Predicting PRIO3 Stock Prices: PatchTST vs. LSTM

Genário Correia de Azevedo[1], **Tiago Buarque A. de Carvalho**[1],
Anderson Pinheiro Cavalcanti[2], **Itamar Bernardo S. Júnior**[3]

[1] Centro de Informática – Universidade Federal do Agreste de Pernambuco (UFAPE)
Av. Bom Pastor, s/n – Boa Vista – 55292-270 – Garanhuns – PE – Brazil

[2]Departamento de Computação – Universidade Federal Rural de Pernambuco (UFRPE)
Rua Dom Manoel de Medeiros, s/n – Dois Irmãos – 52171-900 – Recife – PE – Brazil

[3]Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE)
Campus Garanhuns – Rua Francisco Braga, s/n – Severiano de Moraes Filho
55298-320 – Garanhuns – PE – Brazil

{genario.azevedo,tiago.buarque}@ufape.edu.br

anderson.pinheiro@ufrpe.br, itamarbernardo2013@gmail.com

Abstract. *Financial time series forecasting is highly challenging due to data volatility and noise, limitations frequently encountered by recurrent models such as LSTM. This paper evaluates the Patch Time Series Transformer (PatchTST) model in predicting the closing price of the PRIO3 stock. Through a rigorous walk-forward validation protocol with daily retraining, PatchTST was compared against a baseline LSTM across four experimental configurations involving PETR4 and Brent as covariates. The results demonstrate that PatchTST significantly outperforms LSTM in all scenarios, achieving a 95.5% reduction in Mean Squared Error (MSE) within the univariate approach. Furthermore, incorporating the sector-specific covariate PETR4 further optimized performance, yielding the lowest global MSE of 0.9593. We conclude that the architectural characteristics of PatchTST offer superior robustness and predictive power for short-term financial forecasting.*

Resumo. *A previsão de séries temporais financeiras é desafiadora devido à volatilidade e ao ruído dos dados, limitações frequentemente enfrentadas por modelos como a LSTM. Este artigo avalia o modelo Patch Time Series Transformer (PatchTST) na previsão do preço de fechamento da ação PRIO3. O PatchTST foi comparado a uma LSTM em quatro configurações experimentais envolvendo as covariáveis PETR4 e Brent. Os resultados demonstram que o PatchTST superou a LSTM em todos os cenários, reduzindo o Erro Quadrático Médio (MSE) em 95,5% na abordagem univariada. A inclusão da covariável setorial PETR4 otimizou ainda mais o desempenho, alcançando o menor MSE global (0,9593). Conclui-se que a arquitetura do PatchTST oferece robustez e capacidade preditiva superior para o mercado financeiro de curto prazo.*

1. Introdução

A previsão de séries temporais financeiras constitui um problema clássico e ainda em aberto na interseção entre aprendizado de máquina, econometria e finanças computacionais. A dificuldade decorre, entre outros fatores, da elevada volatilidade dos ativos, da

presença de ruído, da não estacionariedade e da influência simultânea de fatores micro e macroeconômicos [Fama 1970, Cont 2001, Tsay 2010]. Em mercados reais, essas características tornam limitada a eficácia de abordagens que dependem de suposições fortes de linearidade ou estabilidade estatística ao longo do tempo.

Modelos estatísticos clássicos, como o ARIMA, permanecem relevantes em diferentes aplicações de previsão temporal, sobretudo pela interpretabilidade e pelo formalismo consolidado [Box et al. 2015]. Contudo, em séries financeiras, sua capacidade de representação tende a ser restrita diante de relações não lineares, mudanças de regime e estruturas complexas de dependência temporal. Esse cenário impulsionou o uso de técnicas de *Deep Learning*, particularmente redes neurais recorrentes, como as *Long Short-Term Memory* (LSTM), que se tornaram uma referência para modelagem sequencial em diversos domínios.

Embora as LSTM tenham avançado o estado da prática em tarefas temporais, sua natureza recorrente impõe limitações importantes. Em especial, o processamento sequencial reduz o paralelismo computacional e dificulta a captura eficiente de dependências de longo alcance, sobretudo em sequências extensas e ruidosas [Bengio 1994, Hochreiter and Schmidhuber 1997]. Mais recentemente, arquiteturas baseadas em atenção passaram a ocupar papel central na modelagem de sequências. O Transformer [Vaswani et al. 2017], ao substituir recorrência por mecanismos de autoatenção, tornou possível modelar interações globais entre diferentes instantes da série de forma mais direta.

No domínio de séries temporais, uma das propostas recentes mais relevantes é o *Patch Time Series Transformer* (PatchTST) [Nie et al. 2023]. O modelo introduz duas ideias particularmente interessantes para cenários financeiros: a segmentação da janela temporal em *patches*, reduzindo redundância local e favorecendo a extração de padrões mais estáveis, e a independência entre canais, evitando que a mistura precoce entre variáveis introduza ruído ou correlações espúrias. Essas propriedades tornam a arquitetura especialmente atraente para séries multivariadas com covariáveis heterogêneas.

Neste contexto, este trabalho investiga o desempenho do PatchTST na previsão de curtíssimo prazo do preço de fechamento da ação PRIO3, ativo do setor petrolífero brasileiro. A escolha se justifica por se tratar de um papel exposto simultaneamente à dinâmica do mercado doméstico e à oscilação de *commodities* globais, constituindo um caso relevante para avaliar a utilidade de covariáveis financeiras e macroeconômicas. O modelo é comparado a uma LSTM de referência empregada em estudo anterior [Silva Junior 2024], mantendo-se um protocolo experimental controlado, com validação *walk-forward* e retreinamento diário.

Dessa forma, este trabalho realiza uma avaliação empírica controlada do PatchTST no contexto do mercado acionário brasileiro, comparando-o diretamente com um modelo LSTM sob condições experimentais equivalentes. Adicionalmente, são analisadas diferentes configurações de entrada, incluindo cenários univariados e multivariados, com o objetivo de investigar o impacto de covariáveis setoriais e macroeconômicas na qualidade das previsões.

O restante do artigo está organizado da seguinte forma. A Seção 2 discute os fundamentos teóricos e os trabalhos relacionados. A Seção 3 descreve os dados, a arquitetura avaliada e o protocolo experimental. A Seção 4 apresenta os resultados e sua discussão.

Por fim, a Seção 5 resume os principais achados, limitações e direções futuras.

2. Referencial Teórico

Esta seção apresenta, de forma sintética, os principais conceitos que sustentam o trabalho: as propriedades das séries financeiras, as limitações das arquiteturas recorrentes e a motivação para o uso do PatchTST em previsão temporal.

2.1. Séries Temporais Financeiras

Séries temporais financeiras representam a evolução de preços, retornos ou volumes observados ao longo do tempo. Diferentemente de séries oriundas de sistemas físicos mais estáveis, elas refletem decisões agregadas de agentes econômicos, eventos exógenos e mudanças frequentes de regime. Por essa razão, exibem propriedades empíricas bem documentadas na literatura, frequentemente chamadas de "fatos estilizados", como não estacionariedade, assimetria, caudas pesadas e aglomeração de volatilidade [Cont 2001, Tsay 2010, Engle 1982]. Essas características impõem desafios importantes à modelagem preditiva. Em particular, a dependência temporal observada nem sempre é estável ao longo do horizonte amostral, e relações úteis em um intervalo podem perder validade em outro. Em aplicações financeiras, portanto, o desempenho de um modelo depende não apenas de sua capacidade de ajuste, mas também de sua robustez diante de ruído e mudanças estruturais.

2.2. Modelos Recorrentes e LSTM

As Redes Neurais Recorrentes foram concebidas para modelar dependências em dados sequenciais por meio de estados ocultos propagados ao longo do tempo. Apesar de seu apelo conceitual, RNNs simples sofrem com o problema do desvanecimento do gradiente, o que dificulta a aprendizagem de dependências de longo prazo [Bengio 1994]. A arquitetura LSTM [Hochreiter and Schmidhuber 1997] atenuou esse problema ao introduzir uma célula de memória e mecanismos de portas que regulam o fluxo de informação. Apesar de seu sucesso em previsão financeira [Fischer 2018], a LSTM mantém limitações relevantes. O processamento estritamente sequencial dificulta paralelização, aumenta o custo de treinamento e pode restringir a modelagem de interações temporais mais longas ou dispersas. Tais limitações motivaram o avanço de arquiteturas baseadas em atenção.

2.3. Transformers para Séries Temporais

O Transformer [Vaswani et al. 2017] substitui a recorrência por mecanismos de autoatenção, permitindo que o modelo compare diretamente diferentes posições de uma sequência ao construir suas representações internas. No entanto, a aplicação direta de Transformers a séries temporais ponto a ponto apresenta desafios. A granularidade muito fina dos sinais pode introduzir redundância e ruído, além de elevar o custo computacional do mecanismo de atenção. Em séries financeiras, nas quais pequenas oscilações locais podem não conter sinal preditivo consistente, essa limitação se torna ainda mais relevante.

2.4. PatchTST

O PatchTST [Nie et al. 2023] foi proposto para tornar a modelagem Transformer mais adequada ao domínio temporal. Sua principal ideia é representar a série de entrada por meio de *patches*, isto é, subsegmentos contíguos da janela temporal. Em vez de tratar

cada ponto como um token independente, o modelo passa a operar sobre blocos locais, o que reduz o comprimento efetivo da sequência e favorece a extração de padrões de maior nível. Além disso, o PatchTST adota independência entre canais (*channel independence*), processando cada variável separadamente antes da agregação final. Em séries multivariadas, essa escolha pode ser vantajosa quando as covariáveis apresentam escalas distintas, ruído específico ou correlações instáveis. O modelo também pode incorporar *Reversible Instance Normalization* (RevIN), mecanismo útil para mitigar efeitos de mudança de distribuição entre janelas. No contexto deste trabalho, essas características tornam o PatchTST uma alternativa plausível para previsão financeira de curto prazo, especialmente em cenários em que se deseja avaliar se covariáveis externas realmente agregam informação ao ativo-alvo.

3. Metodologia

3.1. Base de dados e configurações experimentais

O estudo utiliza séries diárias referentes ao período de abril de 2022 a abril de 2023. A variável-alvo é o preço de fechamento da ação PRIO3. Como covariáveis, foram consideradas: (i) a ação PETR4, representando informação setorial do mercado brasileiro de petróleo e gás; e (ii) o preço do Brent, representando um fator macroeconômico global associado à dinâmica internacional da commodity.

Após sincronização temporal via *inner join*, a base consolidada totalizou aproximadamente 250 pregões úteis. Para investigar o efeito de diferentes composições de entrada, foram definidas quatro configurações:

1. **Univariada**: apenas PRIO3;
2. **Multivariada-PETR4**: PRIO3 + PETR4;
3. **Multivariada-Brent**: PRIO3 + Brent;
4. **Multivariada-Completa**: PRIO3 + PETR4 + Brent.

Essa organização permite realizar uma análise de ablação simples, com foco em mensurar se a adição de variáveis exógenas melhora efetivamente a previsão de curto prazo.

3.2. Modelo avaliado

Foi empregada a versão supervisionada do PatchTST [Nie et al. 2023], configurada para previsão de um passo à frente. A janela de entrada foi definida como $seq_len = 30$, correspondendo a 30 dias úteis, e o horizonte preditivo como $pred_len = 1$.

A Tabela 1 resume os principais hiperparâmetros utilizados. A configuração segue valores compatíveis com a literatura do modelo e foi mantida fixa entre os cenários, de modo a preservar comparabilidade entre os experimentos.

Tabela 1. Hiperparâmetros do modelo PatchTST

Hiperparâmetro	Valor	Justificativa / Observação
seq_len	30	Janela histórica de 30 pregões
pred_len	1	Previsão de um dia à frente
patch_len	16	Valor compatível com a configuração-base do modelo [Nie et al. 2023]
stride	8	Sobreposição moderada entre *patches*
d_model	128	Dimensão do espaço latente
n_heads	8	Número de cabeças de atenção
e_layers	2	Profundidade do codificador Transformer
revin	ativado	Mitigação de variações de escala e não estacionariedade
subtract_last	ativado	Estratégia comum em séries financeiras de curto prazo
dropout / fc_dropout	0.05	Regularização leve

3.3. Protocolo de avaliação

Adotou-se validação *walk-forward* com retreinamento diário. Nesse protocolo, a cada novo instante de teste, o modelo é treinado apenas com dados passados e utilizado para prever o próximo valor da série. Em seguida, a janela temporal é deslocada e o processo é repetido. Esse desenho reduz o risco de vazamento de informação e produz uma avaliação mais próxima das condições encontradas em uso real.

A opção por retreinamento a cada passo foi motivada pela natureza não estacionária do problema. Em séries financeiras, padrões úteis podem mudar rapidamente, e protocolos mais simples de divisão fixa treino-teste frequentemente oferecem uma estimativa otimista ou pouco realista do desempenho fora da amostra.

Para reduzir o efeito da aleatoriedade associada à inicialização dos pesos e ao processo de otimização, cada configuração foi executada 10 vezes de forma independente. As métricas reportadas correspondem às médias dessas execuções.

3.4. Métricas e ambiente computacional

Foram utilizadas três métricas de erro: Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE) e Erro Percentual Absoluto Médio (MAPE). O MSE penaliza erros maiores com mais intensidade, o MAE fornece uma medida direta da magnitude média do erro, e o MAPE facilita interpretação relativa em termos percentuais.

Os experimentos foram executados em ambiente padronizado, com processador Intel Core i5 de 8ª geração e 8 GB de RAM. Embora não se trate de infraestrutura de alto desempenho, essa escolha reforça a viabilidade do modelo em contexto computacional acessível, particularmente para previsão diária.

4. Resultados e Discussão

4.1. Comparação com o baseline LSTM

O primeiro resultado relevante deste estudo é a vantagem consistente do PatchTST em relação ao baseline LSTM reportado em [Silva Junior 2024]. No cenário univariado, o PatchTST obteve MSE médio de 1,0584, enquanto a LSTM de referência apresentou MSE de 23,40. Considerando o mesmo problema de previsão e o mesmo contexto experimental, essa diferença corresponde a uma redução de 95,5% no MSE.

Esse resultado sugere que a representação por *patches*, aliada ao mecanismo de atenção, é particularmente adequada para capturar padrões úteis no horizonte de um dia à frente. Em vez de processar a série ponto a ponto, o modelo passa a operar sobre blocos locais, o que pode atenuar ruído de alta frequência e favorecer a aprendizagem de estruturas temporais mais informativas.

Ao mesmo tempo, convém interpretar esse ganho com cautela. Em previsão financeira, diferenças expressivas entre modelos podem depender fortemente do ativo, do intervalo amostral e do protocolo de avaliação. Assim, embora o resultado seja robusto dentro do escopo estudado, ele não deve ser automaticamente generalizado para outros ativos ou horizontes sem validação adicional.

4.2. Efeito das covariáveis

A Tabela 2 apresenta os resultados consolidados das quatro configurações avaliadas. Observa-se que o PatchTST superou o baseline LSTM em todos os cenários, com reduções de erro variando conforme o conjunto de entradas.

Tabela 2. Comparativo global dos experimentos com o PatchTST e baseline LSTM

Exp.	Configuração	Covariáveis	MSE (PatchTST)	MAE	MAPE (%)	MSE (LSTM)	Redução vs LSTM
1	Univariado	Nenhuma	1,0584	0,7858	2,194%	23,4000	95,5%
2	Multivariado	PETR4	**0,9593**	**0,7362**	**2,047%**	2,3300	58,8%
3	Multivariado	BRENT	1,0969	0,7736	2,163%	5,5100	80,1%
4	Multivariado	PETR4, BRENT	1,0455	0,7552	2,102%	2,6500	60,5%

A melhor configuração foi a multivariada com PETR4, que atingiu MSE de 0,9593. Esse resultado indica que a informação setorial doméstica parece ser mais útil para a previsão de PRIO3 do que a variável Brent considerada isoladamente. Uma interpretação plausível é que PETR4, por refletir parte da dinâmica do mesmo segmento econômico e do mesmo mercado acionário, compartilhe sinais mais diretamente alinhados ao comportamento de curto prazo da série-alvo.

Por outro lado, a configuração com Brent não melhorou o desempenho em relação ao cenário univariado, e a combinação PETR4 + Brent tampouco superou o uso exclusivo de PETR4. Esse achado é relevante porque mostra que adicionar mais variáveis não implica, necessariamente, maior poder preditivo. Em janelas curtas, covariáveis globais podem introduzir ruído, defasagens ou informação redundante, especialmente quando o modelo já dispõe de uma covariável setorial mais próxima do comportamento do ativo-alvo.

A Figura 1 sintetiza essa comparação e torna evidente que o ganho do PatchTST não depende de uma única configuração, embora a utilidade das covariáveis varie substancialmente.

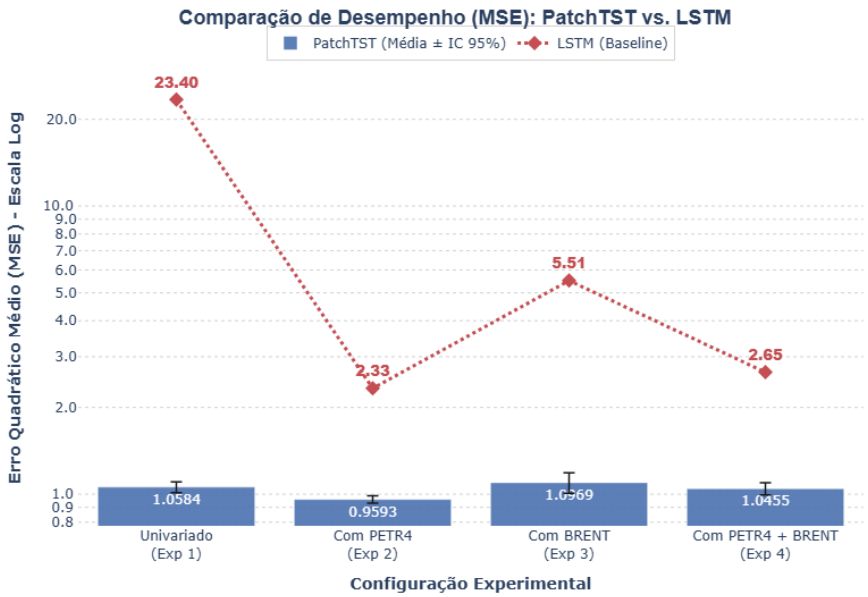


Figura 1. Comparação do MSE médio do PatchTST nas diferentes configurações experimentais, em contraste com o baseline LSTM.

4.3. Estabilidade entre execuções

Além do desempenho médio, um aspecto importante em aplicações financeiras é a estabilidade do modelo. A Figura 2 mostra a evolução do MSE ao longo das 10 execuções independentes do cenário univariado. Os valores permanecem concentrados em torno da média global, sem flutuações abruptas, sugerindo baixa sensibilidade à aleatoriedade da inicialização.

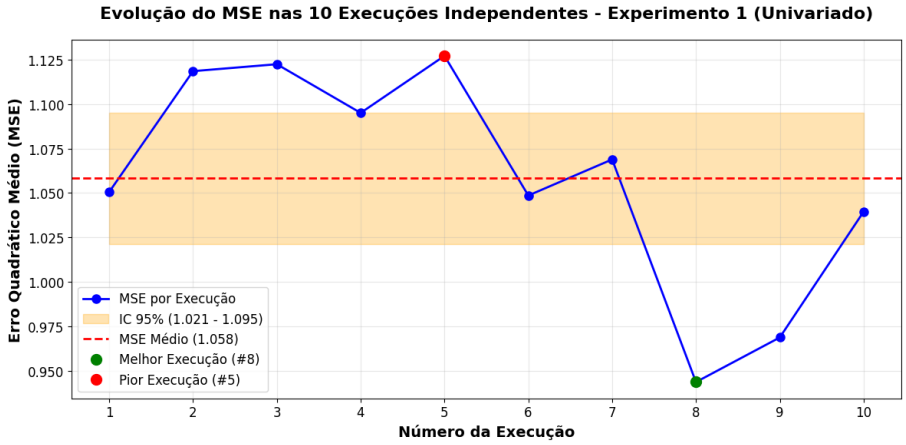


Figura 2. Evolução do MSE nas 10 execuções independentes do PatchTST no cenário univariado.

O intervalo de confiança de 95% para o MSE médio nesse cenário foi estimado em [1,0134, 1,1034], o que reforça a consistência da estimativa. Embora essa análise não substitua testes estatísticos mais abrangentes entre modelos, ela fornece evidência de que o desempenho observado não decorre de um resultado isolado.

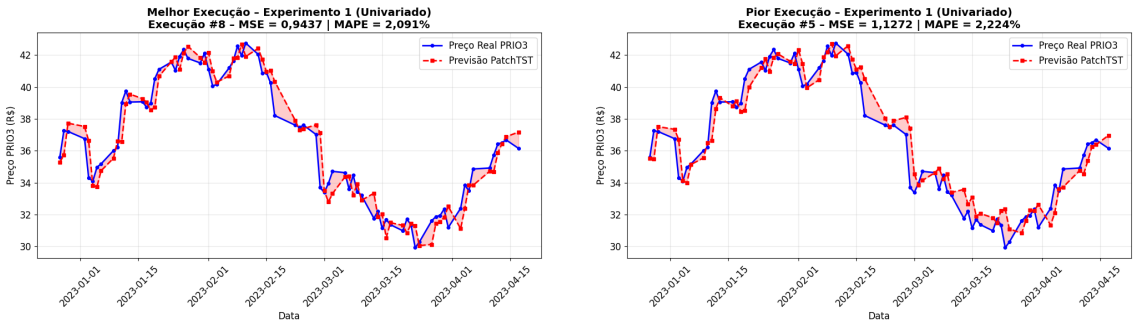
A Tabela 3 detalha os resultados individuais das 10 execuções no cenário univariado. Nota-se que os tempos médios por predição se mantiveram próximos de 3,71 ms, sugerindo estabilidade também no custo computacional do processo de inferência.

Tabela 3. Métricas individuais do PatchTST univariado (PRIO3)

Execução	MSE	MAE	MAPE (%)	Tempo médio por predição (ms)
1	1,0506	0,7852	2,184	3,74
2	1,1186	0,7931	2,219	3,44
3	1,1226	0,8300	2,317	3,46
4	1,0951	0,8042	2,250	3,47
5	1,1272	0,7941	2,224	4,18
6	1,0487	0,7758	2,168	4,34
7	1,0690	0,7956	2,222	3,56
8	0,9437	0,7477	2,091	3,90
9	0,9688	0,7556	2,099	3,53
10	1,0396	0,7768	2,167	3,53
Média	**1,0584**	**0,7858**	**2,194**	**3,71**

4.4. Análise qualitativa das previsões

A Figura 3 apresenta as execuções com melhor e pior desempenho no cenário univariado. Em ambos os casos, o modelo preserva adequadamente a tendência local da série e acompanha a maior parte dos movimentos observados. Os desvios mais visíveis concentram-se em pontos de maior volatilidade, o que é compatível com a dificuldade inerente de prever oscilações abruptas em séries financeiras.



(a) Melhor execução (Execução #8, MSE = 0,9437).

(b) Pior execução (Execução #5, MSE = 1,1272).

Figura 3. Comparação entre preços reais e previstos nas execuções extremas do cenário univariado.

Esse comportamento é importante porque sugere que o PatchTST não apenas reduz erro médio, mas também mantém consistência qualitativa ao longo de diferentes execuções. Ainda assim, os resultados não devem ser confundidos com capacidade de antecipar integralmente movimentos abruptos de mercado. Em aplicações reais, a utilidade operacional do modelo dependeria também de custos de transação, sensibilidade a mudanças de regime e desempenho em períodos mais longos e heterogêneos.

4.5. Síntese dos achados

Em conjunto, os resultados apontam três conclusões principais. Primeiro, o PatchTST mostrou-se superior ao baseline LSTM no problema investigado. Segundo, a inclusão de uma covariável setorial nacional (PETR4) foi mais útil do que a inclusão de uma variável macroeconômica global (Brent) para o horizonte de um dia. Terceiro, o modelo apresentou estabilidade razoável entre execuções, tanto em termos de erro quanto de custo computacional.

Esses achados são coerentes com a hipótese de que, em previsão financeira de curtíssimo prazo, sinais mais próximos do contexto setorial e do próprio mercado doméstico podem ser mais informativos do que variáveis globais agregadas. Também são consistentes com a proposta do PatchTST, que favorece a extração de padrões temporais locais e reduz interferência entre canais heterogêneos.

5. Conclusão

Este artigo avaliou a arquitetura PatchTST na previsão de um dia à frente do preço de fechamento da ação PRIO3, considerando um protocolo experimental rigoroso baseado em validação *walk-forward* com retreinamento diário. Em todos os cenários analisados, o modelo superou o baseline LSTM de referência, indicando que a combinação entre segmentação em *patches* e mecanismos de atenção constitui uma alternativa promissora para previsão financeira de curto prazo.

Os resultados também mostraram que a escolha das covariáveis influencia de maneira relevante o desempenho. Entre os cenários multivariados, a inclusão da PETR4 produziu o melhor resultado, enquanto o uso de Brent, isoladamente ou em combinação, não levou a ganhos adicionais consistentes. Esse achado reforça que a simples ampliação do conjunto de entrada não garante melhora preditiva e que a seleção de variáveis deve considerar proximidade econômica, regime temporal e potencial introdução de ruído.

Do ponto de vista metodológico, o trabalho contribui ao investigar o PatchTST em um contexto ainda pouco explorado na literatura nacional, utilizando um ativo do mercado brasileiro e um protocolo de avaliação mais realista do que divisões estáticas simples de treino e teste. Adicionalmente, a repetição de experimentos permitiu observar estabilidade razoável dos resultados, o que fortalece a confiabilidade das conclusões no escopo avaliado.

Embora os resultados tenham sido consistentes no contexto avaliado, o estudo concentrou-se em um único ativo, em um horizonte de previsão de um passo e em um intervalo temporal específico. Assim, os achados deste trabalho devem ser interpretados no escopo experimental considerado, indicando o potencial do PatchTST para previsão financeira de curto prazo no mercado brasileiro. Estudos futuros com múltiplos ativos,

horizontes preditivos mais longos e janelas temporais mais extensas podem aprofundar a análise da capacidade de generalização do modelo.

Referências

Bengio, Y. (1994). Learning long-term dependencies. *IEEE Transactions on Neural Networks*, 5(2):157–166.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2015). *Time Series Analysis: Forecasting and Control*. Wiley, San Francisco.

Cont, R. (2001). Empirical properties of asset returns. *Quantitative Finance*, 1(2):223–236.

Engle, R. (1982). Autoregressive conditional heteroscedasticity. *Econometrica*, 50(4):987–1007.

Fama, E. (1970). Efficient capital markets. *The Journal of Finance*, 25(2):383–417.

Fischer, T. (2018). Deep learning with lstm. *European Journal of Operational Research*, 270(2):654–669.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Nie, Y., Nguyen, N. H., Sinthong, P., and Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.

Silva Junior, I. B. (2024). Previsão do preço de ações usando lstm. Trabalho de Conclusão de Curso, UFAPE.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. Wiley, Hoboken, 3 edition.

Vaswani, A. et al. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30.