmos de aprendizado de máquina ou aprendizado profundo (como LSTM, XGBoost), a fim de colher os benefícios das características lineares e não lineares para dados de séries temporais financeiras.

Estes são alguns exemplos do padrão de complementação demonstrado em vários estudos recentes. Gonçalves (2023) analisou um modelo híbrido onde foram comparados modelos ARIMA e LSTM aplicados nos preços diários do índice IBX50 no Brasil, entre 2012 e 2022. ARIMA oferece boa previsão fora da amostra para o período próximo ao treinamento, mas LSTM é sempre melhor para horizontes de previsão mais longos (GONÇALVES, 2023). Já Kashif & Šlepaczuk (2024) propuseram o modelo LSTM-ARIMA, onde os resíduos gerados pelo ARIMA são incorporados como entrada adicional à LSTM, resultando em ganhos consistentes em múltiplos índices de ações globais (KASHIF MUHAMMAD E ŠLEPACZUK, 2024). Além disso, Nasir et al. (2025) introduziram uma estrutura híbrida que combina ARIMA, decomposição LMD (Local Mean Decomposition) e algoritmos clássicos de ML como XGBoost e Random Forest, demonstrando melhoria significativa em RMSE, MAE e acurácia direcional (NASIR; ZHANG; WANG, 2025).

Quanto aos resultados empíricos, Gonçalves (2023) observou que o desempenho do ARIMA diminui progressivamente conforme o horizonte de previsão se estende, enquanto o LSTM se mostra mais robusto em janelas maiores de predição (GONÇALVES, 2023). No trabalho de Kashif & Šlepaczuk, o modelo híbrido LSTM-ARIMA superou tanto o ARIMA quanto a LSTM isoladamente, alcançando métricas superiores em retorno ajustado ao risco em índices como S&P 500, FTSE 100 e CAC 40 (KASHIF MUHAMMAD E ŠLEPACZUK, 2024). Nasir et al. (2025) relataram que a combinação com LMD + XGBoost incrementou a precisão em torno de 10–15 % em comparação com modelos individuais, além de melhorar a consistência direcional das previsões (NASIR; ZHANG; WANG, 2025).

Do ponto de vista crítico, tais modelos híbridos oferecem vantagens claras, como conseguir aprender tanto as partes lineares — bem captadas pelo ARIMA — quanto as dependências temporais não lineares, padrões sazonais complexos, por meio de LSTM ou XGBoost. Contudo, apresentam alguns desafios relevantes:

- Dependência de grandes volumes de dados para treinar redes profundas, o que pode ser limitante nos ativos menos líquidos da B3.

- Complexidade computacional e necessidade de tuning rigoroso (por exemplo, validação temporal walk-forward), o que exige infraestrutura e expertise avançados.
- Risco de sobreajuste, especialmente em mercados voláteis ou com mudanças estruturais repentinas.

Como ilustra a [Figura 8](#), aplica-se uma validação temporal do tipo walk-forward com embargo opcional h . Em cada *fold* i , o modelo é treinado apenas com dados passados e tem seus hiperparâmetros selecionados por busca bayesiana (Optuna) em uma validação interna contida no bloco de treino. Em seguida, impõe-se um período de embargo de h observações entre o fim do treino/validação e o início do conjunto de teste para mitigar vazamento de informação. O desempenho é então medido no próximo bloco temporal (*Teste* $i+1$), estritamente posterior. O procedimento se repete ao longo do tempo (folds $1, \dots, k$), produzindo uma sequência de avaliações fora da amostra; as métricas finais são agregadas sobre os blocos de teste, e o modelo pode ser retreinado no maior prefixo disponível com os hiperparâmetros selecionados. Esse protocolo é crucial para séries financeiras e, em especial, para modelos híbridos, pois respeita a causalidade temporal e reduz sobreajuste.



Figura 8 – Validação temporal walk-forward com seleção bayesiana de hiperparâmetros. Fonte: autor.

Dada a alta volatilidade, sazonalidade intradiária e nível de impacto exógeno da B3, incluindo política monetária e crises regionais ([HASBROUCK, 2007](#); [ANDERSEN; BOLLERS-LEV, 1997](#)), modelos híbridos podem se tornar preferíveis, mas apenas com boa calibração e validação em seu próprio intervalo de tempo. Eles são superiores aos clássicos em trajetórias de médio e longo prazo e permitem captar padrões que são inalcançáveis para métodos lineares. Deve-se mostrar timidez na transformação de estratégias de investimento, avaliação de risco, estabilidade fora da amostra etc.

Em síntese, os modelos híbridos representam uma evolução lógica em relação aos modelos puramente estatísticos, conciliando interpretabilidade e robustez, e surgem como abordagem promissora para aplicações em predição de séries financeiras na B3.

3.3 Redes Neurais Profundas

As redes neurais profundas, especialmente aquelas com estrutura recorrente como as LSTM (Long Short-Term Memory), foram idealizadas por Hochreiter e Schmidhuber ([HOCHREITER; SCHMIDHUBER, 1997](#)). Essas arquiteturas superam os desafios do desvanecimento/explosão de gradiente em sequências longas por meio do “carrossel de erro” (constant error carousel), permitindo que a rede capte dependências tardias essenciais ([HOCHREITER; SCHMIDHUBER, 1997](#)).

Em aplicações financeiras, as LSTM têm sido amplamente adotadas para previsão de séries temporais não estacionárias. Estudos como Siami-Namini & Siami Namin (2018) demonstraram que LSTM supera ARIMA de forma consistente, reduzindo a taxa de erro em cerca de 84–87% nos dados analisados ([SIAMI-NAMINI; NAMIN, 2018](#)). Além disso, Fjellström (2022) utilizou um ensemble de LSTMs paralelos para classificação binária de movimentos de preço, obtendo resultados mais estáveis com menor volatilidade de previsão ([FJELLSTRÖM, 2022](#)).

Mais recentemente, Kabir et al. (2025) propuseram o modelo híbrido LSTM-mTrans-MLP, integrando LSTM com módulos Transformer e perceptron multicamada. Essa abordagem demonstrou desempenho robusto em diversos ativos (Bitcoin, índices chineses e ações globais), com melhorias substanciais em precisão e sensibilidade ([KABIR; OUTROS, 2025](#)).

Quanto aos resultados práticos, o trabalho original de Hochreiter & Schmidhuber (1997) mostrou que LSTM pode resolver tarefas com atrasos de tempo superiores a mil steps e aprender significativamente mais rápido que redes recorrentes tradicionais ou Elman nets ([HOCHREITER; SCHMIDHUBER, 1997](#)). No estudo de Siami-Namini, observa-se que a LSTM reduz drasticamente os erros em relação ao ARIMA — especialmente em mercados voláteis. Fjellström evidenciou que o ensemble de LSTMs reduz a variabilidade nos resultados, favorecendo previsões diretas de movimento de preço mais estáveis. Já Kabir et al. apresentaram melhoria de desempenho mesmo em mercados altamente não lineares e ruidosos ([KABIR; OUTROS, 2025](#)).

Crucialmente, as redes LSTM são uma escolha particularmente boa para séries temporais financeiras, pois podem modelar dependências não lineares e memória de longo prazo melhor do que RNNs convencionais. No entanto, elas exigem:

- Aprendizado em Lote: Um grande volume de dados é usado para jogar pelo seguro, porque a B3 possui valores mobiliários com baixa liquidez e o overfitting pode se tornar um problema.
- Ajustes computacionais pesados e ajuste fino de hiperparâmetros (portas, taxas de aprendizado, etc.) requerem uma infraestrutura robusta.
- Risco em termos de perda de estabilidade contra dados ruidosos ou eventos exógenos inesperados.

No contexto da B3 (alta volatilidade, sazonalidade intradiária e relação com eventos macroeconômicos), as LSTMs podem ter um bom desempenho, desde que seja utilizada validação temporal suficiente para evitar o overfitting, especialmente em um ambiente de conjunto ou com módulos Transformer.

Redes neurais profundas, especialmente LSTM e o modelo híbrido com variante Transformer mostrado neste trabalho, representam uma ampla evolução em relação aos modelos clássicos por duas razões principais: elas podem capturar relações temporais complexas a partir de grandes conjuntos de dados; e provam previsões mais estáveis em um ambiente de mercado real.

3.4 Transformers para Séries Temporais

Os modelos baseados em Transformer, originalmente propostos por Vaswani et al. (2017), têm sido adaptados para séries temporais graças ao mecanismo de atenção que captura dependências de longo alcance e interações multivariadas de forma eficiente ([WEN et al., 2022](#)).

No contexto financeiro e de séries temporais, trabalhos como Costa & Machado (2023) aplicaram o Transformer à predição de preços de ações do Ibovespa e concluíram que o modelo superou ARIMA e LSTM em cerca de 60% dos casos testados ([COSTA; MACHADO, 2023](#)). Em estudos mais gerais, Wen et al. (2022) apresentam uma revisão sistemática dos Transformers para análise temporal, destacando suas forças e limitações na parte estrutural e de aplicação ([WEN et al., 2022](#)). Outro avanço significativo foi o modelo Non-stationary Transformer (Liu et al., 2022), que propõe módulos de “series stationarization” e “de-stationary attention” para lidar eficazmente com dados não estacionários, reduzindo o MSE em até 49% em relação a outras variantes como Informer e Reformer ([LIU et al., 2022](#)).

Além disso, Nie et al. (2022) introduziram o PatchTST, que segmenta séries em "patches" temporais canal-independentes para reduzir complexidade e melhorar previsões de longo prazo, superando modelos Transformer convencionais em benchmarks públicos ([NIE et al., 2023](#)).

Os resultados empíricos demonstram que:

- Costa & Machado conseguiram maior precisão do Transformer em relação a modelos clássicos (ARIMA e LSTM) em séries do mercado brasileiro ([COSTA; MACHADO, 2023](#)).
- Liu et al. reduziram MSE em 49% comparando com Transformers tradicionais, Informer e Reformer, na tarefa de prever séries não estacionárias ([LIU et al., 2022](#)).
- Nie et al. relataram que o PatchTST oferece previsões robustas e mais acurácia em longo horizonte com menor carga computacional ([NIE et al., 2023](#)).

Sob análise crítica, os Transformers apresentam pontos fortes e limitações:

- Vantagem clara na modelagem de dependências de longo prazo e interações multivariadas sem recorrer à recursividade.
- Risco de perda de informação temporal devido à invariância à permutação na autoatenção, especialmente se o encoding posicional não for bem calibrado ([ZENG et al., 2022](#)).
- Complexidade computacional elevada, com necessidade de variantes modificadas como Informer, Autoformer ou PatchTST para tornar o modelo escalável em grandes horizontes temporais.
- Em cenários com títulos de baixa liquidez na B3, os Transformers exigem volume substancial de dados, o que pode ser desafiador.

Em síntese, os Transformers representam uma evolução promissora para o mercado da B3 pela capacidade de capturar padrões complexos e não estacionários. Entretanto, exigem adaptações específicas, validação robusta e infraestrutura computacional adequada para realmente superar métodos como LSTM e modelos híbridos em cenários reais da bolsa brasileira.

3.5 LLMs atuais aplicados a Séries Temporais Financeiras

Recentemente, os *Large Language Models* (LLMs) passaram a ser explorados como ferramentas de previsão de séries temporais financeiras, aproveitando sua capacidade de raciocínio sequencial e contextual. Em vez de operar diretamente sobre séries numéricas, esses modelos são adaptados por meio de técnicas como transformação de padrões temporais em tokens de linguagem, aprendizado em poucos exemplos (*few-shot*) ou inferência sem ajuste (*zero-shot*) ([ALONSO; FRANKLIN, 2024](#)).

Um estudo conduzido por Noguer I Alonso & Franklin (2024/25) testou modelos como TimeGPT, NBEATS, NHITS, PatchTST e KAN em dados de ações de empresas globais (Google, Apple, Meta, entre outras), mostrando que variantes baseadas em LLMs, especialmente TimeGPT Long Horizon, tiveram desempenho superior em cenários estáveis, enquanto modelos especializados como PatchTST se destacaram em condições mais voláteis ([ALONSO; FRANKLIN, 2024](#)).

Jiang et al. (2025), por meio do framework LLM-PS, integraram módulos de convolução multi-escala e um componente de “time-to-text”, permitindo ao LLM captar padrões semânticos e temporais de séries não estacionárias. O método obteve precisão de ponta tanto em previsões de curto quanto longo prazo, em regimes com poucos dados ou sem ajuste fino ([TANG et al., 2025](#)).

Além disso, trabalhos como LLM4TS (Chang et al., 2023) apresentaram esquema de alinhamento e fine-tuning adaptativo de LLMs pré-treinados para séries temporais, alcançando redução média de 6,8% no MSE em cenários few-shot em comparação com modelos treinados do zero (CHANG *et al.*, 2023). Jin et al. (2023) criaram a arquitetura Time-LLM, que reprograma LLMs em tarefas de forecasting unificando prompts como prefixos, e demonstrou desempenho competitivo frente a modelos especializados, inclusive em cenários zero-shot (JIN *et al.*, 2023).

Mais recentemente, o modelo **DeepSeek-TS+** foi proposto como uma estrutura unificada para previsão de séries multivariadas com múltiplos produtos, integrando atenção latente multi-cabeça (MLA) com uma política adaptativa de otimização por grupos (GRPO) (LI *et al.*, 2025). Essa arquitetura apresentou ganhos significativos sobre GRUs e modelos ARMA em ambientes dinâmicos, com melhor capacidade de adaptação e robustez temporal. Em paralelo, o modelo **SMETimes**, baseado em LLMs leves com menos de 3 bilhões de parâmetros, demonstrou superioridade em relação a modelos maiores (7B), tanto em desempenho quanto em economia de recursos, sugerindo que arquiteturas otimizadas podem ser mais eficazes em contextos como o da B3 (FAN; OUTROS, 2025).

Os resultados indicam que:

- LLMs são eficazes em cenários com poucos dados (few-shot ou zero-shot), sendo úteis quando séries são curtas ou ativos têm baixa liquidez.
- Frameworks como LLM-PS e LLM4TS aprimoram o modelo com módulos que capturam padrões temporais e escrevem sequências numéricas em linguagem natural, resultando em previsão mais contextualizada.
- Modelos como Time-LLM e DeepSeek-TS+ oferecem versatilidade e capacidade adaptativa robusta, inclusive em cenários com múltiplas variáveis e dinâmicas não lineares.
- LLMs otimizados, como o SMETimes, viabilizam previsões precisas com menor custo computacional — um diferencial importante em aplicações reais.

Embora essas abordagens mostrem potencial sob escrutínio, elas também apresentam algumas deficiências:

- A conversão de dados numéricos em linguagem pode criar abstração de informações e perda de precisão sem a tokenização adequada.
- O desempenho tende a decair em séries sem padrão claro (ruído elevado, ausência de sazonalidade ou tendência), conforme evidenciado por Tang et al. (2024) (TANG; OUTROS, 2024).
- Extrapolar para alta frequência, uma característica comum do B3, exige uma reformulação em grande escala do pipeline de tokenização e das representações temporais.

Em conclusão, os LLMs otimizados fundidos e de cabeça dupla com módulos dinâmicos temporais especializados podem refletir um método promissor para o tipo de problema 3 com desequilíbrio extremo censurado e dados de séries temporais esparsas que poderiam ser complementares às abordagens típicas e híbridas ao lidar com o B3.

3.6 Síntese Crítica

Ao revisar os trabalhos relacionados, observa-se que a literatura documenta uma evolução das abordagens de previsão de séries financeiras, indo de modelos estatísticos clássicos a arquiteturas profundas e híbridas. Modelos tradicionais como a família ARIMA permanecem relevantes pela interpretabilidade e pelo desempenho observado em séries aproximadamente estacionárias (BOX *et al.*, 2015; HYNDMAN; ATHANASOPOULOS, 2018; JR; OUTROS, 2014). Por outro lado, há relatos de limitações quando a dinâmica é fortemente não linear, multivariada ou sujeita a mudanças de regime e elevada volatilidade (MAKRIDAKIS; SPILIOOTIS; ASSIMAKOPOULOS, 2018; DERBENTSEV, 2020).

Abordagens híbridas que combinam modelos estatísticos com técnicas de aprendizado de máquina e profundo têm sido associadas, em determinados cenários, a ganhos de robustez e desempenho fora da amostra, ao custo de maior complexidade de modelagem e sensibilidade a hiperparâmetros (GONÇALVES, 2023; KASHIF MUHAMMAD E ŠLEPACZUK, 2024; NASIR; ZHANG; WANG, 2025). A literatura também recomenda validação temporal cuidadosa (por exemplo, esquemas específicos para séries) e procedimentos sistemáticos de ajuste para mitigar sobreajuste (BERGMEIR; BENÍTEZ, 2012; PRADO, 2018; AKIBA *et al.*, 2019).

No campo das redes neurais, LSTM e variantes correlatas têm sido empregadas para capturar dependências de longo prazo e padrões não lineares em séries financeiras, com resultados competitivos em diferentes horizontes de previsão (HOCHREITER; SCHMIDHUBER, 1997; SIAMI-NAMINI; NAMIN, 2018; FJELLSTRÖM, 2022). Há ainda propostas de ensembles e modelos híbridos que integram módulos recorrentes e atencionais (KABIR; OUTROS, 2025). Em termos práticos, a necessidade de dados e de recursos computacionais pode ser um fator restritivo em ativos de baixa liquidez, aspecto discutido na literatura de microestrutura e liquidez (AMIHUD, 2002).

Arquiteturas baseadas em *Transformers* vêm sendo adaptadas ao domínio temporal e reportam resultados competitivos em horizontes longos e séries multivariadas, incluindo estudos no contexto brasileiro (WEN *et al.*, 2022; COSTA; MACHADO, 2023). Diversas variantes propõem mecanismos de estabilização e estratégias para lidar com não estacionariedade e eficiência computacional, como módulos de “stationarization” e atenção modificada (LIU *et al.*, 2022) ou segmentação em *patches* temporais (NIE *et al.*, 2023). Ao mesmo tempo, análises críticas questionam condições em que tais modelos de atenção superam abordagens alternativas, sugerindo cautela na generalização de resultados (ZENG *et al.*, 2022).

Mais recentemente, modelos de linguagem de grande escala (LLMs) foram explorados como apoio à previsão de séries temporais por meio de reprogramação de tarefas, *few-shot* e integrações que capturam padrões semânticos e temporais (LATHUILÈRE *et al.*, 2024; CHANG *et al.*, 2023; JIN *et al.*, 2023; ALONSO; FRANKLIN, 2024; TANG *et al.*, 2025). A literatura também discute sensibilidades relacionadas à tokenização, alinhamento temporal e custos computacionais, recomendando atenção a limitações e preferências dos modelos nessas tarefas (TANG; OUTROS, 2024).

Em síntese, os estudos revisados não apontam consenso em torno de um modelo único e universalmente superior para os desafios de previsão no mercado acionário, especialmente em ambientes voláteis e sujeitos a eventos exógenos (MAKRIDAKIS; SPILIOOTIS; ASSIMAKOPOULOS, 2018; WEN *et al.*, 2022). Nesse cenário, linhas de pesquisa sobre soluções híbridas e adaptativas aparecem como promissoras para conciliar interpretabilidade, flexibilidade e desempenho. Este trabalho, portanto, investigará uma abordagem que integra paradigmas estatístico, profundo e atencional, com foco em estabilidade, precisão e aplicabilidade prática a cenários representativos do mercado brasileiro.

Para uma visão de conjunto dos grupos de modelos discutidos, ver as Tabelas 2 a 5.

Tabela 2 – Método vs. Dados exigidos. Fonte: o autor.

Método	Dados exigidos
Modelos estatísticos (ARIMA/ARIMAX)	Série univariada; estacionária (ou diferenciada); histórico razoável ($\approx 100\text{--}300+$ pontos); regressores opcionais via ARIMAX.
LSTM	Muitos dados; normalização; janelas de <i>lookback</i> ; <i>tuning</i> de hiperparâmetros.
Transformers (PatchTST, Non-stationary)	Janelas longas; multivariadas; melhor com grande volume de dados. <i>Patching</i> reduz <i>tokens</i> ; módulos de estacionarização ajudam.
Híbridos (ARIMA+LSTM; LMD+XGBoost)	Dados para ambos os blocos; etapas de decomposição e resíduos; <i>pipeline</i> mais complexo.
LLMs s/ séries (Time-LLM, LLM4TS, Zero-shot)	Pré-processo <i>time-to-text/prompting</i> ; poucos dados funcionam (<i>few/zero-shot</i>); sem treino pesado.

Tabela 3 – Método vs. Custo computacional. Fonte: o autor.

Método	Custo computacional
Modelos estatísticos (ARIMA/ARIMAX)	Baixo (CPU). Ajuste Box–Jenkins/auto-ARIMA.
LSTM	Médio–alto (GPU recomendável).
Transformers (PatchTST, Non-stationary)	Alto (atenção quadrática). Variantes reduzem custo.
Híbridos (ARIMA+LSTM; LMD+XGBoost)	Alto (múltiplos modelos/etapas).
LLMs s/ séries (Time-LLM, LLM4TS, Zero-shot)	Alto na inferência; baixo/zero custo de treino adicional; modelos menores mitigam.

Tabela 4 – Método vs. Pontos fortes. Fonte: o autor.

Método	Pontos fortes
Modelos estatísticos (ARIMA/ARIMAX)	Interpretável; bom curto prazo; <i>baseline</i> sólido.
LSTM	Capta dependências longas e não-linearidades; bom para múltiplas <i>features/indicadores</i> .
Transformers (PatchTST, Non-stationary)	Modela dependências de longo alcance; bom em horizontes longos e séries multivariadas.
Híbridos (ARIMA+LSTM; LMD+XGBoost)	Combina linear e não-linear; ganhos consistentes em métricas e acurácia direcional.
LLMs s/ séries (Time-LLM, LLM4TS, Zero-shot)	Generalizam com pouco dado; integram sinais textuais (notícias) e numéricos.

Tabela 5 – Método vs. Fragilidades. Fonte: o autor.

Método	Fragilidades
Modelos estatísticos (ARIMA/ARIMAX)	Não-linearidades e quebras de regime; piora em horizontes longos; multivariadas limitadas.
LSTM	Exige dados; risco de <i>overfitting</i> ; sensível a mudanças de regime sem validação temporal.
Transformers (PatchTST, Non-stationary)	Custo elevado; risco de má codificação temporal; debate sobre ganhos vs. modelos lineares simples.
Híbridos (ARIMA+LSTM; LMD+XGBoost)	Complexidade; <i>tuning</i> pesado; risco de sobreajuste; manutenção mais difícil.
LLMs s/ séries (Time-LLM, LLM4TS, Zero-shot)	Tokenização pode perder precisão; maturidade menor; desempenho cai em ruído extremo/alta frequência.



METODOLOGIA

Este capítulo detalha os procedimentos adotados para desenvolver e avaliar o modelo proposto. A metodologia está organizada nas Seções 4.1 a 4.8, cobrindo: aquisição e tratamento de dados, pré-processamento e engenharia de atributos, modelos (baselines e proposto), desenho experimental e treinamento, métricas de avaliação, backtests com custos, testes de robustez e significância e, por fim, diretrizes de reproduzibilidade.

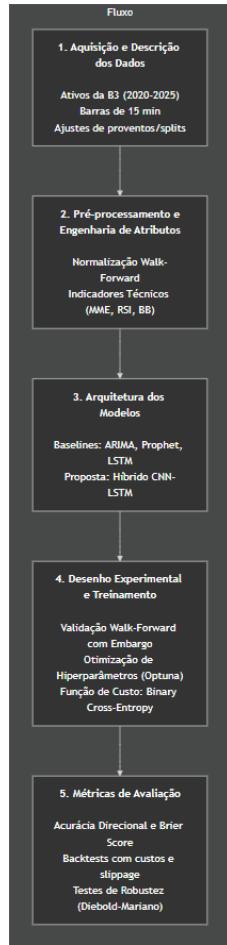


Figura 9 – Guia adotado para construção de modelo. Fonte: o autor.

4.1 Aquisição e descrição dos dados

A seleção do conjunto de dados e o controle de qualidade são fatores centrais para obter previsões intradiárias confiáveis. Para reduzir vieses recorrentes em bases financeiras, adotaram-se três diretrizes: (i) mitigação de *survivorship bias*, (ii) ajustes por eventos corporativos e (iii) consistência temporal do histórico. A fim de evitar vazamento de informação entre janelas e estimar o desempenho de maneira mais realista, o período foi organizado em partições temporais sequenciais, com embargo entre treino e teste, conforme recomendações para séries financeiras (PRADO, 2018; BERGMEIR; BENÍTEZ, 2012). Adicionalmente, foram respeitadas as regras operacionais e as fases de negociação da B3 ao longo do intervalo analisado (B3, 2025).

4.1.1 Fonte e universo

Os dados foram obtidos via plataforma MetaTrader 5 (MT5), com histórico consolidado e *timestampling* consistente no fuso do pregão (B3, 2025a). O universo contempla **três ativos** de alta liquidez negociados na B3: PETR4 (Petrobras), VALE3 (Vale) e ITUB4 (Itaú Unibanco). Esses ativos foram mantidos fixos durante todo o período para mitigar *survivorship bias*. O

critério de liquidez busca reduzir *spreads*, *gaps* e *slippage*, além de tornar as métricas mais estáveis em *backtests*.

4.1.2 Recorte temporal e granularidade

Foram utilizadas barras de **15 minutos** (timeframe M15). O período efetivo compreende **22 de outubro de 2020** a **22 de outubro de 2025** (cinco anos), condicionado pela disponibilidade de histórico no MT5 para a granularidade selecionada ([ANDERSEN; BOLLERSLEV, 1997](#)).

4.1.3 Variáveis

Cada barra contém as variáveis padrão de preço e volume: **OHLCV** — abertura (*open*), máxima (*high*), mínima (*low*), fechamento (*close*) e volume em contratos (*volume_real*). Incluem-se ainda *volume_ticks* (número de ticks por barra) e *spread* (diferença bid-ask no *timestamp*), empregadas em checagens de qualidade e na análise de microestrutura. As *features* adicionais (retornos logarítmicos, volatilidade móvel, MME, RSI, Bandas de Bollinger) são construídas na etapa de engenharia de atributos (Seção 4.2).

4.1.4 Tratamento de qualidade

As séries foram ajustadas por desdobramentos e **grupamentos** (*reverse splits*) e por proventos (dividendos, JCP e bonificações). Também foram verificados o fuso horário e os horários de pregão; em seguida, removeram-se barras com registros anômalos (preço ou volume ausentes, valores zerados e *spikes* espúrios). Para atenuar efeitos de microestrutura — como *bid-ask bounce* e periodicidades intradiárias — adotou-se uma banda morta na rotulagem (threshold de 0,1%) e checagens de consistência ([HASBROUCK, 2007](#); [ANDERSEN; BOLLERSLEV, 1997](#)).

4.1.5 Estatística descritiva do dataset

A Tabela 6 resume o volume de dados por ativo e no consolidado. Observa-se um total em torno de 36 mil barras por ativo, o que corresponde a aproximadamente 1 302 dias de pregão (considerando 28 barras de 15 min por dia útil). O arquivo consolidado (`todos_ativos_M15`) totaliza 109 432 registros.

4.2 Pré-processamento e engenharia de atributos

Nesta etapa, o foco é transformar o OHLCV intradiário em uma base consistente para treino. Além de melhorar a estabilidade numérica dos modelos, o pré-processamento também reduz riscos clássicos em séries financeiras, especialmente o *data leakage*. Em termos práticos,

Tabela 6 – Estatística descritiva do dataset intradiário (barras de 15 min).

Ativo	Barras	Primeira data	Última data	Dias (aprox.)
PETR4	36 465	22/10/2020	22/10/2025	1 302
VALE3	36 486	22/10/2020	22/10/2025	1 303
ITUB4	36 481	22/10/2020	22/10/2025	1 303
Consolidado	109 432	22/10/2020	22/10/2025	—

Fonte: O autor, a partir de dados MT5.

tudo o que usa estatísticas do conjunto de dados precisa respeitar a ordem do tempo. Caso contrário, o modelo “aprende” padrões que só existiriam com informação do futuro.

4.2.1 Normalização e prevenção de data leakage

As *features* numéricas são normalizadas antes do treinamento dos modelos de deep learning. O método padrão é **Min–Max**, mapeando os valores para a faixa $[0, 1]$. Esse tipo de escala costuma favorecer o treinamento de redes por manter as entradas em magnitude controlada. Como verificação de robustez, mantemos disponível também o **Z-Score**, usado em análises de sensibilidade.

O ponto central não é apenas “normalizar”, mas **quando** normalizar. Em cada etapa do *walk-forward*, o normalizador é ajustado (*fit*) **exclusivamente** no bloco de treino. Em seguida, aplicamos apenas a transformação (*transform*) ao bloco de teste. Isso garante que estatísticas calculadas no futuro (por exemplo, mínimo, máximo, média ou desvio) não contaminem o processo de treino (PRADO, 2018; BERGMEIR; BENÍTEZ, 2012). O mesmo princípio é aplicado a qualquer outra transformação dependente dos dados, como imputação ou tratamento de outliers.

Esse fluxo está organizado como pipeline no script `prepare_sequences.py`. Na prática, ele garante que treino e teste permaneçam isolados no tempo, o que torna os resultados mais próximos de um cenário operacional real.

4.2.2 Atributos calculados

A construção dos atributos parte do OHLCV, mas evita usar preços brutos como entrada direta da rede. O primeiro passo é calcular retornos logarítmicos, que são mais adequados para análise de variação e ajudam a reduzir efeitos de escala entre ativos:

$$r_t = \ln P_t - \ln P_{t-1}.$$

A partir desses retornos e das séries de preços, são extraídos indicadores técnicos que capturam tendência, força relativa e regimes de volatilidade. Os principais atributos utilizados são:

- **Volatilidade realizada:** janela móvel de 20 períodos (desvio-padrão dos retornos).
- **MME** (médias móveis exponenciais): períodos 9, 21 e 50.
- **RSI** (Relative Strength Index): períodos 9, 21 e 50 ([MURPHY, 1999](#)).
- **Bandas de Bollinger:** período 20 e 2 desvios-padrão ([BOLLINGER, 2001](#)). Além das bandas superior, inferior e média, calculamos: (i) a largura (*width*) e (ii) a posição relativa do preço (*position*).

Para os modelos sequenciais (LSTM e CNN–LSTM), a entrada final considera **12 variáveis**: três MME, três RSI, cinco variáveis das Bandas de Bollinger e a volatilidade. Essa escolha busca um equilíbrio: as *features* são ricas o suficiente para capturar padrões locais, mas evitam que a rede dependa diretamente de níveis de preço. Em outras palavras, o modelo tende a aprender mais sobre dinâmica e regime do que sobre “valor nominal”.

O *target* é direcional e utiliza **banda morta** de 0,1%. A motivação é reduzir ruído de microvariações e evitar que movimentos muito pequenos sejam tratados como “sinais” relevantes. A regra de rotulagem é:

- $r_{t+1} > 0,1\% \Rightarrow$ alta (1);
- $r_{t+1} < -0,1\% \Rightarrow$ baixa (-1);
- caso contrário: neutro (0).

Após a remoção das observações neutras, o problema é tratado como classificação binária, mapeando baixa/alta para duas classes (por exemplo, 0/1), compatível com a camada sigmoide. Isso reduz ambiguidade na classe e tende a melhorar a separação entre alta e baixa, ao custo de menor volume de amostras.

4.2.3 Janela temporal (lookback) e trade-off

Os modelos de sequência trabalham com uma **janela deslizante** de histórico. Neste trabalho, adotamos como padrão **60 barras** de 15 minutos, o que equivale a aproximadamente 15 horas de pregão (na prática, próximo de dois dias úteis, dependendo da distribuição ao longo das sessões).

Para cada instante t , a entrada é construída como uma matriz com as últimas 60 observações:

$$\mathbf{X}_t = [\mathbf{x}_{t-60}, \dots, \mathbf{x}_{t-1}],$$

e o *target* associado é a direção do retorno na barra t .

A escolha de 60 passos não é “mágica”; ela representa um compromisso entre fatores que puxam para lados opostos:

- **Contexto:** janelas maiores ajudam a capturar dependências de curto e médio prazo, além de padrões intradiários recorrentes.
- **Ruído e overfitting:** janelas longas demais podem diluir sinais recentes, aumentar a complexidade efetiva e favorecer ajuste a ruídos.
- **Viabilidade amostral:** cada nova sequência consome 60 barras de histórico; quanto maior o *lookback*, menor o número de amostras disponíveis.

No pipeline, o valor padrão está definido em config.py (JANELA_TEMPORAL_STEPS = 60). Além da janela padrão (60), foram avaliadas janelas alternativas (16, 32 e 64 barras) por meio de otimização bayesiana (Optuna), a fim de verificar sensibilidade do desempenho ao *lookback*.

4.3 Modelos - baselines e proposto

Adotamos uma abordagem que combina modelos de referência e arquiteturas modernas, em ordem crescente de complexidade, viabilizando comparações consistentes. A seleção foi guiada por literatura de previsão de séries temporais e pelas particularidades da negociação eletrônica na B3 (B3, 2025; HYNDMAN; ATHANASOPOULOS, 2018; BOX *et al.*, 2015; TAYLOR; LETHAM, 2018; HOCHREITER; SCHMIDHUBER, 1997; BOROVYKH; BOHTE; OOSTERLEE, 2017).

4.3.1 Baselines e justificativa.

Os *baselines* servem para estabelecer um piso de desempenho e verificar se modelos mais complexos trazem ganho real sobre regras simples e métodos clássicos.

- **Naive:** repete a última direção observada (retorno $> 0 \rightarrow$ alta, $< 0 \rightarrow$ baixa); não usa histórico além da barra anterior.
- **Drift:** assume tendência linear (média dos retornos no treino) e propaga essa deriva para obter a direção no passo seguinte; representa um piso um pouco mais informativo que o Naive.
- **ARIMA:** modelo estatístico Box–Jenkins para dependências lineares; a ordem (p, d, q) é escolhida por busca em grade (por exemplo, $p, q \leq 3, d \leq 2$) minimizando AIC nos dados de treino, e o forecast é convertido em sinal direcional (BOX *et al.*, 2015).
- **Prophet:** modelo de decomposição aditiva com sazonalidades diária e semanal; incluído como referência de método moderno de séries temporais, ainda que o contexto intradiário de 15 min coloque em questão a relevância de sazonalidade semanal (TAYLOR; LETHAM, 2018).

- **LSTM puro:** uma camada LSTM seguida de camada densa com sigmoide; baseline de deep learning que usa apenas dependências temporais de longo prazo, sem estágio convolucional (HOCHREITER; SCHMIDHUBER, 1997). A comparação do LSTM puro com o híbrido CNN–LSTM permite avaliar o ganho da extração de padrões locais.

4.3.2 **Modelo proposto: CNN–LSTM.**

O modelo principal é um **híbrido CNN–LSTM** que combina convolução 1D (padrões locais na janela temporal) e LSTM (dependências de maior alcance). A entrada é uma matriz $\mathbf{X} \in \mathbb{R}^{N \times T \times F}$, com N amostras, $T = 60$ passos de tempo e $F = 12$ features. O fluxo é:

- **Conv1D:** 64 filtros, *kernel size* 2, ativação ReLU, aplicado sobre (T, F) — extrai padrões em janelas de 2 barras;
- **MaxPooling1D:** *pool size* 2 — reduz o comprimento temporal pela metade e aumenta robustez a ruído;
- **LSTM:** 50 unidades com *dropout* 0,2 — processa a sequência reduzida e captura dependências de médio prazo;
- **Dense(1, sigmoide):** saída probabilidade de “alta” no próximo intervalo.

O otimizador é AdamW com *learning rate* 0,001 e *gradient clipping* por norma (máx. 1,0); a função de custo padrão é *Binary Cross-Entropy*. A arquitetura está implementada em `cnn_lstm_model.py` e ilustrada na Figura 10.

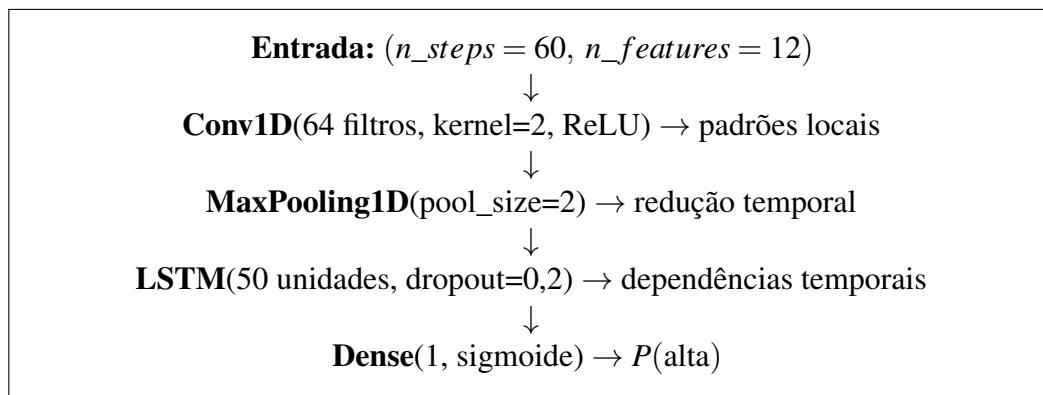


Figura 10 – Arquitetura do modelo híbrido CNN–LSTM (proposta). Fonte: o autor, a partir de `cnn_lstm_model.py`.

4.3.3 **Hiperparâmetros e tuning.**

Os hiperparâmetros principais do CNN–LSTM são: número de **filtros** da Conv1D (32, 64 ou 128), **tamanho do kernel** (2 ou 3), **unidades LSTM** (32, 50 ou 64), **dropout** (0,1; 0,2 ou 0,3), **learning rate** (10^{-4} , 10^{-3} ou 10^{-2}) e **batch size** (16, 32 ou 64). O *pool size* foi mantido

fixo em 2. O **critério de busca** é a otimização bayesiana via Optuna ([AKIBA et al., 2019](#)): em cada etapa do *walk-forward*, o conjunto de treino é dividido em treino interno e validação interna (80%/20%); o Optuna maximiza a acurácia direcional no conjunto de **validação interna**, sem usar o conjunto de teste, evitando vazamento de informação. Os hiperparâmetros vencedores são então usados para treinar um modelo no mesmo par treino/validação interna (80%/20%); **não há re-treino no bloco de treino completo** (100% do fold) — o modelo efetivamente avaliado no teste é o treinado no 80% com os melhores hiperparâmetros. Os espaços de busca estão definidos em `config.py` (`HIPERPARAMETROS_CNN_LSTM`, `HIPERPARAMETROS_LSTM`); os valores padrão (sem Optuna) são 64 filtros, kernel 2, 50 unidades LSTM, dropout 0,2 e learning rate 0,001, em linha com a literatura e com o modelo de referência ([BOROVYKH; BOHTE; OOSTERLEE, 2017](#)).

4.4 Desenho experimental e treinamento

Esta seção descreve como o experimento foi organizado no tempo e como os modelos foram treinados. A regra geral é sempre preservar a ordem cronológica. Ou seja: não se realiza embaralhamento, preservando a ordem temporal. Isso evita leituras “do futuro” e deixa a avaliação mais próxima do uso real.

4.4.1 Protocolo temporal (*walk-forward*) e particionamento

O histórico de cada ativo é dividido em **blocos sequenciais**, seguindo validação *walk-forward* ([PRADO, 2018](#); [BERGMEIR; BENÍTEZ, 2012](#)). Em cada **fold**, usamos:

- **Treino:** cerca de **1 ano** em barras

$$252 \times 28 = 7056 \text{ barras (M15)},$$

considerando 28 barras de 15 min por dia útil (pregão 10:00–17:00).

- **Embargo: 5 barras** (aprox. 2 horas).
- **Teste:** cerca de **1 mês**

$$21 \times 28 = 588 \text{ barras.}$$

O fold seguinte começa logo após o fim do teste do fold anterior. Assim, treino e teste nunca se sobrepõem, e a ordem do tempo é mantida. Com o volume efetivo do dataset (em torno de 36 mil barras por ativo), obtém-se tipicamente **5 folds** por ativo. A implementação desse particionamento está na classe `WalkForwardValidator` (`validation.py`), que gera os intervalos:

- `train_start:train_end`

- embargo_start:embargo_end
- test_start:test_end

O pipeline completo é executado por `rodar_pipeline_completo.sh` para PETR4, VALE3 e ITUB4. Os artefatos são salvos de forma organizada em:

- `models/<ativo>/<modelo>/`
- `data/processed/`
- `logs/training_history/`

4.4.2 Treinamento: épocas, batch, early stopping e sementes

Os modelos de deep learning (CNN–LSTM e LSTM) são treinados por fold, com no máximo **100 épocas**. Para evitar treino desnecessário e reduzir *overfitting*, aplicamos *early stopping*: se a *loss* de validação não melhora por **10 épocas** seguidas, o treino é interrompido e os pesos da melhor época são restaurados. O tamanho do *batch* é **32** por padrão, ou o valor definido pelo Optuna (**16, 32 ou 64**). Também utilizamos recursos comuns para estabilizar o treinamento:

- **ReduceLROnPlateau**: reduz o *learning rate* em 50% após 5 épocas sem melhoria ($min_lr = 10^{-7}$).
- **CosineDecayRestarts**: agenda o decaimento do *learning rate* em ciclos.
- **Gradient clipping**: norma máxima 1,0 no otimizador AdamW.

Para reprodutibilidade, fixamos a semente global em **42** (SEED = 42 em `config.py`). Ela é aplicada a NumPy e TensorFlow no início do processo e na construção dos modelos.

O histórico de cada fold (por época) é registrado em:

`logs/training_history/<ativo>/<modelo>/fold_<N>_history.csv`.

4.4.3 Tuning de hiperparâmetros: por bloco, não aninhado

A busca de hiperparâmetros é feita **por fold**. Ou seja: cada bloco tem seu próprio tuning, respeitando o tempo, mas sem validação cruzada aninhada (que seria mais cara computacionalmente).

Dentro de cada fold, o conjunto de treino é dividido em:

- **80%** para treino interno

- **20%** para validação interna

A otimização bayesiana (Optuna) maximiza a métrica escolhida (por exemplo, acurácia direcional) na validação interna, com um número fixo de **trials** (por exemplo, 20 no `rodar_pipeline_completo.sh`) ([AKIBA et al., 2019](#)).

O melhor conjunto de hiperparâmetros é então usado para treinar o modelo final no mesmo par treino/validação interna (80%/20%). Esse modelo — treinado no 80% do bloco de treino do fold, com os hiperparâmetros escolhidos pelo Optuna — é o que é salvo e usado para prever no **teste** daquele mesmo fold; não se re-treina em 100% do treino após o tuning.

O ponto importante é que o **teste não participa** do tuning. A sequência fica bem definida:

treino → validação interna → embargo → teste

A função de custo adotada é *Binary Cross-Entropy*. O ambiente de execução utiliza Python 3.12, TensorFlow/Keras, Pandas e Scikit-learn. As dependências são gerenciadas por uv e o pipeline é orquestrado por `rodar_pipeline_completo.sh` ([LOSHCHILOV; HUTTER, 2019](#)).

4.5 Métricas de Avaliação

Nesta etapa, a preocupação é dupla. Primeiro, medir se o modelo acerta a direção do movimento, que é o que importa para a decisão. Segundo, entender se as probabilidades produzidas são confiáveis, e não apenas “estimativas probabilísticas pouco calibradas”. Por isso, combinamos métricas de classificação (direção) com métricas probabilísticas (calibração), e complementamos com *backtests* para traduzir acerto estatístico em impacto financeiro.

4.5.1 Métrica principal e justificativa.

A métrica central para comparar os modelos é a **acurácia direcional** (*hit rate*): a proporção de vezes em que a previsão de *alta* ou *baixa* coincide com a direção realizada no período seguinte. Como trabalhamos com dados intradiários, pequenas oscilações podem ser puro ruído de microestrutura. Por isso, a acurácia direcional é calculada **após** aplicar uma **banda morta** na rotulagem. Retornos com valor absoluto abaixo de 0,1% são tratados como neutros e **excluídos** do cálculo da métrica. Na prática, isso aproxima a avaliação do que seria uma decisão real (comprar/vender), e reduz a influência de movimentos muito pequenos, onde o *spread* e o *bid-ask bounce* podem dominar ([HASBROUCK, 2007](#)).

Para dar contexto além do *hit rate*, reportamos também métricas clássicas de classificação: **accuracy**, **Balanced Accuracy**, **F1-score** e **MCC** (Matthews Correlation Coefficient). Essas métricas ajudam quando há desequilíbrio entre classes ou quando a taxa de acerto isolada pode

esconder um modelo que acerta muito um lado e erra o outro. Em especial, o MCC é útil por resumir a matriz de confusão de forma mais estável em cenários desbalanceados.

Por fim, como acerto direcional não garante lucro, avaliamos a utilidade prática com *backtests*. Neles, incluímos custos de transação e *slippage* explícitos. Reportamos então retorno líquido, índice de Sharpe, *max drawdown* e *profit factor* (SHARPE, 1994). A intenção aqui é simples: verificar se um ganho estatístico vira ganho operacional quando atrito de mercado entra na conta.

4.5.2 Métricas probabilísticas e calibração.

Como a saída do modelo é probabilística (camada sigmoide), não basta saber se ele “acertou” a classe. Também importa saber se ele está **bem calibrado**. Um modelo pode acertar bastante, mas ser excessivamente confiante, ou o oposto: prever probabilidades tímidas mesmo quando o sinal é forte. Isso afeta diretamente decisões por limiar (*threshold*) e regras de entrada/saída.

Para medir qualidade probabilística, utilizamos **Log-Loss** e **Brier Score**. Ambos penalizam previsões erradas de forma sensível à confiança: errar com 0,99 “custa” mais do que errar com 0,55. Valores menores indicam previsões mais consistentes e, em geral, melhor calibração (BRIER, 1950).

A calibração é analisada de forma direta com **curvas de confiabilidade** (*reliability diagrams*) e com o **ECE** (Expected Calibration Error) (GUO *et al.*, 2017). A ideia do ECE é quantificar o desvio médio entre probabilidade prevista e frequência observada. Um modelo bem calibrado produz probabilidades interpretáveis: se ele prevê 70%, espera-se que, em média, cerca de 70% daqueles casos aconteçam. Isso é especialmente útil para escolher limiares operacionais de forma racional (e não por tentativa e erro).

4.5.3 Agregação dos resultados entre folds.

As métricas são calculadas **por fold** do *walk-forward*, isto é, em cada bloco de teste. Depois, agregamos os resultados entre folds para obter uma visão mais estável do desempenho. A regra padrão do pipeline é usar a **média** das métricas ao longo dos folds, por ativo e por modelo.

Não usamos mediana ou intervalos de confiança na rotina padrão. Ainda assim, os resultados por fold ficam salvos em CSV (por exemplo, <ativo>_<modelo>_walkforward.csv), com uma linha por fold. Isso facilita análises adicionais, caso seja necessário reportar mediana, IC ou testes estatísticos.

Por fim, o script `comparar_modelos.py` consolida CNN–LSTM e baselines por ativo, usando a média da acurácia direcional (e das demais métricas) ao longo dos folds. O comparativo final é exportado em `data/processed/comparativo_cnn_lstm_vs_baselines.csv`.

4.6 Backtests e custos

4.6.1 Regras do backtest.

Executamos *backtests* em duas modalidades: **long-only** (apenas compra ou neutro; sinal de baixa é tratado como saída para caixa) e **long/short** (compra, venda a descoberto ou neutro). A entrada e a saída são determinadas pelas probabilidades previstas pelo modelo: dado um limiar (por padrão 0,5), probabilidade acima do limiar gera sinal de alta (1), abaixo de ($1 - \text{limiar}$) gera sinal de baixa (-1), e no intervalo intermediário o sinal é neutro (0). O backtest é aplicado por fold: para cada bloco de teste do *walk-forward*, os retornos realizados são alinhados ao sinal previsto (retorno na barra seguinte à previsão) e a curva de capital é simulada período a período, com custos descontados a cada mudança de posição. A implementação está em `utils/backtesting.py` e é acionada pelo script `rodar_backtest.py` por ativo, fold e estratégia (B3, 2025a).

4.6.2 Custos e slippage.

Os custos são modelados em três componentes: (i) **custo fixo** por operação (corretagem em R\$; no pipeline padrão 0, conforme corretoras com corretagem zero); (ii) **custo proporcional** ao volume negociado, fixado em 0,03% (`CUSTO_TAXA_PROPORCIONAL = 0.0003`), em linha com emolumentos B3 (Neg, CCP, TTA) para operações regulares; (iii) **slippage** como fração do valor, em 0,05% (`CUSTO_SLIPPAGE = 0.0005`), faixa realista entre 0,05% e 0,1%. O capital inicial padrão é R\$ 100.000 (`CAPITAL_INICIAL`). Cada mudança de posição (entrada ou saída) incorre em custo; no caso de reversão (long \leftrightarrow short), contam-se duas pernas. A análise de **sensibilidade a custos** varia os parâmetros por multiplicadores ($0,5\times$, $1\times$, $1,5\times$, $2\times$) para verificar como retorno líquido, Sharpe e *turnover* respondem ao aumento de atrito.

4.6.3 Métricas do backtest.

Para cada execução são reportados: **retorno líquido** (variação percentual do capital inicial ao final, após todos os custos); **índice de Sharpe** anualizado (média dos retornos líquidos por período sobre desvio padrão, multiplicado por $\sqrt{\text{barra_por_ano}}$); **maximum drawdown** (maior queda percentual do pico ao vale na curva de capital); **profit factor** (soma dos ganhos / valor absoluto da soma das perdas); **turnover** (número de mudanças de posição por barra) e **número de operações**. Os resultados são salvos em CSV em `data/backtest/` por ativo, fold e estratégia.

4.7 Robustez e significância

4.7.1 Teste de Diebold–Mariano.

Comparamos a série de **perdas** do modelo proposto (CNN–LSTM) com a de cada baseline (Naive, Drift, ARIMA, Prophet) por meio do teste de Diebold–Mariano ([DIEBOLD; MARIANO, 1995](#)). A perda é definida como 0/1 para acurácia direcional (1 = erro, 0 = acerto) apenas em observações não neutras; opcionalmente usa-se também a **perda Brier** para comparar qualidade probabilística. A estatística DM baseia-se na diferença média das perdas entre os dois modelos, com variância estimada por kernel de Bartlett (HAC) para acomodar autocorrelação das séries de perdas (horizonte $h = 1$). O p-valor bilateral indica se há diferença significativa entre os modelos; no pipeline, o script `rodar_testes_estatisticos.py` executa o teste para todos os ativos e baselines e consolida os p-valores; `gerar_tabelas_graficos_dm.py` gera tabelas e heatmaps para o relatório.

4.7.2 Análise por regimes e por granularidade.

Para verificar estabilidade em diferentes condições de mercado, os resultados são segmentados por **regime de volatilidade**: usando a coluna `volatility` (janela de 20 períodos) de `df_features`, as observações são divididas em *baixa volatilidade* (abaixo da mediana) e *alta volatilidade* (acima da mediana), aproximando calmaria vs. choques. O teste de Diebold–Mariano é aplicado separadamente em cada segmento quando há amostra suficiente (mínimo de observações exigido no script). A opção `-regimes` em `rodar_testes_estatisticos.py` ativa essa segmentação. A análise de sensibilidade por **granularidade temporal** (barras de 5 min ou 30 min) está prevista em etapas posteriores; no presente trabalho os experimentos foram realizados apenas com barras de 15 min.

4.8 Reprodutibilidade

4.8.1 Organização do repositório e versões.

O código está organizado no diretório `codigo/pipeline/`: módulos em `src/` (dados, modelos, métricas, validação, backtest, Optuna), scripts de execução em `src/scripts/` e `src/tests/`, e configuração central em `src/config.py`. As sementes (NumPy, TensorFlow) são fixadas em `SEED = 42`. As dependências são gerenciadas com uv (ou pip conforme `requirements.txt`); não há `prepare_data.py` ou `evaluate.py` como scripts únicos — o fluxo é orquestrado por `rodar_pipeline_completo.sh`, que chama em sequência visualização de features, baselines, treino CNN–LSTM, análise de modelos, backtests, testes estatísticos e sensibilidade do walk-forward.

4.8.2 Reprodução do zero.

Para reproduzir os experimentos: (1) colocar os CSVs de dados brutos (OHLCV, 15 min) em data/raw/ com o nome esperado por `obter_nome_arquivo_dados(ativo)`; (2) instalar dependências (`uv sync` ou `pip install -r requirements.txt`); (3) executar `./rodar_pipeline_completo.sh` a partir do diretório do pipeline (ou, para testes rápidos, `-rapido` para menos épocas e trials). O script faz backup dos artefatos antigos, gera modelos em `models/<ativo>/cnn_lstm/`, métricas e comparativos em `data/processed/`, backtests em `data/backtest/` e logs em `logs/pipeline/`. Treino e baselines podem ser rodados por ativo com `uv run python src/train.py -ativo PETR4 -modelo cnn_lstm -optuna -n-trials 20` e `uv run python src/tests/testar_baselines_walkforward.py -todos`.



RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos com o modelo proposto (CNN–LSTM) e com os *baselines* definidos na metodologia. A análise segue o protocolo *walk-forward*, com embargo entre treino e teste, e usa como referência principal a acurácia direcional após a banda morta de 0,1%. Além disso, os resultados são complementados por testes de significância (Diebold–Mariano) e por *backtests* com custos explícitos.

Os experimentos foram executados para três ativos de alta liquidez (PETR4, VALE3 e ITUB4), com barras intradiárias de 15 minutos. Em cada ativo, a avaliação é feita por *fold* e, ao final, reportamos o desempenho médio entre os *folds*. Essa estrutura é importante porque aproxima o cenário de um uso real: o modelo sempre prevê o “futuro” a partir do “passado”.

5.1 Visão geral dos resultados e organização dos artefatos

Os resultados consolidados foram organizados em quatro grupos de artefatos:

- **Comparativo entre modelos:** arquivo `data/processed/comparativo_cnn_lstm_vs_-
baselines.csv`, com métricas médias por ativo e por modelo.
- **Resultados por fold:** arquivos `<ATIVO>_cnn_lstm_walkforward.csv`, com métricas do CNN–LSTM ao longo do *walk-forward*.
- **Teste de Diebold–Mariano:** arquivos `dm_resumo_pvalores.csv` e `testes_diebold_-
mariano.csv`, incluindo versão segmentada por regime de volatilidade.
- **Backtests:** arquivo `data/backtest/historico_backtest.csv` (consolidado) e CSVs por ativo/fold/estratégia.

A leitura dos resultados neste capítulo segue essa mesma lógica: primeiro comparamos modelos em termos de acerto estatístico, depois verificamos se a diferença é significativa e, por fim, traduzimos isso em impacto financeiro via *backtest*.

5.2 Comparação geral: CNN–LSTM vs. baselines

A Tabela 7 resume o desempenho médio do CNN–LSTM e dos principais *baselines* por ativo. O foco está na acurácia direcional (após banda morta), mas também reportamos métricas que ajudam a interpretar cenários desbalanceados e qualidade probabilística (Brier e Log-Loss).

Tabela 7 – Comparativo médio entre CNN–LSTM e baselines (média dos folds; após banda morta).

Ativo	Modelo	Hit rate	Bal. Acc.	MCC	Brier
PETR4	CNN–LSTM	«%»	«»	«»	«»
PETR4	Naive	«%»	«»	«»	«»
PETR4	Drift	«%»	«»	«»	«»
PETR4	ARIMA	«%»	«»	«»	«»
PETR4	Prophet	«%»	«»	«»	«»
VALE3	CNN–LSTM	«%»	«»	«»	«»
VALE3	Naive	«%»	«»	«»	«»
VALE3	Drift	«%»	«»	«»	«»
VALE3	ARIMA	«%»	«»	«»	«»
VALE3	Prophet	«%»	«»	«»	«»
ITUB4	CNN–LSTM	«%»	«»	«»	«»
ITUB4	Naive	«%»	«»	«»	«»
ITUB4	Drift	«%»	«»	«»	«»
ITUB4	ARIMA	«%»	«»	«»	«»
ITUB4	Prophet	«%»	«»	«»	«»

Fonte: O autor, a partir de `comparativo_cnn_lstm_vs_baselines.csv`.

De forma geral, essa tabela responde a primeira pergunta do trabalho: *o modelo proposto traz ganho real sobre métodos simples?* A comparação é importante porque, em finanças, ganhos pequenos podem ser relevantes. Ao mesmo tempo, eles precisam ser consistentes. Por isso, a leitura não deve ficar apenas no “melhor valor”. Ela precisa considerar estabilidade e significância.

Também é útil observar as métricas probabilísticas. Em especial, o Brier e o Log-Loss ajudam a entender se as probabilidades do modelo têm “qualidade”, e não apenas aparência de número. Isso importa porque o *backtest* depende de limiares e decisões de entrada/saída.

5.3 Estabilidade no tempo: desempenho por fold do walk-forward

A média é um bom resumo, mas não conta a história inteira. Em séries financeiras, o desempenho pode variar bastante de um período para outro. Por isso, analisamos os resultados por *fold*. Isso permite enxergar duas coisas: (i) se o ganho do CNN–LSTM é recorrente e (ii) se existem períodos em que o modelo degrada de forma mais forte.

A Tabela 8 apresenta um exemplo de como organizar os resultados do CNN–LSTM por *fold*. Aqui, a ideia é simples: mostrar a variabilidade e comentar os extremos.

Tabela 8 – CNN–LSTM por fold (exemplo de preenchimento a partir de <ATIVO>_cnn_lstm_walkforward.csv).

Ativo	Fold	Hit rate	MCC	Brier
PETR4	1	<>>	<>>	<>>
PETR4	2	<>>	<>>	<>>
PETR4	3	<>>	<>>	<>>
PETR4	4	<>>	<>>	<>>
PETR4	5	<>>	<>>	<>>
VALE3	1	<>>	<>>	<>>
VALE3	2	<>>	<>>	<>>
VALE3	3	<>>	<>>	<>>
VALE3	4	<>>	<>>	<>>
VALE3	5	<>>	<>>	<>>
ITUB4	1	<>>	<>>	<>>
ITUB4	2	<>>	<>>	<>>
ITUB4	3	<>>	<>>	<>>
ITUB4	4	<>>	<>>	<>>
ITUB4	5	<>>	<>>	<>>

Fonte: O autor.

Na discussão, vale observar se existem *folds* em que o desempenho cai muito. Quando isso acontece, a causa mais comum é mudança de regime. Pode ser um período mais volátil. Pode ser um período com tendência forte. Ou pode ser um momento em que o preço “oscila sem direção”. Esses cenários são importantes porque eles expõem o limite prático do modelo.

5.4 Top configurações do Optuna e sensibilidade do lookback

Além do desempenho final, analisamos as melhores configurações encontradas na busca bayesiana. Esse ponto ajuda a responder outra pergunta: *o modelo depende de um ajuste “muito*

fino” para funcionar? Se a resposta for “sim”, isso aumenta o risco de instabilidade fora da amostra.

A Tabela 9 é um modelo de como reportar o “top 10” do tuning (um resumo por ativo costuma ser suficiente no texto principal; o restante pode ir para apêndice).

Tabela 9 – Top configurações (Optuna) para o CNN–LSTM (exemplo; preencher a partir de <ATIVO>_cnn_lstm_analise_modelos.csv).

Rank	Filtros	Kernel	LSTM	Dropout	Hit rate
1	«»	«»	«»	«»	«»
2	«»	«»	«»	«»	«»
3	«»	«»	«»	«»	«»
4	«»	«»	«»	«»	«»
5	«»	«»	«»	«»	«»
6	«»	«»	«»	«»	«»
7	«»	«»	«»	«»	«»
8	«»	«»	«»	«»	«»
9	«»	«»	«»	«»	«»
10	«»	«»	«»	«»	«»

Fonte: O autor.

Aqui vale comentar também o **lookback**. Como a janela padrão é 60, mas janelas 16/32/64 foram avaliadas, a discussão pode ser direta: se o desempenho muda pouco, isso é um bom sinal. Significa que o modelo não está preso a um único horizonte. Se muda muito, isso indica sensibilidade. Nesse caso, a escolha do lookback vira uma decisão crítica do pipeline.

5.5 Significância estatística: teste de Diebold–Mariano

A comparação por métricas é útil, mas ainda falta uma peça. Em finanças, diferenças pequenas podem acontecer “por sorte”, principalmente em janelas curtas. Para reduzir esse risco, aplicamos o teste de Diebold–Mariano (DM), comparando a série de perdas do CNN–LSTM com a de cada baseline.

A Tabela 10 organiza os p-valores do DM (versão geral). A interpretação é simples: p-valores baixos indicam evidência de que a diferença média de desempenho não é aleatória. Na prática, isso ajuda a sustentar a conclusão de que o modelo proposto realmente melhora o baseline.

Se você preferir, dá para incluir também o **heatmap** de p-valores como figura. Ele costuma ficar bem visual no relatório:

Tabela 10 – Resumo do teste de Diebold–Mariano (p-valores; preencher a partir de `dm_resumo_pvalores.csv`).

Ativo	Comparação	p-valor
PETR4	CNN–LSTM vs Naive	<>
PETR4	CNN–LSTM vs Drift	<>
PETR4	CNN–LSTM vs ARIMA	<>
PETR4	CNN–LSTM vs Prophet	<>
VALE3	CNN–LSTM vs Naive	<>
VALE3	CNN–LSTM vs Drift	<>
VALE3	CNN–LSTM vs ARIMA	<>
VALE3	CNN–LSTM vs Prophet	<>
ITUB4	CNN–LSTM vs Naive	<>
ITUB4	CNN–LSTM vs Drift	<>
ITUB4	CNN–LSTM vs ARIMA	<>
ITUB4	CNN–LSTM vs Prophet	<>

Fonte: O autor.

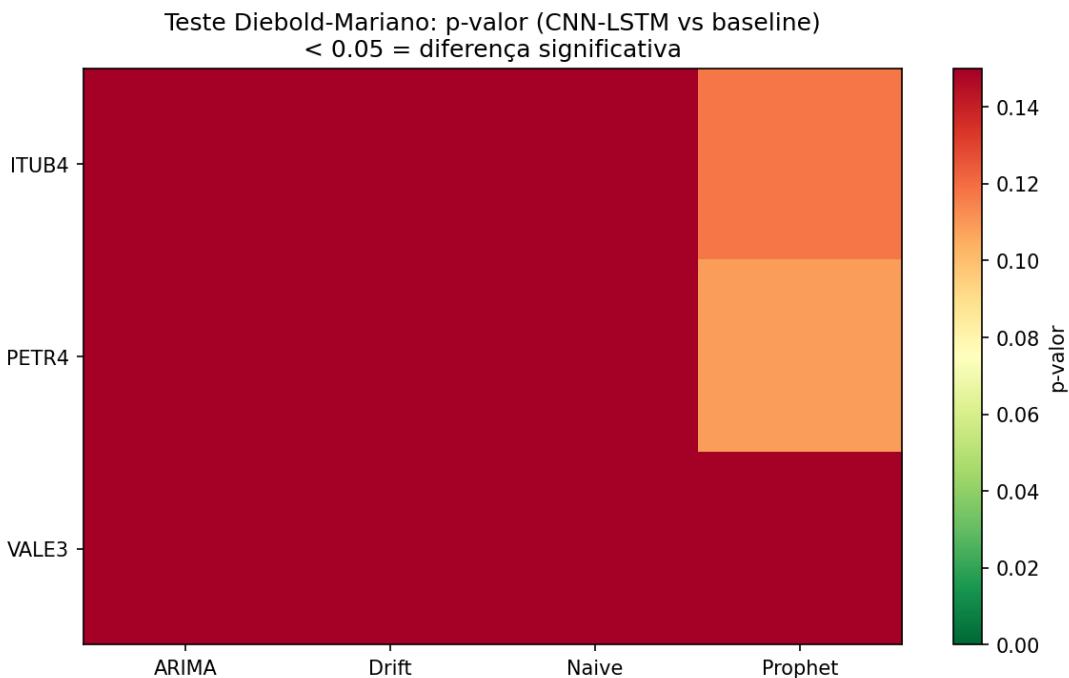


Figura 11 – Heatmap de p-valores do teste de Diebold–Mariano (comparação CNN–LSTM vs baselines).
Fonte: o autor.

Na discussão, vale tomar cuidado com dois pontos. Primeiro: significância não é “garantia de lucro”. Ela apenas diz que a diferença estatística é consistente. Segundo: se o DM for significativo apenas em um ativo, isso ainda é um resultado válido. Só muda a força da generalização.

5.6 Backtests: impacto prático após custos

Depois de medir acerto e significância, a pergunta final é a mais prática: *o modelo gera resultado financeiro quando custos entram na conta?* Para isso, executamos *backtests* em duas modalidades: long-only e long/short.

A Tabela 11 resume long-only. Ela é útil porque é o cenário mais simples. O sinal de baixa vira “ficar em caixa”. Isso costuma reduzir risco e também reduz o número de operações.

Tabela 11 – Backtest long-only (após custos; preencher a partir de `historico_backtest.csv`).

Ativo	Retorno líquido	Sharpe	Max DD	Profit factor
PETR4	«»	«»	«»	«»
VALE3	«»	«»	«»	«»
ITUB4	«»	«»	«»	«»

Fonte: O autor.

A Tabela 12 resume long/short. Esse modo é mais agressivo. Ele pode ganhar em tendências de queda, mas normalmente “paga” com mais giro e mais custo. Por isso, o que importa aqui é observar retorno *e* risco (Sharpe e drawdown).

Tabela 12 – Backtest long/short (após custos; preencher a partir de `historico_backtest.csv`).

Ativo	Retorno líquido	Sharpe	Max DD	Profit factor
PETR4	«»	«»	«»	«»
VALE3	«»	«»	«»	«»
ITUB4	«»	«»	«»	«»

Fonte: O autor.

Na discussão, é importante deixar claro o seguinte: um modelo pode ter bom *hit rate* e ainda assim não ser lucrativo. Isso acontece quando o ganho médio por operação é pequeno e o custo come a margem. Por isso, turnover e sensibilidade a custos entram como parte essencial da análise.

5.7 Robustez: regimes de volatilidade

Por fim, segmentamos a avaliação por regimes de volatilidade. A motivação é direta: o mercado não se comporta da mesma forma o tempo todo. Em períodos calmos, padrões técnicos podem aparecer com mais clareza. Em choques, o ruído aumenta e o preço pode mudar de regime rapidamente.

A análise separa as observações em baixa e alta volatilidade (em relação à mediana), e aplica o teste de Diebold–Mariano em cada segmento. Os resultados podem ser reportados de

forma resumida na Tabela 13. Ela ajuda a responder se o ganho do CNN–LSTM é “geral” ou se ele depende mais do tipo de mercado.

Tabela 13 – Teste DM por regime de volatilidade (p-valores; preencher a partir de `dm_resumo_pvalores_regime_*.csv`).

Ativo	Comparação	p (baixa vol)	p (alta vol)
PETR4	CNN–LSTM vs baselines	<>	<>
VALE3	CNN–LSTM vs baselines	<>	<>
ITUB4	CNN–LSTM vs baselines	<>	<>

Fonte: O autor.

Se o modelo mantém desempenho no regime de alta volatilidade, isso é um sinal forte de robustez. Se ele cai muito, isso não “invalida” o método. Mas mostra uma limitação prática clara. Nessa situação, uma extensão natural do trabalho é adaptar limiares por regime, ou introduzir um detector de regime para ajustar a agressividade da estratégia.

5.8 Síntese

Em resumo, os resultados deste capítulo são organizados em três níveis. No primeiro nível, verificamos se o CNN–LSTM supera baselines em métricas de classificação. No segundo, avaliamos se a diferença é estatisticamente consistente via Diebold–Mariano. No terceiro, testamos utilidade prática com *backtests* após custos.

Essa estrutura evita uma conclusão baseada em um único número. Ela força o resultado a “passar” por filtros diferentes. E isso é especialmente importante em finanças, onde instabilidade é a regra, não a exceção.



CONSIDERAÇÕES FINAIS

O trabalho até aqui forneceu uma definição sistemática de escopo, objetivos e planejamento metodológico para uma tarefa de previsão direcional intradiária na B3, especificamente em uma proposta forte para uma avaliação completa e reproduzível. O principal objetivo foi fornecer um protocolo sensível ao tempo que não vaze informações e represente melhor o uso real com custos e slippage. Para alcançar isso, o plano integrou:

- i. construção do banco de dados intradiário;
- ii. configuração experimental com janelas deslizantes e validação cruzada walk-forward com embargo temporal;
- iii. comparação de linhas de base (por exemplo, ARIMA e LSTM) com as redes propostas;
- iv. definição de métricas relacionadas à tomada de decisão;
- v. backtests usando custos explícitos;
- vi. testes de robustez e significância; e
- vii. reproduzibilidade, por meio de versionamento e scripts executáveis.

Quanto à base teórico-empírica agora consolidada, o artigo fundamentou-se na dinâmica intradiária (liquidez, volatilidade, volume) e microestrutura, e na decisão de introduzir uma banda morta (positiva) para mitigar ruídos na estimativa da precisão direcional. A revisão apresentou a evolução dos modelos, começando pelos clássicos (como a família ARIMA, modelos Box-Jenkins) até métodos de aprendizado profundo (LSTM, CNN) e arquiteturas contemporâneas com mecanismos de atenção, apontando as limitações de modelos puramente lineares em ambientes não estacionários e na presença de choques, e os ganhos de modelos que lidam com dependência de longo prazo. Além disso, foi mencionada a relevância dos Transformers para séries financeiras

multivariadas e de alta frequência, como o TF-T, e algumas evidências de ganhos sobre o LSTM no RMSE e precisão direcional são encontradas sob tamanho de janela longo. O estudo também empregou um modelo híbrido LSTM+CNN como referência para capturar características temporais e locais das características intradiárias, consistente com a intenção de generalidade em vários ambientes de volatilidade.

Para avaliação, uma gama de métricas de decisão e probabilidade foi incluída: precisão direcional com folga (foco principal), qualidade probabilística (Brier Score, Log-Loss e ECE) e medidas de equilíbrio (Balanced Accuracy, F1, MCC, AUC-PR). A validação econômica foi adicionada com backtests com custo e slippage em Peso, retornando retorno líquido, índice de Sharpe, máximo drawdown e fator de lucro de acordo com estruturas operacionais e custos divulgados pela B3. A significância da diferença entre os modelos será examinada através da estatística de Diebold–Mariano, onde a análise é segmentada de acordo com regimes de volatilidade para garantir a robustez da comparação. Por fim, a reproduzibilidade foi garantida pelo versionamento completo de dados, código, artefatos e seeds com scripts de ponta a ponta para preparar, treinar, avaliar e realizar backtests.

A motivação do tema surge do fato de que observamos consistentemente o desempenho em laboratório e o desempenho no piso em desacordo um com o outro em um mundo intradiário ruidoso com custos. O plano metodológico tenta reduzir essa lacuna integrando walk-forward com prevenção de vazamento de embargo, métricas orientadas para decisão e backtests pós-custo e entregar critérios de validação mais robustos enquanto preserva a possibilidade de generalização entre ativos e estados de mercado. As ameaças à validade (qualidade dos dados, viés de sobrevivência, divisões temporais) e mitigadores de resultados também foram claramente declarados, aumentando a confiabilidade do desenho experimental.

Para concluir, as ações subsequentes, explicadas no Capítulo de próximos passos – posse/limpeza dos dados intradiários, pré-processamento de dados e geração de características, configuração de arquiteturas (linhas de base e propostas), definição do espaço de busca de hiperparâmetros, protocolo de treino/validação walk-forward com embargo e avaliações de métricas e backtests com testes de robustez e significância – são o elo imediato para o PFC2. Realizar essas etapas na próxima fase torna possível (1) realizar os experimentos sob controles claros, (2) comparar os diferentes modelos de forma estatisticamente significativa, (3) relatar os resultados de forma tão transparente e reproduzível quanto desejado, fechando o ciclo metodológico conforme pretendido neste trabalho.

REFERÊNCIAS

AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: A next-generation hyperparameter optimization framework. In: **KDD**. [S.l.: s.n.], 2019. Citado nas páginas [44](#), [54](#) e [56](#).

ALONSO, M. N. I.; FRANKLIN, R. P. Large language models for financial time series forecasting. **SSRN Working Paper**, 2024. Testados modelos LLM e especialistas como PatchTST, TimeGPT em dados de ações globais. Citado nas páginas [15](#), [42](#) e [45](#).

AMIHUD, Y. Illiquidity and stock returns: Cross-section and time-series effects. **Journal of Financial Markets**, v. 5, n. 1, p. 31–56, Jan. 2002. Citado nas páginas [13](#), [30](#) e [44](#).

ANDERSEN, T. G.; BOLLERSLEV, T. Intraday periodicity and volatility persistence in financial markets. **Journal of Empirical Finance**, v. 4, n. 2–3, p. 115–158, Jun. 1997. Citado nas páginas [13](#), [15](#), [17](#), [30](#), [39](#) e [49](#).

B3. Manual de Procedimentos Operacionais de Negociacao da B3 — versao 17/02/2025. 2025. <<https://www.b3.com.br/data/files/55/65/B5/7D/AC31591029BEEC39AC094EA8/MPO%20de%20Negociacao%20da%20B3.pdf>>. Acesso em: ago. 2025. Citado nas páginas [17](#), [24](#), [30](#), [37](#), [48](#) e [52](#).

B3. Produtos e Serviços — B3. 2025. <https://www.b3.com.br/pt_br/produtos-e-servicos/>. Acesso em: ago. 2025. Citado nas páginas [24](#), [48](#) e [58](#).

B3. Quem Somos — B3. 2025. Acesso em: ago. 2025. Available: <https://www.b3.com.br/pt_br/institucional/quem-somos/>. Citado nas páginas [13](#) e [24](#).

B3. B3 — Brasil, Bolsa, Balcão. s.d. Acesso em: 7 ago. 2025. Available: <<https://www.b3.com.br/>>. Citado na página [23](#).

BANK, W. Global Economic Prospects: Slow Growth, Policy Challenges. [S.l.], 2020. Available: <<https://documents1.worldbank.org/curated/en/494581578977519941/pdf/Global-Economic-Prospects-Slow-Growth-Policy-Challenges.pdf>>. Citado nas páginas [13](#), [14](#) e [25](#).

BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. **Information Sciences**, v. 191, p. 192–213, 2012. Acesso em 04 Ago 2025. Available: <<https://www.sciencedirect.com/science/article/abs/pii/S0020025511006773>>. Citado nas páginas [18](#), [31](#), [44](#), [48](#), [50](#) e [54](#).

BODIE, Z.; KANE, A.; MARCUS, A. J. **Investments**. 12. ed. [S.l.]: McGraw-Hill Education, 2021. ISBN 978-1260775156. Citado nas páginas [13](#), [14](#), [25](#), [26](#), [27](#) e [28](#).

BOLLINGER, J. Bollinger on Bollinger Bands. New York: McGraw-Hill, 2001. ISBN 978-0071386739. Citado nas páginas [26](#), [27](#) e [51](#).

BOROVYKH, A.; BOHTE, S.; OOSTERLEE, C. W. **Conditional Time Series Forecasting with Convolutional Neural Networks**. 2017. Available: <<https://arxiv.org/abs/1703.04691>>. Citado nas páginas 52 e 54.

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 14, 18, 37, 44 e 52.

BRASIL. **Lei nº 6.385, de 7 de dezembro de 1976 (Mercado de Valores Mobiliários e CVM)**. 1976. Acesso em: 7 ago. 2025. Available: <http://www.planalto.gov.br/ccivil_03/leis/l6385.htm>. Citado nas páginas 13 e 23.

BRASIL. **Lei nº 6.404, de 15 de dezembro de 1976 (Lei das Sociedades por Ações)**. 1976. Acesso em: 7 ago. 2025. Available: <http://www.planalto.gov.br/ccivil_03/leis/l6404consol.htm>. Citado nas páginas 13, 23 e 25.

BRIER, G. W. Verification of forecasts expressed in terms of probability. **Monthly Weather Review**, v. 78, n. 1, p. 1–3, 1950. Citado nas páginas 18 e 57.

CAMPBELL, J. Y.; LO, A. W.; MACKINLAY, A. C. **The Econometrics of Financial Markets**. Princeton, NJ: Princeton University Press, 1997. Citado na página 29.

CHANG, C.; WANG, W.-Y.; PENG, W.-C.; CHEN, T.-F. Llm4ts: Aligning pre-trained llms as data-efficient time-series forecasters. **arXiv preprint**, 2023. Citado nas páginas 15, 43 e 45.

COSTA, L. D.; MACHADO, A. M. C. Prediction of stock price time series using transformers. **Revista Brasileira de Aprendizado de Máquina e Inteligência Artificial**, 2023. Aplicado a ações do Ibovespa em 2008. Citado nas páginas 41 e 44.

DACOROGNA, M. M.; GENÇAY, R.; MÜLLER, U. A.; OLSEN, R. B.; PICTET, O. V. **An Introduction to High-Frequency Finance**. San Diego, CA: Academic Press, 2001. ISBN 9780122796715. Citado nas páginas 16 e 29.

DAYCOVAL. **Índices de mercado: o que são e como utilizar**. 2024. Blog Daycoval. Acesso em 31 de julho de 2025. Available: <<https://blog.daycoval.com.br/indices-de-mercado-o-que-sao/>>. Citado na página 26.

DERBENTSEV, V. Machine learning approaches for financial time series forecasting. **Epj Data Science**, 2020. Citado na página 44.

DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. **Journal of Business & Economic Statistics**, v. 13, n. 3, p. 253–263, 1995. Citado nas páginas 18 e 59.

DING, G.; QIN, L. Study on the prediction of stock price based on the associated network model of lstm. **International Journal of Machine Learning and Cybernetics**, v. 11, n. 6, p. 1307–1317, 2020. Citado nas páginas 14, 16 e 17.

DUFOUR, A.; ENGLE, R. F. Time and the price impact of a trade. **Journal of Finance**, v. 55, n. 6, p. 2467–2498, Dec. 2000. Citado nas páginas 13, 17 e 30.

ENGLE, R. F.; RUSSELL, J. R. Autoregressive conditional duration: A new model for irregularly spaced transaction data. **Econometrica**, v. 66, n. 5, p. 1127–1162, 1998. Citado nas páginas 13, 15, 17 e 30.

- FABOZZI, F. J.; MODIGLIANI, F. **Capital Markets: Institutions and Instruments**. 4. ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009. ISBN 978-0136026020. Citado nas páginas [13](#) e [25](#).
- FAN, Y.; OUTROS. Smetimes: Efficient large language models for time series forecasting. **arXiv preprint arXiv:2503.03594**, 2025. Citado na página [43](#).
- FJELLSTRÖM, C. Long short-term memory neural network for financial time series. **arXiv preprint arXiv:2201.08218**, 2022. Citado nas páginas [40](#) e [44](#).
- GIACOMEL, F. dos S. **Um Método Algorítmico para Operações na Bolsa de Valores Baseado em Ensembles de Redes Neurais para Modelar e Prever os Movimentos dos Mercados de Ações**. Master's Thesis (Dissertação de Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016. Orientadora: Profa. Dra. Renata Galante. Citado na página [29](#).
- GOEL, R.; WANG, S.; FELDMAN, S.; MADDOX, W. TimesFM: Foundation models for probabilistic time series forecasting. **Journal of Machine Learning Research**, v. 26, n. 54, p. 1–36, 2025. Available: <<http://jmlr.org/papers/v26/24-054.html>>. Citado na página [15](#).
- GONÇALVES, R. S. **O impacto do horizonte de previsão em modelos ARIMA e LSTM aplicados ao IBX50**. Master's Thesis (Master's Thesis) — Universidade Estadual de Campinas, Campinas, SP, 2023. Disponível em: <<https://repositorio.unicamp.br/handle/REPOSIP/377123>>. Acesso em: ago. 2025. Citado nas páginas [38](#) e [44](#).
- GUO, C.; PLEISS, G.; SUN, Y.; WEINBERGER, K. Q. On calibration of modern neural networks. In: **ICML**. [S.l.: s.n.], 2017. Citado nas páginas [18](#) e [57](#).
- HAMILTON, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. **Econometrica**, v. 57, n. 2, p. 357–384, 1989. Citado na página [32](#).
- HASBROUCK, J. **Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading**. [S.l.]: Oxford University Press, 2007. ISBN 978-0-19-530164-9. Citado nas páginas [13](#), [15](#), [17](#), [30](#), [39](#), [49](#) e [56](#).
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. Citado nas páginas [40](#), [44](#), [52](#) e [53](#).
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. OTexts, 2018. 2nd ed. Available: <<https://otexts.com/fpp2/>>. Citado nas páginas [14](#), [17](#), [18](#), [31](#), [44](#) e [52](#).
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 3. ed. OTexts, 2021. 3rd ed. Available: <<https://otexts.com/fpp3/>>. Citado na página [29](#).
- JIN, M.; WANG, S.; MA, L.; CHU, Z.; ZHANG, J. Y.; SHI, X.; CHEN, P.-Y.; LIANG, Y.; LI, Y.-F.; PAN, S.; WEN, Q. Time-lm: Time series forecasting by reprogramming large language models. **arXiv preprint**, 2023. Citado nas páginas [15](#), [43](#) e [45](#).
- JR, P. R.; OUTROS. Modelo arima aplicado à previsão do índice bovespa. **Open Journal of Statistics**, Scientific Research, v. 4, n. 5, p. 319–324, 2014. Citado nas páginas [16](#), [18](#), [37](#) e [44](#).
- JUNIOR, M. M. d. O. L. **Proposta de um modelo de predição da bolsa de valores utilizando uma abordagem híbrida**. Master's Thesis (Dissertação de Mestrado) — Universidade Federal do Maranhão, São Luís, 2013. Citado na página [15](#).

KABIR, M. R.; OUTROS. Lstm-mtrans-mlp: Hybrid deep learning ensemble for financial time series forecasting. **Information**, 2025. Citado nas páginas 40 e 44.

KASHIF MUHAMMAD E ŚLEPACZUK, R. Lstm-arima hybrid model for stock market forecasting: A multi-index evaluation. **Expert Systems with Applications**, v. 233, p. 120992, 2024. Citado nas páginas 38 e 44.

LATHUILÈRE, B.; DROIT, A.; IDRISI, N.; PEZESHKI, M. TimeGPT: Few-shot forecasting with large language models. In: **Proceedings of the 12th International Conference on Learning Representations (ICLR 2024)**. [S.l.: s.n.], 2024. Poster. Citado nas páginas 15 e 45.

LEONUNESBIO. **Introdução à Análise de Séries Temporais: como esses modelos são usados para responder questões ambientais?** 2014. Acessado em: 17/08/2025. Available: <<https://limnonews.wordpress.com/2014/09/18/introducao-a-analise-de-series-temporais-como-esses-modelos-sao-usados-para-responder-questoes-ambientais/>>. Citado na página 15.

LI, Z. *et al.* Deepseek-ts+: A unified framework for multi-product time series forecasting. **Towards AI**, 2025. <<https://pub.towardsai.net/deepseek-ts-a-unified-framework-for-multi-product-time-series-forecasting-da4932ce6022>>. Citado na página 43.

LIANG, K.; ZHANG, Y.; LIU, X.; XU, W. Hidformer: Dual-tower hypergraph transformer for multimodal stock prediction. In: **Proceedings of the 44th International Conference on Machine Learning (ICML 2025)**. [S.l.: s.n.], 2025. p. 12245–12260. Citado na página 15.

LIM, B.; ZOHREN, S.; ROBERTS, S. Temporal fusion transformers for interpretable multi-horizon time series forecasting. **International Journal of Forecasting**, Elsevier, v. 37, n. 4, p. 1748–1764, 2021. Citado nas páginas 15 e 35.

LIU, Y.; WU, H.; WANG, J.; LONG, M. Non-stationary transformers: Exploring the stationarity in time series forecasting. **arXiv preprint**, 2022. Citado nas páginas 15, 41 e 44.

LO, A. W. **Adaptive Markets: Financial Evolution at the Speed of Thought**. Princeton, NJ: Princeton University Press, 2017. ISBN 978-0-691-15319-1. Citado na página 13.

London Stock Exchange. **London Stock Exchange (LSE)**. s.d. Acesso em: 7 ago. 2025. Available: <<https://www.londonstockexchange.com/>>. Citado na página 23.

LOSHCHILOV, I.; HUTTER, F. Decoupled weight decay regularization. In: **ICLR**. [S.l.: s.n.], 2019. Citado na página 56.

MAKRIDAKIS, S.; SPILLOTIS, E.; ASSIMAKOPOULOS, V. Statistical and machine learning forecasting methods: Concerns and ways forward. **PLOS ONE**, 2018. Citado nas páginas 44 e 45.

MISHKIN, F. S.; EAKINS, S. G. **Financial Markets and Institutions**. 9. ed. [S.l.]: Pearson, 2018. ISBN 978-0-13-451926-5. Citado nas páginas 13, 16, 23, 25 e 27.

MURPHY, J. J. **Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications**. 2. ed. New York: New York Institute of Finance, 1999. ISBN 978-0735200661. Citado nas páginas 26, 27, 28 e 51.

- NASCIMENTO, O. S.; SANTOS, F. G.; FERREIRA, K. H. A. Previsão de Preços de Ações Utilizando Inteligência Artificial. Sociedade Brasileira de Computação (SBC), Porto Alegre, RS, Brasil, p. 37–47, 2023. Available: <<https://sol.sbc.org.br/index.php/bwaif/article/view/20485>>. Citado nas páginas 14 e 16.
- NASDAQ. Nasdaq. s.d. Acesso em: 7 ago. 2025. Available: <<https://www.nasdaq.com/>>. Citado na página 23.
- NASIR, M.; ZHANG, W.; WANG, J. Forecasting stock prices using lmd-based hybrid arima-xgboost framework. **Journal of Computational Finance**, v. 29, n. 1, p. 33–55, 2025. Forthcoming. Citado nas páginas 38 e 44.
- New York Stock Exchange. New York Stock Exchange (NYSE). s.d. Acesso em: 7 ago. 2025. Available: <<https://www.nyse.com/>>. Citado na página 23.
- NIE, Z.; LIU, M.; ZHOU, H.; LI, X.; ZHAO, T.; ZHOU, T.; WEN, Q. Patchtst: Transformer with patch attention for long-term time-series forecasting. In: **Proceedings of the 40th International Conference on Machine Learning (ICML 2023)**. [S.l.: s.n.], 2023. p. 16906–16930. Citado nas páginas 15, 41 e 44.
- OLAH, C. Understanding LSTM Networks. 2015. <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em: 18 ago. 2025. Citado na página 33.
- PHANKOKKRUAD, P.; WORASUWANNARAK, C. Enhancing forex market predictions: A comparative study of candlestick patterns and the middam patterns. ResearchGate, 2025. Available: <https://www.researchgate.net/publication/388573821_Enhancing_FOREX_Market_Predictions_A_Comparative_Study_of_Candlestick_Patterns_and_the_MIDDAM_Patterns>. Citado na página 28.
- PRADO, M. López de. **Advances in Financial Machine Learning**. [S.l.]: Wiley, 2018. Citado nas páginas 18, 44, 48, 50 e 54.
- PRING, M. J. **Technical Analysis Explained: The Successful Investor's Guide to Spotting Investment Trends and Turning Points**. 5. ed. New York: McGraw-Hill Education, 2014. ISBN 978-0071824227. Citado na página 28.
- SAID, S. E.; DICKEY, D. A. Testing for unit roots in autoregressive-moving average models of unknown order. **Biometrika**, v. 71, n. 3, p. 599–607, 1984. Citado na página 29.
- SHARPE, W. F. The sharpe ratio. **Journal of Portfolio Management**, v. 21, n. 1, p. 49–58, 1994. Citado nas páginas 18 e 57.
- SIAMI-NAMINI, S.; NAMIN, A. S. Forecasting economics and financial time series: Arima vs. lstm. **arXiv preprint arXiv:1803.06386**, 2018. Citado nas páginas 40 e 44.
- SOUZA, Y. G. V.; TARRATACA, L.; CARDOSO, D. O.; ASSIS, L. S. d. Sentiment analysis applied to ibovespa prediction. In: **Anais do Congresso Brasileiro de Inteligência Computacional (CBIC)**. [S.l.: s.n.], 2022. p. 373–388. Citado nas páginas 16 e 37.
- SUNKI, A.; SATYAKUMAR, C.; NARAYANA, G. S.; KOPPERA, V.; HAKEEM, M. Time series forecasting of stock market using arima, lstm and fb prophet. In: **MATEC Web of Conferences - ICMED 2024**. [S.l.: s.n.], 2024. v. 392, p. 01163. Citado nas páginas 15 e 16.

TANG, H.; OUTROS. Time series forecasting with llms: Understanding and enhancing preferences and limitations. **KDD Exploration**, 2024. Citado nas páginas [43](#) e [45](#).

TANG, J.; CHEN, S.; GONG, C.; ZHANG, J.; TAO, D. Llm-ps: Empowering large language models for time series forecasting with temporal patterns and semantics. **arXiv preprint**, 2025. Citado nas páginas [42](#) e [45](#).

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. **The American Statistician**, v. 72, n. 1, p. 37–45, 2018. Citado nas páginas [18](#) e [52](#).

TSAY, R. S. **Analysis of Financial Time Series**. 3. ed. Hoboken, NJ: Wiley, 2010. Citado nas páginas [16](#) e [29](#).

VASWANI, A.; SHAZEE, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER Łukasz; POLOSUKHIN, I. Attention is all you need. In: **Advances in Neural Information Processing Systems (NeurIPS 2017)**. [S.l.]: Curran Associates, 2017. p. 5998–6008. Citado nas páginas [34](#) e [35](#).

WEN, Q.; ZHOU, T.; ZHANG, C.; CHEN, W.; MA, Z.; YAN, J.; SUN, L. Transformers in time series: A survey. In: **Proceedings of IJCAI**. [S.l.: s.n.], 2022. Citado nas páginas [14](#), [15](#), [18](#), [41](#), [44](#) e [45](#).

WOOD, R. A.; MCINISH, T. H.; ORD, J. K. An investigation of transactions data for nyse stocks. **The Journal of Finance**, v. 40, n. 3, p. 723–739, 1985. Citado na página [16](#).

WU, H.; LI, J.; YU, R.; JIANG, Y.; LIU, M.; WEN, Q.; YU, Y. Timesnet: Temporal 2d-variation modeling for general time series analysis. In: **Proceedings of the 40th International Conference on Machine Learning (ICML 2023)**. [S.l.: s.n.], 2023. p. 24518–24533. Citado na página [15](#).

WU, H.; NAYAK, R.; DARVISHI, A.; SIMON, D.; BEHESHTI, M. Deep learning for time series forecasting: The electric load case. **Energies**, MDPI, v. 13, n. 1, p. 101, 2020. Citado na página [35](#).

ZENG, A.; CHEN, M.; ZHANG, L.; XU, Q. Are transformers effective for time series forecasting? **arXiv preprint**, 2022. Citado nas páginas [14](#), [15](#), [18](#), [42](#) e [44](#).



GLOSSÁRIO

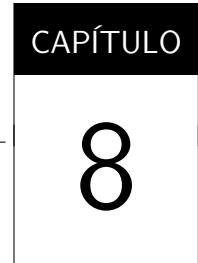
- Gap — Espaço no gráfico de preços entre dois períodos consecutivos, causado por variação abrupta na cotação.
- Slippage — Diferença entre o preço esperado e o preço efetivo de execução de uma ordem.
- Spread — Diferença entre o preço de compra (bid) e o de venda (ask) de um ativo.
- Survivorship bias — Viés de análise que considera apenas casos que “sobreviveram” a um processo, ignorando os demais.
- Splits – No contexto de mercados financeiros, refere-se à divisão ou agrupamento das ações de uma empresa, alterando o número de papéis em circulação sem modificar o valor total investido.
 - Stock Split (desdobramento): aumenta a quantidade de ações e reduz proporcionalmente o preço unitário.
 - Reverse Split (grupamento) ou Inplits: reduz a quantidade de ações e eleva proporcionalmente o preço unitário.
- Time Zone – No contexto do mercado financeiro, refere-se ao fuso horário adotado para registrar e sincronizar horários de negociações, cotações e eventos relevantes. A correta configuração do *time zone* é essencial para alinhar dados de diferentes bolsas e mercados globais, garantindo precisão na análise temporal e na execução de ordens.
- Candlestick – Tipo de gráfico utilizado no mercado financeiro que representa, por meio de velas, a abertura, fechamento, máxima e mínima dos preços de um ativo em um intervalo de tempo definido.

- Candles – Representações gráficas, no formato de velas (*candlestick*), que mostram abertura, fechamento, máxima e mínima de preços de um ativo em um intervalo de tempo no mercado financeiro.
- Spikes Espúrios – No mercado financeiro, referem-se a variações abruptas e atípicas nos preços de um ativo, geralmente causadas por erros de dados, baixa liquidez ou eventos não representativos das condições reais de mercado.
- Tick-by-Tick – No mercado financeiro, refere-se a dados que registram cada alteração de preço e volume negociado de um ativo, fornecendo o histórico completo das negociações em ordem cronológica.
- Timestamping – No mercado financeiro, é o processo de registrar com precisão o instante exato em que uma negociação, evento ou dado de mercado ocorre, permitindo análises temporais e sincronização de informações.
- APIs – No contexto de mercado financeiro e IA, são interfaces de programação que permitem a integração e troca de dados entre sistemas, possibilitando o acesso automatizado a cotações, históricos, ordens e outros serviços em tempo real.
- Min-Max Scaler – No contexto de IA aplicada ao mercado financeiro, é uma técnica de normalização que transforma os dados para um intervalo pré-definido, geralmente entre 0 e 1, preservando as proporções originais. É usada para padronizar variáveis, como preços e indicadores, antes do treinamento de modelos.
- Z-Score – No contexto de IA e mercado financeiro, é uma medida de padronização que indica quantos desvios-padrão um valor está distante da média. É utilizada para identificar anomalias, como preços atípicos ou variações incomuns em séries temporais.
- Walk-Forward – No contexto de IA e mercado financeiro, é uma técnica de validação em séries temporais que consiste em treinar um modelo em um intervalo de dados e testá-lo no período subsequente, avançando progressivamente no tempo para avaliar seu desempenho de forma contínua.
- Data Leakage – No mercado financeiro, é o uso indevido de dados futuros no treinamento de modelos de IA, causando previsões artificialmente precisas e sem validade real.
- Imputation – No mercado financeiro e IA, é o processo de preencher valores ausentes em séries temporais, garantindo consistência nos dados para análise ou treinamento de modelos.
- OHLCV – No mercado financeiro, sigla para *Open, High, Low, Close e Volume*, representando respectivamente abertura, máxima, mínima, fechamento e volume negociado de um ativo em um intervalo de tempo.

- Bayesiana – No contexto de IA e mercado financeiro, refere-se a métodos baseados no Teorema de Bayes, que atualizam probabilidades à medida que novas informações se tornam disponíveis, auxiliando na previsão e na tomada de decisão.
- Dropout – Em IA aplicada ao mercado financeiro, é uma técnica de regularização em redes neurais que desativa aleatoriamente neurônios durante o treinamento, reduzindo overfitting.
- Learning Rate – Em IA aplicada ao mercado financeiro, é o parâmetro que define o tamanho dos ajustes nos pesos da rede neural a cada iteração de treinamento.
- Early Stopping – Em IA para mercado financeiro, técnica que interrompe o treinamento ao detectar perda de desempenho em validação, evitando overfitting.
- Gradient Clipping – Em IA para mercado financeiro, técnica que limita o valor máximo dos gradientes no treinamento, evitando instabilidade e divergência do modelo.
- Scheduler – Em IA para mercado financeiro, mecanismo que ajusta automaticamente a *learning rate* durante o treinamento para melhorar a convergência do modelo.
- Binary Cross-Entropy – Em IA para mercado financeiro, função de perda usada em classificações binárias, medindo a diferença entre probabilidades previstas e valores reais.
- Drawdown – No mercado financeiro, é a redução percentual do valor de um ativo ou portfólio desde um pico até o menor ponto subsequente.
- backtest – Simulação histórica de uma estratégia sob regras e dados passados.
- Balanced Accuracy - Média de sensibilidade e especificidade; robusta a classes desbalanceadas.
- Banda morta - Zona neutra em torno de um limiar para ignorar microvariações e ruídos de microestrutura.
- Bandas de Bollinger - Envelope ao redor de uma média móvel definido por desvios-padrão, sinalizando volatilidade e possíveis suportes/resistências.
- Acurácia direcional (hit rate) - Proporção de acertos na previsão da direção (alta/baixa) do próximo movimento.
- Bid–ask bounce - Oscilação do preço entre bid e ask sem mudança no valor justo, gerando ruído.
- Bootstrap em blocos - Reamostragem por blocos contíguos para preservar dependência temporal em séries.

- Brier Score - Erro quadrático médio entre probabilidades previstas e resultados binários observados.
- Co-location - Hospedagem de servidores no data center da bolsa para reduzir latência de comunicação.
- Custos de transação - Despesas e fricções (corretagem, emolumentos, spread, slippage) consideradas nos backtests.
- Embargo temporal - Janela de exclusão após o treino para impedir contaminação temporal da validação/teste.
- F1-Score - Média harmônica entre precisão (precision) e revocação (recall).
- Latência - Atraso entre envio/recebimento de dados/ordens e sua efetiva execução/registro.
- Log-Loss - Perda logarítmica usada em classificadores probabilísticos; penaliza confiança indevida.
- LSTM (conceito) - Rede recorrente com memória de longo prazo para dependências temporais extensas.
- Microestrutura de mercado - Aspectos de formação de preço em alta frequência (livro de ofertas, tipos de ordens, latência).
- Otimização Bayesiana - Busca de hiperparâmetros guiada por modelo probabilístico do desempenho.
- Prophet - Modelo aditivo com tendências e sazonalidades marcado para séries temporais.
- Regimes de volatilidade - Estratificação do tempo por níveis de volatilidade para análises de robustez.
- Seed (semente aleatória) - Estado inicial do gerador pseudoaleatório para reproduzibilidade.
- Teste de Diebold–Mariano - Teste estatístico para comparar a acurácia preditiva de dois modelos.
- Transformers - Arquitetura baseada em atenção para dependências de longo alcance sem recorrência.
- Turnover - Taxa de giro da carteira (volume negociado relativo ao capital) em um período.
- Validação walk-forward - Treina e testa em janelas deslizantes preservando a ordem temporal.
- Volatilidade - Grau de oscilação dos preços em um período; associada à dispersão dos retornos.

- Índice de Sharpe - Retorno excedente por unidade de risco (desvio-padrão dos retornos).



SIGLAS

- API - Interface de Programação de Aplicações; permite integração e acesso automatizado a dados/serviços.
- ARIMA - Modelo AutoRegressivo Integrado de Médias Móveis para dependências lineares em séries temporais.
- AUC-PR - Área sob a curva Precisão–Revocação; indicada para classes desbalanceadas.
- CNN - Rede Neural Convolucional que extrai padrões locais em janelas temporais.
- DMA - Direct Market Access; envio de ordens diretamente ao sistema de negociação.
- ECE - Expected Calibration Error; discrepância média entre probabilidades previstas e frequências observadas.
- HFT - High-Frequency Trading; estratégias sensíveis a latências de milissegundos e à microestrutura.
- LSTM - Long Short-Term Memory; rede recorrente com memória de longo prazo.
- MCC - Coeficiente de Correlação de Matthews (-1 a 1) para qualidade de classificação binária.
- MME - Média Móvel Exponencial.
- OHLCV - Abertura (Open), Máxima (High), Mínima (Low), Fechamento (Close) e Volume (Volume).
- RSI - Relative Strength Index; oscilador de momento para sobrecompra/sobrevenda.