MODELOS DE PROJEÇÃO DA TAXA DE CÂMBIO DA MOEDA BRASILEIRA CONTRA A MOEDA AMERICANA

um estudo comparativo de quatro arquiteturas de redes neurais em alta frequência

Caroline Bernal Penna (IC) e Eli Hadad Junior (Orientador)

Apoio: PIBIC CNPq

RESUMO

O presente artigo compara o desempenho preditivo de quatro arquiteturas de redes neurais (LSTM, N-BEATS, N-HiTS e TCN) frente ao modelo de passeio aleatório, na projeção da taxa de câmbio entre o real e o dólar em frequência intradiária. O problema de pesquisa é verificar se essas arquiteturas superam o benchmark ingênuo que a literatura consagrou como critério mínimo para a previsão cambial. Utilizaram-se dados de um minuto do Sistema Bloomberg, entre janeiro de 2024 e abril de 2025, submetidos a tratamento que incluiu a reconstrução da linha temporal, a correção da conversão da cotação, a delimitação da sessão de negociação a partir de fontes oficiais do Banco Central e da B3 e a exclusão das variações que atravessam a virada de pregão. A série foi caracterizada por testes de raiz unitária, razão de variâncias, agrupamento de volatilidade e normalidade, calculados exclusivamente sobre a amostra de treino. As quatro arquiteturas foram treinadas sob protocolo experimental idêntico e avaliadas em quatro frequências, com comparação formal por meio do teste de Diebold-Mariano e de alternativas livres de suposição distributiva. Em quinze das dezesseis combinações, o erro superou o do passeio aleatório, mas, em nenhuma delas, a diferença mostrou-se estatisticamente significativa. A hipótese de superioridade das redes neurais não se confirmou. A vantagem do benchmark concentra-se integralmente nos intervalos em que o preço não se altera; nos intervalos em que há movimento, o desempenho das arquiteturas não revela vantagem sistemática.

Palavras-chave: Câmbio. Redes neurais. Alta frequência.

ABSTRACT

This article compares the predictive performance of four neural network architectures (LSTM, N-BEATS, N-HiTS and TCN) against the random walk model in forecasting the Brazilian real to United States dollar exchange rate at intraday frequency. The research problem is to determine whether these architectures outperform the naive benchmark that the literature has established as the minimum criterion for exchange rate forecasting. One-minute data from the Bloomberg System, covering January 2024 to April 2025, were processed in order to correct the reconstruction of the time stamp and the quotation convention, to delimit the trading session on the basis of official sources from the Central Bank of Brazil and from B3, and to

exclude the variations that span the interval between trading sessions. The series was characterized through unit root tests, the variance ratio test, and tests of volatility clustering and normality, all computed exclusively on the training sample. The four architectures were trained under an identical experimental protocol and evaluated at four frequencies, with formal comparison carried out through the Diebold-Mariano test and through alternatives free of distributional assumptions. In fifteen of the sixteen combinations the error exceeded that of the random walk, and in none of them the difference was statistically significant. The hypothesis that the neural networks would be superior was therefore not confirmed. The advantage of the benchmark is concentrated entirely in the intervals in which the price does not change; where movement does occur, the performance of those architectures is equivalent to chance.

Keywords: Exchange rate. Neural networks. High Frequency.

INTRODUÇÃO

A taxa de câmbio entre o real e o dólar é um dos preços mais influentes da economia brasileira. Sua variação atravessa custos de produção, inflação, comércio exterior e decisões de investimento (CAMPA; GOLDBERG, 2005; KRUGMAN; OBSTFELD, 2018). Prever seu comportamento com maior precisão interessa a empresas, investidores e formuladores de política, mas trata-se de tarefa notoriamente difícil: a série apresenta não estacionariedade, volatilidade variável e ruído, características que limitam os métodos econométricos tradicionais (HAMILTON, 1994).

A dificuldade tem nome na literatura. Meese e Rogoff (1983) mostraram que modelos macroeconômicos estruturais não superam consistentemente uma previsão ingênua, o passeio aleatório (*random walk*), segundo o qual a melhor estimativa para o próximo valor é o valor corrente. Desde então, superar esse referencial tornou-se o critério mínimo para que um modelo de câmbio seja considerado útil.

As redes neurais profundas (*deep learning*) surgem como alternativa por aprenderem relações não lineares diretamente dos dados, sem exigir especificação prévia de forma funcional. Quatro arquiteturas se destacaram na previsão de séries temporais: a LSTM (HOCHREITER; SCHMIDHUBER, 1997), o N-BEATS (ORESHKIN et al., 2020), o N-HiTS (CHALLU et al., 2022) e a TCN (BAI; KOLTER; KOLTUN, 2018). Estudos aplicados relatam bom desempenho dessas redes em séries financeiras (CHEN; ZHOU; DAI, 2015; SAGHEER; KOTB, 2019), mas raramente as comparam entre si sob protocolo idêntico, e mais raramente ainda no câmbio brasileiro em frequência intradiária.

O trabalho mais próximo é o de Rossi et al. (2022), que investigou a previsibilidade do câmbio brasileiro a 1, 5 e 15 minutos empregando regressão linear e aprendizado de máquina (*machine learning*), e concluiu que superar o passeio aleatório com preditores defasados exige incorporar o mercado futuro de real. Não há, até onde se verificou, comparação sistemática das quatro arquiteturas de aprendizado profundo acima sobre o câmbio brasileiro intradiário, lacuna que esta pesquisa busca preencher.

Problema de pesquisa: Qual o desempenho comparativo dos modelos de projeção da taxa de câmbio baseados nas redes neurais LSTM, N-BEATS, N-HiTS e TCN, frente a um modelo de passeio aleatório, para o mercado brasileiro, utilizando dados de alta frequência?

Objetivo geral: Analisar e comparar o desempenho dessas quatro arquiteturas na projeção da taxa de câmbio BRL/USD.

Objetivos específicos: Coletar e pré-processar dados de alta frequência da taxa de câmbio BRL/USD; desenvolver e treinar os modelos sob protocolo experimental idêntico; avaliar o

desempenho preditivo por métricas de erro adequadas; e comparar os resultados para identificar o modelo de melhor desempenho para a série em estudo.

Hipótese: Espera-se que os modelos de redes neurais apresentem capacidade preditiva superior à do passeio aleatório para a taxa de câmbio BRL/USD.

REFERENCIAL TEÓRICO

O passeio aleatório como referencial

O modelo de passeio aleatório postula que a melhor previsão para o próximo valor de uma série é o seu valor corrente. Apesar da simplicidade, tornou-se referência obrigatória na previsão cambial após Meese e Rogoff (1983), que demonstraram empiricamente a dificuldade de modelos estruturais em superá-lo fora da amostra em horizontes curtos.

A dificuldade persiste na literatura brasileira. Gaglianone e Marins (2016), em estudo do Banco Central do Brasil, avaliaram previsões de ponto e de densidade para o câmbio brasileiro e verificaram, por meio dos testes de Diebold e Mariano (1995) e de Giacomini e White, em quais horizontes é possível superar o passeio aleatório.

Mais do que um competidor, o passeio aleatório é uma hipótese estatística testável. Lo e MacKinlay (1988) propuseram o teste de razão de variâncias, que explora uma propriedade direta: sob passeio aleatório a variância cresce em proporção ao intervalo, de modo que a variância do retorno acumulado em q períodos deve ser exatamente q vezes a de um período. A razão entre ambas vale 1 sob a hipótese nula; valores acima indicam persistência e valores abaixo, reversão. Os autores apresentam versão da estatística robusta à heterocedasticidade, apropriada a séries financeiras.

Arquiteturas de redes neurais para séries temporais

A LSTM (HOCHREITER; SCHMIDHUBER, 1997), evolução das redes recorrentes de Elman (1990), emprega unidades de memória com portões que regulam seletivamente o fluxo de informação, permitindo reter dependências longas e descartar o que não é útil (GERS; SCHMIDHUBER; CUMMINS, 2000).

O N-BEATS (ORESHKIN et al., 2020) adota abordagem distinta: rede puramente feed-forward organizada em blocos e pilhas, em que cada bloco explica parte do sinal e transmite adiante o resíduo. As funções-base podem ser aprendidas dos dados ou pré-definidas pelo usuário.

O N-HiTS (CHALLU et al., 2022) evolui do N-BEATS acrescentando amostragem multi-taxa: cada pilha processa a série em resolução distinta, especializando-se numa escala temporal. Essa é sua contribuição própria; sem ela, o modelo reduz-se ao N-BEATS. A aplicação do N-HiTS à previsão cambial ainda é incipiente (APTE; HARIBHAKTA, 2024).

A TCN (BAI; KOLTER; KOLTUN, 2018) adapta convoluções a dados sequenciais por meio de três elementos: convoluções causais, que garantem que a previsão em t use apenas informação até t−1; convoluções dilatadas, que ampliam o alcance sem multiplicar parâmetros; e blocos residuais, que permitem empilhar camadas sem degradar o treinamento.

Redes neurais aplicadas ao câmbio

Estudos relatam desempenho favorável de redes neurais em séries financeiras, frequentemente superando modelos mais simples (CHEN; ZHOU; DAI, 2015; PATEL et al., 2015; KIM; PARK; CHO, 2018; CHEN; GE, 2019; SAGHEER; KOTB, 2019). Boa parte dessa literatura, contudo, trata de ações, e nem sempre adota o passeio aleatório como referencial. Abouzaid e Boussedra (2025), por exemplo, comparam MLP, Random Forest e LSTM no par EUR/USD em frequência mensal, com melhor desempenho da LSTM, mas sem confrontá-las com o benchmark ingênuo.

No câmbio brasileiro em alta frequência, Rossi et al. (2022) investigaram a previsibilidade a 1, 5 e 15 minutos com regressão linear, Lasso, Elastic Net, Random Forest e Gradient Boosting, empregando o teste de Diebold e Mariano (1995). Com preditores contemporâneos, superaram o passeio aleatório em todas as frequências. Com preditores exclusivamente defasados, isto é, no exercício de previsão propriamente dito, só o superaram ao incorporar o futuro de real como preditor adicional, na frequência de um minuto e até dois minutos à frente, o que os autores interpretam como evidência do papel do mercado futuro na formação do câmbio à vista.

Fatos estilizados de séries financeiras

Séries de retornos financeiros apresentam regularidades documentadas, os chamados fatos estilizados (*stylized facts*) (CONT, 2001): caudas mais pesadas que a normal, agrupamento de volatilidade (*volatility clustering*), em que períodos agitados sucedem períodos agitados, e aproximação gradual da normalidade conforme se agregam intervalos maiores.

Em frequências muito altas, soma-se um efeito de natureza distinta. O preço observado passa a refletir o funcionamento do próprio mercado: a alternância entre pontas de compra e venda produz oscilação sem conteúdo informacional, manifestada como

autocorrelação negativa de primeira ordem (ROLL, 1984). Distinguir esse ruído de microestrutura de previsibilidade genuína é condição para interpretar corretamente qualquer resultado em alta frequência.

METODOLOGIA

Esta pesquisa adota abordagem quantitativa para comparar o desempenho preditivo de quatro arquiteturas de redes neurais e do passeio aleatório.

Fonte e coleta dos dados

Foram utilizados dados da taxa de câmbio BRL/USD com frequência de um minuto, obtidos pelo Sistema Bloomberg, cobrindo de 2 de janeiro de 2024 a 3 de abril de 2025. O arquivo original contém 397.863 registros e, além do real, traz as cotações de euro, dólar canadense, franco suíço e dólar australiano, empregadas neste trabalho como moedas de controle para fins de diagnóstico.

Correções aplicadas à base

Duas correções foram necessárias antes de qualquer análise, e ambas são declaradas por afetarem a interpretação dos resultados.

A primeira é o carimbo de tempo. A coluna de data e hora do arquivo original emprega relógio de doze horas sem marcador de período: existem apenas horas de 1 a 12. Disso resultam 228.598 carimbos repetidos e uma coluna que não é crescente. Como a ordem das linhas está preservada, a linha temporal foi reconstruída: para cada registro, há dois horários possíveis (dia e relógio ou dia e relógio acrescido de doze horas) e, percorrendo o arquivo em ordem, adota-se o que não retrocede no tempo. A reconstrução foi submetida a dois testes independentes, ambos aprovados: nenhuma violação de monotonia em 397.863 linhas e coincidência integral com a data do arquivo original. Trata-se, ainda assim, de inferência e não de recuperação: a informação de período não existe no arquivo de origem, e a limitação é registrada na seção final.

A segunda é a convenção da cotação. A coluna do real traz valores entre 0,1586 e 0,2069, isto é, dólar por real, e não real por dólar. Como o objeto do estudo é a taxa BRL/USD, aplicou-se a inversão, que produz valores entre 4,8333 e 6,3052. As demais moedas já se encontram na convenção usual e não foram alteradas. A inversão foi verificada por reconstrução do valor original, com erro máximo da ordem de 10^{-17}.

Nenhum registro foi removido do arquivo. Todo tratamento é aplicado por marcação, de modo que a base final preserva as 397.863 linhas originais e permite conferência registro a registro.

Delimitação da sessão de negociação

O mercado de câmbio à vista é de balcão e não possui pregão com abertura e fechamento divulgados. A B3 mantém página de horários intitulada "Câmbio e Dólar Pronto", mas a tabela publicada contém apenas contratos futuros.

Duas instituições publicam horários que delimitam a negociação de câmbio no país. O Banco Central do Brasil, nos Horários de Sistemas Informatizados, informa que o sistema de registro das operações interbancárias (Interbancário Eletrônico, PCAM380) fica indisponível das 0h às 8h e das 17h01 às 24h para operações de compra, o que corresponde a funcionamento das 8h às 17h. A B3 publica o horário de negociação do futuro de dólar comercial, das 9h às 18h30, único instrumento de câmbio com sessão formalmente divulgada.

Para fins operacionais do estudo, adotou-se a interseção das duas janelas, das 9h às 17h, inclusive. Essa janela foi definida como o período em que o sistema de registro está disponível e o futuro de dólar se encontra em negociação.

Os dados são compatíveis com essa delimitação. Fora do intervalo adotado, a proporção de observações em que o preço efetivamente muda é de 0,00%; dentro dele, 32,89%. A hora das 8h, embora pertencente ao intervalo regulatório do Banco Central, apresenta apenas 0,02% de observações com variação; o sistema está disponível e o mercado não opera. A partir das 9h, quando o futuro abre, a proporção salta para 43,15%.

Calendário de dias sem negociação

A referência adotada é a lista de feriados nacionais da ANBIMA, escolha amparada pelos horários divulgados pelo Banco Central do Brasil, os Horários de Sistemas Informatizados registram que os sistemas de câmbio interbancário ficam indisponíveis em sábados, domingos e feriados nacionais. Os ofícios circulares da B3 foram consultados para os horários dos dias de expediente reduzido, informação que a lista da ANBIMA não fornece. No período estudado resultam treze datas sem sessão e três com sessão reduzida.

Registre-se que o calendário mudou dentro do período por força de lei: a Lei nº 14.759, de 21 de dezembro de 2023, declarou o 20 de novembro feriado nacional, e a B3 revogou seu calendário de 2024 no dia seguinte à publicação. A data pertence à amostra.

Caracterização da amostra

Após o tratamento, a amostra compreende 260 sessões de negociação e 85.982 observações de um minuto. A taxa variou de 4,8333, em 12 de janeiro de 2024, a 6,2933, em 19 de dezembro de 2024, o que representa amplitude de 30,2%, com variação líquida de 15,9% no período e volatilidade anualizada de 13,8%. Trata-se, portanto, de período de forte movimentação, circunstância considerada na discussão dos resultados.

Identificaram-se ainda 55 sessões em que a cotação do real permanece constante durante todo o pregão enquanto as moedas de controle variam normalmente. Todas contêm exatamente 124 observações válidas, contra mediana de 332 nas demais, e 35 delas caem em uma quarta-feira. A uniformidade e a concentração em um único dia da semana sugerem um padrão pouco compatível com o comportamento usual de mercado, e sua origem não pôde ser determinada. As sessões permanecem identificadas na base por marcação própria, e todo o estudo é replicado com e sem elas.

Alvo, variáveis e divisão da amostra

O alvo dos modelos é o retorno logarítmico, escolha fundamentada nos testes de raiz unitária reportados na seção seguinte. O preço é reconstruído a partir da previsão de retorno, de modo que as métricas permaneçam na escala original e diretamente comparáveis ao benchmark.

Como variáveis de entrada, além do próprio histórico de retornos, empregam-se os minutos decorridos desde a abertura da sessão e a volatilidade realizada, calculada exclusivamente sobre retornos passados. As moedas de controle foram avaliadas e não incorporadas: a correlação defasada com o retorno do real situa-se entre −0,0446 e +0,0432.

Os retornos que atravessam a virada de sessão foram removidos. A variação entre o fechamento de um pregão e a abertura do seguinte cobre cerca de dezesseis horas, e não o intervalo de amostragem; mantida, seria apresentada ao modelo como se fosse movimento intradiário. Seu peso cresce à medida que aumenta o intervalo de amostragem, alcançando 11,1% dos retornos a 60 minutos. A divisão é cronológica: 70% para treino, 10% para validação e 20% para teste.

Protocolo experimental

As quatro arquiteturas compartilham protocolo idêntico, implementado sobre a biblioteca Darts. A padronização das variáveis é ajustada exclusivamente sobre o conjunto de treino. A busca de hiperparâmetros emprega o algoritmo TPE, avaliado no erro quadrático médio de validação; o conjunto de teste não participa de nenhuma decisão. Cada configuração é treinada novamente com três sementes distintas, de modo a separar o desempenho da arquitetura do acaso da inicialização. O orçamento de busca, de vinte tentativas por arquitetura, foi fixado por medição, e não por convenção, conforme reportado na seção de resultados.

O benchmark é o passeio aleatório: a previsão para t é o valor observado em t−1, sobre exatamente os mesmos alvos, o que equivale a prever variação nula.

As métricas são RMSE, MAE, MAPE e acurácia direcional média, todas na escala de preço. A acurácia direcional compara o sinal da variação prevista com o da realizada; empates são excluídos e reportados, e a métrica não se aplica ao passeio aleatório, que por definição não emite direção.

Técnicas estatísticas

A série foi caracterizada antes de qualquer modelo, sobre a amostra de treino apenas, de modo que nenhuma decisão de modelagem incorpore informação do período de avaliação.

A presença de raiz unitária foi avaliada pelos testes de Dickey-Fuller aumentado e KPSS, empregados em conjunto por possuírem hipóteses nulas opostas (KWIATKOWSKI et al., 1992). A hipótese de passeio aleatório foi testada diretamente pela razão de variâncias de Lo e MacKinlay (1988), na versão robusta à heterocedasticidade. A existência de estrutura linear foi verificada pelo teste de Ljung-Box, e o agrupamento de volatilidade pelo teste ARCH (ENGLE, 1982). A constância da variância ao longo da sessão foi avaliada pelo teste de Levene, e a normalidade dos retornos pelo teste de Jarque-Bera.

A comparação entre cada modelo e o benchmark emprega o teste de Diebold e Mariano (1995) com a correção de amostra pequena de Harvey, Leybourne e Newbold (1997), acompanhado dos testes de sinais e de Wilcoxon propostos no mesmo trabalho de 1995, que dispensam suposição distributiva. O viés das previsões é avaliado pela regressão de Mincer e Zarnowitz (1969). Como são realizadas dezesseis comparações (quatro arquiteturas em quatro frequências), o nível de significância de 5% é ajustado por Bonferroni, resultando em 0,3125%.

RESULTADO E DISCUSSÃO

Caracterização da série e definição do alvo

Uma série com raiz unitária não retorna a uma média: o choque de hoje incorpora-se permanentemente ao nível, de modo que prever o nível implica modelar uma série sem mecanismo de reversão à média. Os testes de Dickey-Fuller aumentado e KPSS respondem a essa questão em conjunto, e o fato de suas hipóteses nulas serem opostas é o que torna a leitura conclusiva.

Tabela 1 – Testes de raiz unitária, proporção de retornos nulos e curtose dos retornos, por frequência de amostragem (amostra de treino)

frequência	ADF (preço)	ADF (retorno)	KPSS (preço)	KPSS (retorno)	retornos nulos	curtose
1 min	0,7498	0,0000	≤ 0,01	≥ 0,10	65,07 %	49,14
5 min	0,7486	0,0000	≤ 0,01	≥ 0,10	44,11 %	13,09

15 min	0,7469	0,0000	≤ 0,01	≥ 0,10	28,58 %	8,97
30 min	0,7579	0,0000	≤ 0,01	≥ 0,10	22,69 %	8,24
60 min	0,7673	0,0000	≤ 0,01	≥ 0,10	20,97 %	6,36

Fonte: elaborada pelos autores.

O resultado é unânime nas cinco frequências e nos dois testes: o preço possui raiz unitária e o retorno não. A série é integrada de ordem um, o que fundamenta a adoção do retorno logarítmico como alvo.

A curtose decresce monotonicamente à medida que a frequência diminui, comportamento conhecido como gaussianidade por agregação (CONT, 2001), mas nenhuma frequência se aproxima da normalidade: a menor curtose observada é mais que o dobro da normal, e o teste de Jarque-Bera rejeita a normalidade em todas elas. Daí a necessidade de acompanhar o teste de Diebold-Mariano com os testes de sinais e de Wilcoxon, que dispensam essa suposição.

O passeio aleatório testado diretamente na série

O passeio aleatório não é apenas o modelo de comparação: é hipótese que pode ser testada na própria série, antes de qualquer rede ser treinada.

Tabela 2 – Teste de razão de variâncias: p-valores da estatística robusta à heterocedasticidade, e indicadores de microestrutura, por frequência

frequência	**q = 2**	**q = 4**	**q = 8**	**q = 16**	**AC(1)**	**VR(2)**
1 min	0,0000	0,0000	0,0000	0,0000	−0,0693	0,9307
5 min	0,0003	0,0124	0,1502	0,5518	−0,0409	0,9591
15 min	0,6112	0,5350	0,4517	0,4084	+0,0091	1,0090
30 min	0,5470	0,4661	0,3310	0,2301	+0,0138	1,0134
60 min	0,2913	0,1223	0,0984	0,7709	+0,0309	1,0312

Fonte: elaborado pelos autores.

A 1 minuto, a hipótese é rejeitada para todos os valores de q examinados. A 5 minutos, a evidência depende do horizonte: há rejeição para q=2 e q=4, mas não para q=8 e q=16. A partir de 15 minutos, a hipótese não é rejeitada em nenhum dos valores de q.

A rejeição em alta frequência é microestrutura, não previsibilidade

A rejeição observada nas duas frequências mais altas exige leitura cuidadosa, e a própria tabela fornece os elementos. Nas duas frequências, o VR(2) situa-se abaixo de 1 e a autocorrelação de primeira ordem é negativa; isto é, a rejeição ocorre por reversão, que é a assinatura da alternância entre pontas de compra e venda descrita por Roll (1984), e não por previsibilidade aproveitável. Soma-se a isso a proporção de retornos nulos: a 1 minuto, o preço permanece inalterado em 65,07% dos intervalos.

A partir de 15 minutos, os sinais associados à microestrutura se atenuam: a autocorrelação de primeira ordem torna-se pequena e positiva, o VR(2) aproxima-se de 1 e a hipótese de passeio aleatório deixa de ser rejeitada nos valores de q examinados. A leitura conjunta é direta: a série é estatisticamente indistinguível de um passeio aleatório nas frequências em que a microestrutura não domina.

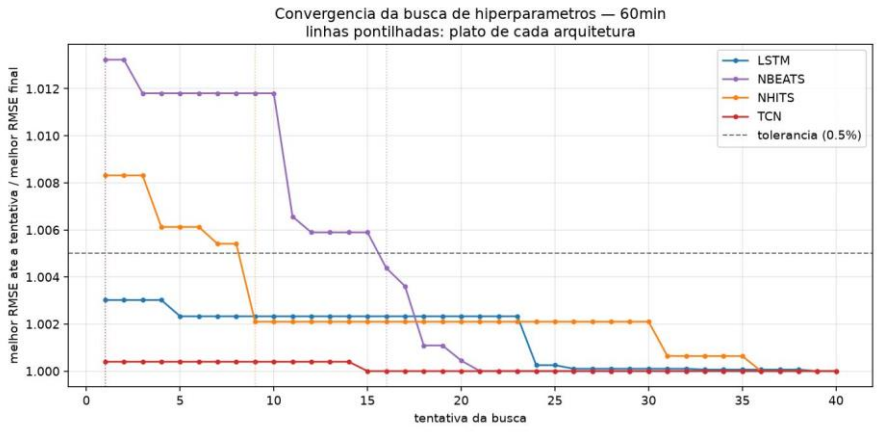
A frequência de 1 minuto foi excluída por concentrar a manifestação mais intensa desses efeitos, com 65,07% de retornos nulos e rejeição para todos os q; a frequência de 5 minutos foi mantida como escala limítrofe para verificar se tais efeitos ainda repercutem no desempenho preditivo.

O espaço de busca de hiperparâmetros é plano

Antes de comparar as arquiteturas, verificou-se quantas tentativas de busca cada uma exige para deixar de melhorar. A finalidade era assegurar tratamento equitativo: arquiteturas com mais hiperparâmetros poderiam ter seu espaço de busca explorado de forma mais superficial, e um desempenho inferior deixaria de ser atribuível à arquitetura.

A medição, com quarenta tentativas por arquitetura, mostrou algo distinto do esperado, apresentado no Gráfico 1. O melhor erro de validação melhora, no máximo, 1,3% entre a primeira e a quadragésima tentativa. Configurações muito diferentes de hiperparâmetros produzem erro praticamente igual.

Gráfico 1 – Convergência da busca de hiperparâmetros por arquitetura: melhor RMSE de validação até cada tentativa, relativo ao melhor valor alcançado em quarenta tentativas



Fonte: elaborado pelos autores.

Duas consequências decorrem disso. A primeira é operacional: o orçamento de busca foi fixado em vinte tentativas, iguais para todas, valor que cobre com folga o maior platô observado. A segunda é substantiva: a baixa sensibilidade do erro às configurações de

hiperparâmetros sugere que, dentro do espaço de busca examinado, as arquiteturas dispõem de pouca estrutura preditiva adicional a explorar.

Nenhuma arquitetura supera o passeio aleatório

As dezesseis combinações (quatro arquiteturas em quatro frequências) foram executadas sob protocolo idêntico. A Tabela 3 apresenta o RMSE do passeio aleatório em cada frequência e, para cada arquitetura, a diferença percentual em relação a ele; valores positivos indicam erro maior que o do benchmark.

Tabela 3 – RMSE do passeio aleatório e diferença percentual de cada arquitetura em relação a ele, por frequência

frequência	Random Walk	LSTM	N-BEATS	N-HiTS	TCN
5 min	0,00420856	+0,03	+0,10	+0,03	−0,07
15 min	0,00684163	+0,22	+0,04	+0,01	+0,15
30 min	0,00902807	+0,17	+0,56	+0,22	+1,06
60 min	0,01244192	+0,37	+0,04	+0,11	+0,59

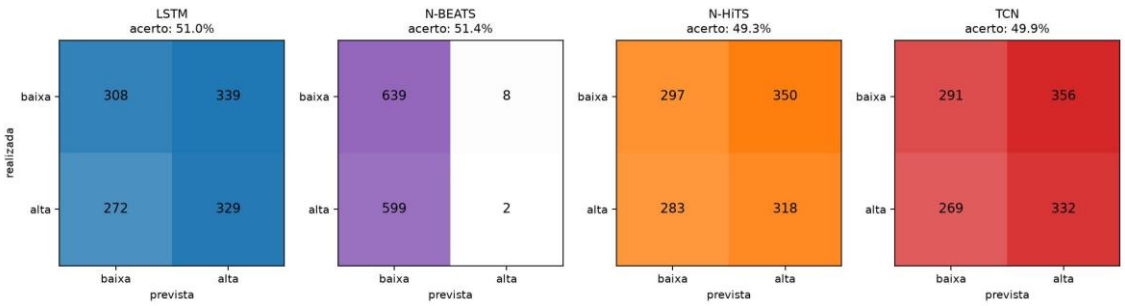
Fonte: elaborado pelos autores.

Em quinze das dezesseis combinações o erro é maior que o do passeio aleatório, e na única em que é menor, a TCN a cinco minutos, a vantagem é de 0,07%. As diferenças, em qualquer direção, situam-se abaixo de 1,1%.

O teste de Diebold-Mariano, ajustado por Bonferroni para as dezesseis comparações, não rejeita a igualdade de acurácia em nenhuma delas. A acurácia direcional média oscila entre 0,4760 e 0,5239, próxima de 0,5 e sem evidência de ganho direcional sistemático. A hipótese de que as redes neurais superariam o passeio aleatório não se confirma.

O Gráfico 2 detalha o desempenho direcional das quatro arquiteturas por meio das respectivas matrizes de confusão, na frequência de quinze minutos. Comparam-se o sentido da variação prevista e o da variação realizada, excluídos os intervalos em que o preço não se alterou, nos quais não há direção realizada a comparar.

Gráfico 2 – Matrizes de confusão direcional das quatro arquiteturas, frequência de quinze minutos, com o percentual de acerto de cada uma



Fonte: elaborado pelos autores.

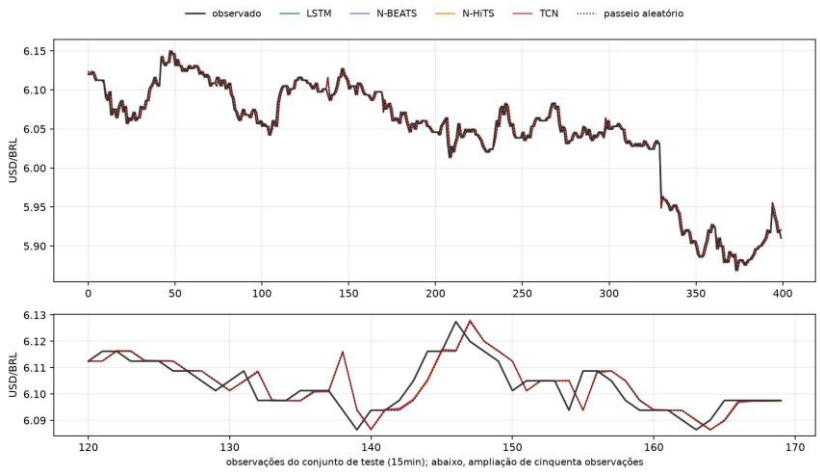
As quatro distribuições são semelhantes entre si e próximas do equilíbrio, o que é coerente com as acurácias direcionais mencionadas anteriormente.

As previsões colapsam para próximo de uma constante

A razão entre a dispersão das previsões e a dispersão dos retornos realizados situa-se entre 0,0072 e 0,1137. Os modelos preveem variações de nove a cento e trinta e nove vezes menores que as efetivamente observadas.

O Gráfico 3 permite verificar a consequência disso. Nele estão a série observada, as previsões das quatro arquiteturas e a previsão do passeio aleatório, na escala de preço, sobre o conjunto de teste. As cinco linhas se sobrepõem visualmente na escala do gráfico, e o painel inferior amplia cinquenta observações para que a coincidência seja verificável de perto.

Gráfico 3 – Valor observado, previsões das quatro arquiteturas e previsão do passeio aleatório na escala de preço, conjunto de teste, frequência de quinze minutos; painel inferior com ampliação de cinquenta observações



Fonte: elaborado pelos autores.

A leitura do gráfico exige um cuidado que ele próprio evidencia. Na escala de preço, a previsão um passo à frente sempre parece acompanhar a série observada, porque o preço em um instante é próximo do preço no instante anterior; qualquer previsão próxima do último valor produz aderência visual. É por isso que a linha do passeio aleatório precisa constar da figura. Sem ela, a aderência das redes sugeriria capacidade preditiva que não se sustenta; com ela, verifica-se que o benchmark faz exatamente o mesmo.

A distância média entre a previsão de cada arquitetura e a do passeio aleatório, expressa como fração da variação efetivamente ocorrida, confirma a leitura: 4,80% para a LSTM, 0,65% para o N-BEATS, 2,90% para o N-HiTS e 4,11% para a TCN. As redes praticamente reproduzem o benchmark.

O comportamento não indica falha de treinamento: é o que a teoria prevê. Sob um martingale, a previsão que minimiza o erro quadrático médio é aproximadamente nula, e um modelo bem ajustado converge para ela. Esse comportamento não indica, por si só, falha de treinamento. Sob a hipótese de martingale, a previsão condicional do retorno tende a zero. Nesse contexto, a baixa amplitude das previsões é compatível com a escassa estrutura preditiva identificada no experimento, sobretudo nas frequências a partir de 15 minutos.

A vantagem do benchmark concentra-se nos intervalos sem movimento

Os testes livres de suposição distributiva revelam um padrão que o Diebold-Mariano não captura. Enquanto este não rejeita a igualdade em nenhuma das dezesseis comparações, o teste de sinais a rejeita em todas as dezesseis, e sempre em favor do passeio aleatório: as arquiteturas erram menos que o benchmark em apenas 29,3% a 40,0% das observações, conforme a combinação de arquitetura e frequência.

A divergência entre os dois testes não é contradição. O teste de sinais conta com que frequência cada previsão é melhor; o Diebold-Mariano pondera pela magnitude da diferença. O passeio aleatório vence com frequência, mas por margens pequenas, e o efeito se dilui na média. A origem dessa vantagem pode ser localizada com precisão, separando as observações conforme o preço tenha ou não se alterado no intervalo:

Tabela 4 – Proporção de observações em que o modelo erra menos que o passeio aleatório, conforme o preço tenha ou não se alterado no intervalo

frequência	**intervalos sem variação**	**modelo erra menos onde o preço parou**	**modelo erra menos onde o preço moveu**
5 min	40,3 %	0,0 %	49,0 % a 52,4 %
15 min	24,7 %	0,0 %	49,3 % a 51,4 %
30 min	21,5 %	0,0 %	49,2 % a 51,0 %
60 min	21,2 %	0,0 %	47,7 % a 50,2 %

Fonte: elaborado pelos autores.

O resultado é categórico. Nos intervalos em que o preço não se altera, o passeio aleatório acerta exatamente, pois sua previsão de variação nula é o valor realizado, e qualquer previsão não nula erra; nenhuma das dezesseis combinações erra menos que o benchmark em uma única dessas observações. Nos intervalos em que o preço efetivamente se move, os modelos erram menos em torno de metade das vezes, sem vantagem sistemática em relação ao benchmark.

Toda a superioridade do passeio aleatório provém, portanto, dos intervalos sem variação de preço, os mesmos que caracterizam a microestrutura discutida anteriormente. Onde há movimento a explicar, nenhuma das arquiteturas explica mais que o acaso.

Diálogo com a literatura

O resultado converge com Rossi et al. (2022) por caminho independente. Aqueles autores, empregando regressão linear e aprendizado de máquina sobre o mesmo mercado e a mesma fonte de dados, só superaram o passeio aleatório com preditores defasados ao incorporar o futuro de real, e ainda assim apenas na frequência de um minuto. A previsibilidade que encontraram exigiu informação originada em outro mercado.

Esta pesquisa examina a questão complementar, a de saber se a série à vista prevê a si mesma, e a resposta é negativa, sob quatro arquiteturas de aprendizado profundo e quatro frequências. As duas conclusões são compatíveis entre si: se, como propõem Rossi et al. (2022), parte relevante do processo de formação de preços ocorre no mercado futuro e é transmitida ao mercado à vista por arbitragem, então a informação relevante pode não estar contida no histórico da própria série à vista.

O contraste com a literatura que relata sucesso de redes neurais em séries financeiras (CHEN; ZHOU; DAI, 2015; SAGHEER; KOTB, 2019) admite leitura semelhante. Boa parte desses estudos trata de ações e nem sempre confronta os modelos com o benchmark ingênuo; Abouzaid e Boussedra (2025) comparam três modelos entre si sem utilizar o passeio aleatório como benchmark. O câmbio, desde Meese e Rogoff (1983), é reconhecidamente o caso em que essa comparação é mais severa, e o resultado aqui obtido é consistente com essa tradição.

Robustez

Toda a caracterização estatística foi replicada sobre a amostra que mantém as sessões sem variação de preço. Os p-valores permanecem estáveis: 0,5370 com as sessões e 0,6112 sem elas a 15 minutos; 0,4938 e 0,5470, respectivamente, a 30 minutos; e 0,3655 e 0,2913, respectivamente, a 60 minutos. A decisão de tratamento quanto a essas sessões não altera as conclusões.

CONSIDERAÇÕES FINAIS

Os objetivos específicos foram integralmente atingidos. Os dados de alta frequência foram coletados e pré-processados, com duas correções de origem (a reconstrução da linha temporal e a inversão da convenção da cotação), documentadas e verificadas. As quatro arquiteturas foram desenvolvidas e treinadas sob protocolo experimental idêntico, As quatro arquiteturas foram desenvolvidas e treinadas sob protocolo experimental idêntico, condição necessária para uma comparação controlada entre elas. O desempenho preditivo foi avaliado por métricas de erro e por testes estatísticos formais. E os resultados foram comparados para identificar o modelo de melhor desempenho.

A hipótese, contudo, não se confirmou. Esperava-se que as redes neurais apresentassem capacidade preditiva superior à do passeio aleatório; em quinze das dezesseis combinações o erro foi maior que o do benchmark, e em nenhuma a diferença é estatisticamente significativa ao nível corrigido. Atingir os objetivos e rejeitar a hipótese não são resultados contraditórios: o trabalho se propôs a testar, e testou.

Três evidências independentes, nenhuma delas dependente da comparação entre modelo e benchmark, sustentam a conclusão. O teste de razão de variâncias não rejeita a hipótese de passeio aleatório a partir de quinze minutos, antes de qualquer rede ser treinada O espaço de busca de hiperparâmetros mostrou-se pouco sensível às configurações examinadas, com ganho máximo de 1,3% em quarenta tentativas, o que sugere pouca estrutura preditiva adicional explorável dentro do espaço de busca considerado. A comparação com o benchmark confirma o que as três já indicavam.

O trabalho identificou ainda dois resultados não previstos. O primeiro é a fronteira da microestrutura: A evidência contra o passeio aleatório é mais forte a 1 minuto e ainda aparece, para os menores valores de q, a 5 minutos; a partir de 15 minutos, ela deixa de ser observada nos q examinados, ao mesmo tempo que os indicadores associados à microestrutura se atenuam. O segundo é a localização precisa da vantagem do benchmark, concentrada nos intervalos sem variação de preço, nos quais nenhuma arquitetura erra menos que ele em uma única observação.

Limitações: A reconstrução da linha temporal é inferência, e não recuperação: a informação de período não existe no arquivo de origem. Cinquenta e cinco sessões sem variação de preço permanecem sem explicação, com regularidade incompatível com comportamento de mercado. Os modelos não observam o movimento ocorrido entre o fechamento de uma sessão e a abertura da seguinte. O estudo de convergência da busca foi medido apenas na frequência de sessenta minutos O período analisado apresenta elevada volatilidade, com amplitude de 30,2%, e as conclusões podem não se estender a períodos mais calmos. Por fim, o estudo cobre um par de moedas, um período e um horizonte de previsão.

Trabalhos futuros: O próprio resultado indica a direção. Se, nas condições analisadas, o histórico da série à vista não apresentou capacidade preditiva adicional, e se a previsibilidade documentada por Rossi et al. (2022) exigiu informação do mercado futuro, o passo natural é incorporar o contrato futuro de dólar como variável preditora e verificar se as arquiteturas aqui avaliadas conseguem explorá-la.

REFERÊNCIAS

ABOUZAID, Oumaima; BOUSSEDRA, Faouzi. Artificial intelligence and exchange rate forecasting: assessing predictive accuracy and macroeconomic sensitivity. **Frontiers in Applied Mathematics and Statistics**, Lausanne, v. 11, 2025. DOI: 10.3389/fams.2025.1654093.

AKIBA, Takuya; SANO, Shotaro; YANASE, Toshihiko; OHTA, Takeru; KOYAMA, Masanori. Optuna: a next-generation hyperparameter optimization framework. In: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. New York: ACM, 2019. p. 2623-2631. DOI: 10.1145/3292500.3330701.

APTE, Manish; HARIBHAKTA, Yashodhara. **Advancing financial forecasting**: a comparative analysis of neural forecasting models N-HiTS and N-BEATS. arXiv:2409.00480, 2024. DOI: 10.48550/arXiv.2409.00480.

ASSOCIAÇÃO BRASILEIRA DAS ENTIDADES DOS MERCADOS FINANCEIRO E DE CAPITAIS. **Feriados nacionais para o ano de 2024**. Disponível em: https://www.anbima.com.br/feriados/fer_nacionais/2024.asp. Acesso em: 17 ago. 2026.

ASSOCIAÇÃO BRASILEIRA DAS ENTIDADES DOS MERCADOS FINANCEIRO E DE CAPITAIS. **Feriados nacionais para o ano de 2025**. Disponível em: https://www.anbima.com.br/feriados/fer_nacionais/2025.asp. Acesso em: 17 ago. 2026.

B3. **Horário de negociação**: derivativos, câmbio e dólar pronto. Disponível em: https://www.b3.com.br/pt_br/solucoes/plataformas/puma-trading-system/para-participantes-e-traders/horario-de-negociacao/derivativos/cambio-e-dolar-pronto/. Acesso em: 17 ago. 2026.

B3. **Ofício Circular 048/2024-PRE**: retificação do calendário de feriados em 2024 e registro de operações na Câmara de Câmbio B3 em 24/12/2024. São Paulo: B3, 2024.

B3. **Ofício Circular 149/2024-PRE**: calendário de feriados em 2025 e funcionamento da B3 em 05/03/2025 (Quarta-Feira de Cinzas). São Paulo: B3, 5 nov. 2024.

BAI, Shaojie; KOLTER, J. Zico; KOLTUN, Vladlen. **An empirical evaluation of generic convolutional and recurrent networks for sequence modeling**. arXiv:1803.01271, 2018. DOI: 10.48550/arXiv.1803.01271.

BANCO CENTRAL DO BRASIL. **Horários de sistemas informatizados**. Disponível em: https://www.bcb.gov.br. Acesso em: 17 ago. 2026.

BRASIL. Lei nº 14.759, de 21 de dezembro de 2023. Declara feriado nacional o Dia Nacional de Zumbi e da Consciência Negra. **Diário Oficial da União**, Brasília, DF, 22 dez. 2023.

CAMPA, José Manuel; GOLDBERG, Linda S. Exchange rate pass-through into import prices: a macro or micro phenomenon? **The Review of Economics and Statistics**, Cambridge, v. 87, n. 4, p. 679-690, 2005. DOI: 10.1162/003465305775098189.

CHALLU, Cristian; ORESHKIN, Boris N.; CARRON, Federico Garza; SEITZER, Max. **N-HiTS**: neural hierarchical interpolation for time series forecasting. arXiv:2201.12886, 2022. DOI: 10.48550/arXiv.2201.12886.

CHEN, Kai; ZHOU, Yi; DAI, Fangyan. A LSTM-based method for stock returns prediction: a case study of China stock market. In: **IEEE International Conference on Big Data**. Santa Clara: IEEE, 2015. p. 2823-2824. DOI: 10.1109/BigData.2015.7364089.

CHEN, Shun; GE, Lei. Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. **Quantitative Finance**, London, v. 19, n. 9, p. 1507-1515, 2019. DOI: 10.1080/14697688.2019.1622287.

CONT, Rama. Empirical properties of asset returns: stylized facts and statistical issues. **Quantitative Finance**, London, v. 1, n. 2, p. 223-236, 2001. DOI: 10.1080/713665670.

DICKEY, David A.; FULLER, Wayne A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, Alexandria, v. 74, n. 366, p. 427-431, 1979. DOI: 10.2307/2286348.

DIEBOLD, Francis X.; MARIANO, Roberto S. Comparing predictive accuracy. **Journal of Business & Economic Statistics**, Alexandria, v. 13, n. 3, p. 253-263, 1995. DOI: 10.1080/07350015.1995.10524599.

ELMAN, Jeffrey L. Finding structure in time. **Cognitive Science**, Hoboken, v. 14, n. 2, p. 179-211, 1990. DOI: 10.1207/s15516709cog1402_1.

ENGLE, Robert F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, New Haven, v. 50, n. 4, p. 987-1007, 1982. DOI: 10.2307/1912773.

GAGLIANONE, Wagner Piazza; MARINS, Jaqueline Terra Moura. **Evaluation of exchange rate point and density forecasts**: an application to Brazil. Working Paper Series, n. 446. Brasília: Banco Central do Brasil, nov. 2016.

GERS, Felix A.; SCHMIDHUBER, Jürgen; CUMMINS, Fred. Learning to forget: continual prediction with LSTM. **Neural Computation**, Cambridge, v. 12, n. 10, p. 2451-2471, 2000. DOI: 10.1162/089976600300015015.

HAMILTON, James D. **Time series analysis**. Princeton: Princeton University Press, 1994.

HARVEY, David; LEYBOURNE, Stephen; NEWBOLD, Paul. Testing the equality of prediction mean squared errors. **International Journal of Forecasting**, Amsterdam, v. 13, n. 2, p. 281-291, 1997. DOI: 10.1016/S0169-2070(96)00719-4.

HERZEN, Julien *et al.* Darts: user-friendly modern machine learning for time series. **Journal of Machine Learning Research**, [S. l.], v. 23, n. 124, p. 1-6, 2022.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. **Neural Computation**, Cambridge, v. 9, n. 8, p. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

JARQUE, Carlos M.; BERA, Anil K. A test for normality of observations and regression residuals. **International Statistical Review**, Haia, v. 55, n. 2, p. 163-172, 1987. DOI: 10.2307/1403192.

KIM, Misuk; PARK, Eun Lee; CHO, Sungzoon. Stock price prediction through sentiment analysis of corporate disclosures using distributed representation. **Intelligent Data Analysis**, Amsterdam, v. 22, n. 6, p. 1395-1413, 2018. DOI: 10.3233/IDA-173670.

KRUGMAN, Paul R.; OBSTFELD, Maurice. **Economia internacional**: teoria e política. 10. ed. São Paulo: Pearson, 2018.

KWIATKOWSKI, Denis; PHILLIPS, Peter C. B.; SCHMIDT, Peter; SHIN, Yongcheol. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? **Journal of Econometrics**, Amsterdam, v. 54, n. 1-3, p. 159-178, 1992. DOI: 10.1016/0304-4076(92)90104-Y.

LEVENE, Howard. Robust tests for equality of variances. In: OLKIN, Ingram (org.). **Contributions to probability and statistics**: essays in honor of Harold Hotelling. Stanford: Stanford University Press, 1960. p. 278-292.

LJUNG, Greta M.; BOX, George E. P. On a measure of lack of fit in time series models. **Biometrika**, Oxford, v. 65, n. 2, p. 297-303, 1978. DOI: 10.1093/biomet/65.2.297.

LO, Andrew W.; MACKINLAY, A. Craig. Stock market prices do not follow random walks: evidence from a simple specification test. **The Review of Financial Studies**, Oxford, v. 1, n. 1, p. 41-66, 1988. DOI: 10.1093/rfs/1.1.41.

MEESE, Richard A.; ROGOFF, Kenneth. Empirical exchange rate models of the seventies: do they fit out of sample? **Journal of International Economics**, Amsterdam, v. 14, n. 1-2, p. 3-24, 1983. DOI: 10.1016/0022-1996(83)90017-X.

MINCER, Jacob; ZARNOWITZ, Victor. The evaluation of economic forecasts. In: MINCER, Jacob (org.). **Economic forecasts and expectations**: analysis of forecasting behavior and performance. New York: National Bureau of Economic Research, 1969. p. 3-46.

ORESHKIN, Boris N.; CARPOV, Dmitri; CHAPADOS, Nicolas; BENGIO, Yoshua. **N-BEATS**: neural basis expansion analysis for time series forecasting. arXiv:1905.10437, 2020. DOI: 10.48550/arXiv.1905.10437.

PATEL, Jigar; SHAH, Sahil; THAKKAR, Priyank; KOTECHA, K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. **Expert Systems with Applications**, Amsterdam, v. 42, n. 1, p. 259-268, 2015. DOI: 10.1016/j.eswa.2014.07.040.

ROLL, Richard. A simple implicit measure of the effective bid-ask spread in an efficient market. **The Journal of Finance**, Hoboken, v. 39, n. 4, p. 1127-1139, 1984. DOI: 10.1111/j.1540-6261.1984.tb03897.x.

ROSSI, José Luiz; PICCIONI, Carlos; ROSSI, Marina; CAJUEIRO, Daniel. **Brazilian exchange rate forecasting in high frequency**. Nota Técnica IDB-TN-2561. Washington: Banco Interamericano de Desenvolvimento, set. 2022. DOI: 10.18235/0004488.

SAGHEER, Alaa; KOTB, Mostafa. Time series forecasting of petroleum production using deep LSTM recurrent networks. **Neurocomputing**, Amsterdam, v. 323, p. 203-213, 2019. DOI: 10.1016/j.neucom.2018.09.082.

Contatos: carolinepenna13@hotmail.com/ hadad.junior@gmail.com