

idp

idp

MESTRADO PROFISSIONAL EM ECONOMIA

**A INFLUÊNCIA DE VARIÁVEIS PREDITORAS NA PERFORMANCE
DE REDES NEURAIS RECORRENTES PARA ESTIMATIVA DO
PREÇO DE AÇÕES INTEGRANTES DO IBOVESPA**

PEDRO ALEXANDRE SANTANA REIS

Brasília-DF, 2023

PEDRO ALEXANDRE SANTANA REIS

A INFLUÊNCIA DE VARIÁVEIS PREDITORAS NA PERFORMANCE DE REDES NEURAIS RECORRENTES PARA ESTIMATIVA DO PREÇO DE AÇÕES INTEGRANTES DO IBOVESPA

Dissertação apresentada como requisito para obtenção do título de Mestre em Economia pelo Programa de Mestrado Profissional em Economia, Políticas Públicas e Desenvolvimento do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP).

Orientador

Professor Doutor Mathias Schneider Tessmann

Brasília-DF 2023

PEDRO ALEXANDRE SANTANA REIS

A INFLUÊNCIA DE VARIÁVEIS PREDITORAS NA PERFORMANCE DE REDES NEURAIS RECORRENTES PARA ESTIMATIVA DO PREÇO DE AÇÕES INTEGRANTES DO IBOVESPA

Dissertação apresentada como requisito para obtenção do título de Mestre em Economia pelo Programa de Mestrado Profissional em Economia, Políticas Públicas e Desenvolvimento do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP).

Aprovado em 17 / 01 / 2023

Banca Examinadora

Prof. Dr. Mathias Schneid Tessmann - Orientador

Prof. Dr. Alexandre Xavier Ywata de Carvalho

Prof. Dr. João Gabriel de Moraes Souza

R375i Reis, Pedro Alexandre Santana
A influência de variáveis preditoras na performance de redes neurais recorrentes para estimativa do preço de ações integrantes do IBOVESPA / Pedro Alexandre Santana Reis. – Brasília: IDP, 2024.

48 p.
Inclui bibliografia.

Trabalho de Conclusão de Curso (Dissertação) – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa – IDP, Curso de Mestrado Profissional em Economia, Brasília, 2023.
Orientador: Prof. Dr. Mathias Schneider Tessmann.

1. RNN. 2. LSTM. 3. Preço de ações. 4. MAPE. 5. Valores Shapley. I. Título.

CDD: 330

Ficha catalográfica elaborada pela Biblioteca Ministro Moreira Alves
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa



RESUMO

Este trabalho busca compreender a influência de variáveis preditivas nas estimativas de redes neurais recorrentes para previsão do preço de ações negociadas na B3 no ano de 2022. Considerando cinco ações que compõem o índice Bovespa e dados de treinamento de 2012 a 2021, examina-se a efetividade das variáveis preditoras e de suas defasagens no erro percentual absoluto médio. São utilizadas defasagens de 5, 15, 30, 60 e 90 pregões, como também combinações de preço de abertura, máxima, mínima, fechamento, volume e dos indicadores técnicos bandas de Bollinger e o índice de força relativa (RSI). O horizonte preditivo se estende de um a cinco pregões imediatamente posteriores ao pregão mais recentemente observado. A contribuição global das variáveis e das suas defasagens para as estimativas produzidas pelos modelos é descrita pela utilização de valores de Shapley. Os resultados obtidos revelam que modelos mais simples que utilizam somente a série histórica do preço de fechamento tendem a produzir os melhores resultados. Isso não obstante, a utilidade da inclusão de mais defasagens parece depender das características de cada ativo considerado. Por fim, há evidência de que indicadores da análise técnica podem ser úteis para melhorar as estimativas dos modelos.

Palavras-chaves: RNN, LSTM, preço de ações, MAPE, valores Shapley.

ABSTRACT

This work aims to describe the influence of the predictive variables in the estimates produced by recurrent neural networks with respect to the prediction of stock prices that were negotiated in the B3 stock exchange in 2022. Taking into account 5 stocks listed in the composition of the Bovespa index, the effectiveness of the predicting variables and its lags is evaluated with respect to the mean absolute percentage error. Lags of 5, 15, 30, 60 and 90 trading days were used, along with different mixes of the following variables: open, high, low and close prices, volume, Bollinger Bands and relative strength index (RSI). The time horizon of the predictions ranges from 1 to 5 time steps. The overall contribution of each variable and its lags for the outcomes predicted by the model is described with the use of Shapley values. The results suggest that simpler models tend to produce the best results with respect to all the stocks considered. The usefulness of larger lags, in turn, seems to be conditioned to the characteristics of each asset. Furthermore, there is evidence that some technical indicators may contribute significantly to the estimates produced by the models.

Keywords: RNN, LSTM, stock price, MAPE, Shapley values.

LISTA DE ABREVIATURAS E SIGLAS

ARIMA	Autoregressive, Integrated Moving Average (ARIMA)
BB	Bolinger Band
BPTT	Backpropagation Through Time
BSE	Bombay Stock Exchange
CNN	Convolutional Neural Network
CVM	Comissão de Valores Mobiliários
GRU	Gated Recurrent Unit
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
NASDAQ	National Association of Securities Dealers Automated Quotations
OHLC	Open, High, Low and Close
NSE	National Stock Exchange of India
NYSE	New York Stock Exchange
RELU	Rectified Linear Unit
RNA	Rede Neural Artificial
RNN	Recurrent Neural Network
RSI	Relative Strength Index
RSPE	Root Mean Squared Percentage Error

LISTA DE ILUSTRAÇÕES

Figura 1 Arquitetura LSTM	24
Figura 2 Distribuição MAPE por horizonte preditivo	29
Figura 3 Distribuição do MAPE em função da defasagem das variáveis preditoras	30
Figura 4 Distribuição do MAPE em função da quantidade de variáveis preditoras	30
Figura 5 Distribuição do MAPE em função da defasagem, segmentando-se por horizonte preditivo e quantidade de variáveis preditoras	32
Figura 6 Distribuição do MAPE por ação e modelo utilizado	33
Figura 7 Contribuição global das defasagens nos melhores modelos de cada ação – percentual de valores Shapley absolutos	35
Figura 8 Contribuição global de valores Shapley absolutos para variáveis preditoras e suas defasagens relativamente ao modelo ITUB4 T15 D8	36
Figura 9 Distribuição dos resíduos por modelo e por horizonte preditivo	36
Figura 10 Função de autocorrelação dos resíduos	38
Figura 11 Função de autocorrelação parcial dos resíduos	39

LISTA DE TABELAS

Tabela 1

Mediana do MAPE por ação e horizonte preditivo**29**

Tabela 2

Mediana do MAPE por ação e defasagem das variáveis preditoras**30**

Tabela 3

Mediana do MAPE por ação e quantidade de variáveis preditoras**31**

Tabela 4

MAPE por ação, defasagem e quantidade de variáveis preditoras**33**

Tabela 5

Estatística e p-valor do teste Shapiro-Wilk dos resíduos por horizonte preditivo**37**

SUMÁRIO

1 INTRODUÇÃO	13
2 FUNDAMENTAÇÃO TEÓRICA.....	17
3 METODOLOGIA.....	22
3.1 Dados.....	22
3.2 Da arquitetura LSTM.....	23
3.3 Dos modelos.....	26
3.4 Dos valores Shapley.....	27
4 RESULTADOS.....	30
5 CONCLUSÃO	43
REFERÊNCIAS.....	45



1

INTRODUÇÃO

A obtenção de melhores estimativas do preço ou da variação do preço de um ativo pode ser útil para otimizar a alocação de recursos tanto na perspectiva de decisões de investimento como na relativa à gestão do risco por parte dos investidores. Nesse sentido, cumpre destacar que o aumento da quantidade de dados disponíveis e o aumento da capacidade computacional têm tido importantes efeitos na modelagem financeira (Dixon, Halperin e Bilokon, 2020).

O mercado de ações brasileiro contava em 2022 com aproximadamente 450 companhias de capital aberto negociando valores mobiliários do tipo ações ordinárias/preferenciais. As cotações desses ativos podem ser prontamente obtidas na internet com relação a vários períodos. Caso se considere uma empresa que em 2012 tenha tido 250 cotações diárias de preços de abertura, máxima, mínima, fechamento e volume, então somente para esse período já haveria aproximadamente 560 mil observações. Multiplique-se esse valor por 10 anos e já se chega a uma quantidade considerável de dados.

Aliado ao aumento da disponibilidade de dados, existe o aumento da capacidade computacional que tem viabilizado a implementação de modelos de inteligência artificial, notadamente de redes neurais artificiais (RNA). Emerge, assim, a oportunidade de mineração de consideráveis massas de dados no intuito de produzir informação útil à tomada de decisão. Alguns métodos tradicionais de produção de estimativas estão ancorados em pressupostos como a estacionariedade da série a ser modelada e a relação linear entre variáveis preditoras e variável resposta (Neusser, 2016). Ocorre que, na prática, nem sempre esse pressuposto se observa, sendo razoável supor que a relação entre as variáveis de interesse exiba relação não-linear. Assim, emerge a necessidade de considerar modelos que não dependam de tantos pressupostos e que sejam ao mesmo tempo dirigidos a dados. Esse é o caso das redes neurais artificiais.

Em especial, redes neurais recorrentes (RNN) alcançaram proeminência em aplicações de linguagem natural (*Natural Language Processing - NLP*) como a classificação de textos (Minaee et al., 2021). A escolha de redes neurais recorrentes se justifica na medida em que sua arquitetura leva em consideração de modo inerente a ordem de

apresentação dos dados. Sua aplicação alcançou reconhecimento em aplicações de linguagem natural e permitiu a elaboração de arquiteturas como a *many-to-one* e a *many-to-many*. Naquela, uma sequência é utilizada para a produção de um output; nesta, uma sequência é insumo para a produção de outra sequência.

Considerando que informações sobre contexto e ordem de apresentação dos dados é importante em aplicações de linguagem natural, vislumbrou-se também sua aplicabilidade a dados de séries temporais (Gers, Eck e Schmidhuber, 2001), no que se incluem naturalmente dados relativos a preços de ações.

Ainda que modelos de inteligência artificial, notadamente redes neurais, tenham alcançado novas marcas na qualidade preditiva, remanesce para o usuário a importância de saber quando a previsão produzida pelo algoritmo é confiável, como também compreender as razões que levaram a determinada estimativa. Essa a razão para a utilização de valores Shapley.

Diante disso, o presente trabalho busca compreender a influência da utilização de mais variáveis preditoras e os aumentos das suas defasagens na qualidade preditiva de redes neurais artificiais do tipo LSTM. Adicionalmente, busca-se compreender o quanto, de modo global, cada variável e sua defasagem contribuíram para a formação das estimativas produzidas pela rede neural artificial.

Um dos motivos para avançar nessa direção consiste em que há muitas abordagens aplicáveis à otimização de modelos de RNN que se dirigem a aspectos como quantidade de camadas ocultas, quantidade de neurônios por camada, funções de ativação, otimizadores, taxa de aprendizagem, quantidade de épocas de treinamento, tamanho do lote de treinamento e regularização. Ocorre que os dados utilizados podem ter impacto considerável na qualidade das predições. Pode ser útil, portanto, colher evidências de como os dados utilizados no treinamento impactam a performance do modelo, em especial no que diz respeito à utilização de mais variáveis preditoras e do aumento da sua defasagem.

Outro motivo para o presente trabalho consiste em utilizar RNN como uma alternativa a modelos tradicionais de previsão em séries temporais integrantes da abordagem ARIMA. Ainda que essa abordagem seja clássica e testada pelo tempo, ela impõe várias restrições que podem significar o descarte de informação relevante para o cumprimento da tarefa preditiva tão somente porque seus

pressupostos não foram atendidos. Dentre esses pressupostos podem-se mencionar a estacionariedade da série, a eventual necessidade de diferenciação, a quantidade de variáveis preditoras, sua defasagem e o horizonte preditivo. As RNN podem tratar esses aspectos de uma forma mais transparente, na medida em que sua aplicabilidade não está de antemão condicionada a pressupostos como estacionariedade, diferenciação da série e remoção de sazonalidade.

Além disso, motiva o trabalho a análise técnica ou grafista. Com efeito, ela tem por pressuposto a suficiência do valor histórico da ação (OHLC e volume) para poder fazer inferência quanto à sua tendência de variação futura. Valores históricos também são utilizados para elaborar indicadores, o que pode ser compreendido como uma tarefa de *feature engineering* ou engenharia de características. Assim, a análise técnica tem em sua essência uma tarefa de reconhecimento de padrões úteis para a previsão de preços, sendo, portanto, o campo da RNN, na medida em que o aprendizado consiste na identificação de padrões que levem à otimização das estimativas produzidas pela rede neural.

A abordagem grafista desafia, ainda, a hipótese do mercado eficiente (HME), segundo a qual não haveria espaço para arbitragem em um espaço em que toda a informação relevante estaria disponível para os tomadores de decisão. (“Efficient Markets Hypothesis”, [s.d.]

Tecidas essas considerações, e para alcançar o objetivo proposto, este estudo utiliza redes neurais artificiais na previsão do valor de ações negociadas na bolsa de valores do Brasil. Trata-se, mais especificamente, de utilizar redes neurais artificiais recorrentes do tipo Long Short-Term Memory (LSTM) para efetuar a previsão do preço de ações integrantes do IBOVESPA no ano de 2022. Sublinhe-se que se trata de um período pós-pandemia em que a volatilidade dos preços ainda pode estar elevada, aumentando, assim, o desafio da produção de estimativas mais acuradas.

O trabalho encontra-se organizado da seguinte forma. A seção 2 traz a fundamentação teórica, no que se incluem estudos que se debruçaram sobre aspectos tangenciados nesta pesquisa. A seção 3 contém a descrição do método, no que incluem dados, configuração da RNA e valores Shapley. A seção 4 apresenta resultados obtidos. A seção 5 consiste na conclusão, em que se destacam algumas restrições do presente estudo e são sinalizadas possíveis linhas de desenvolvimento futuro.



?

2

FUNDAMENTAÇÃO TEÓRICA

A aplicação de redes neurais recorrentes para a previsão do preço de ações já foi objeto de estudo de algumas pesquisas. Bathla, Rani e Aggarwal (2022) analisaram a capacidade de redes neurais LSTM preverem a variação de preços de ações entre janeiro de junho de 2020. Os dados de treino se estenderam de janeiro de 2010 a janeiro de 2020. Foi utilizada a previsão de vários índices (NSE, BSE, NASDAQ, NYSE, Dow Jones e Nikkei) e uma janela de 60 dias. Foram utilizadas 5 camadas ocultas e os dados foram escalonados entre 0 e 1. O treinamento utilizou 100 épocas e houve a utilização de *dropout*. A eficácia do modelo foi aquilatada utilizando-se o erro médio absoluto percentual (MAPE), alcançando valores entre 0,86% e 3,89%. Concluíram, assim, que modelos de *deep learning* podem prever o preço de ações de maneira efetiva mesmo em cenário de elevada volatilidade.

Pinheiro (2022) examinou modelos de redes neurais recorrentes de aprendizado profundo univariado e multivariado do tipo RNN e LSTM relativamente às ações PETR4, ITUB4 e ABEV3. Buscou avaliar se as ações selecionadas seguem um processo estocástico *random walk* i.e. se era válida a hipótese dos mercados eficientes em sua forma fraca. Também buscou aplicar técnica de validação cruzada para séries temporais no intuito de avaliar o poder preditivo de modelos RNN e LSTM univariados e multivariados na previsão do preço de ações. Concluiu que os modelos RNN LSTM univariado e multivariado reportaram maior poder preditivo do que os modelos RNN-U e RNN-M.

Bhandari et al. (2022) utilizaram LSTM para prever o preço de fechamento do dia seguinte do índice S&P 500. Foram utilizadas nove variáveis preditoras relativas a dados fundamentais do mercado, dados macroeconômicos e indicadores técnicos. O desempenho do modelo foi avaliado com a utilização do Root Mean Square Error (RMSE), MAPE e coeficiente de correlação. Concluiu-se que modelo com uma única camada de LSTM ofereceu um melhor ajuste e acurácia do que modelos com várias camadas.

Budiharto (2021) utilizou modelo LSTM para prever preços na bolsa de valores da Indonésia durante período sob influência da crise sanitária de 2020. Observou-se uma melhor adequação do modelo à

medida em que a quantidade de épocas de treinamento aumentou, ainda que dados de mais curta duração tenham sido utilizados.

Chandola et al. (2021) utilizaram modelo de rede neural LSTM com processamento de linguagem natural para prever a direção do movimento de preços no mercado de ações. Foram considerados papéis negociados por 5 (cinco) empresas de diferentes setores de atuação. O insumo do modelo consistiu no preço de fechamento da série temporal e em títulos de notícias (*news headlines*). Foram consideradas séries de Apple, PepsiCo, NRG, AT&T e APEI, estendendo-se de agosto/2008 a julho/2016. Utilizou-se a variação percentual do ativo como variável preditora. A LSTM foi utilizada para fazer o processamento de linguagem natural. A acurácia obtida gravitou entre 51% (APEI) e 54,9% (AT&T) nos dados de validação.

Kaushik et al. (2020) aplicaram *deep neural networks* para a tarefa preditiva em séries temporais não estacionárias. No estudo, foram consideradas redes neurais formada por perceptrons em múltiplas camadas, redes convolucionais (CNN), GRU-RNN e LSTM-RNN. Foram considerados 10 ativos financeiros da Índia. A avaliação da performance dos modelos considerou a previsão de um e vários passos à frente em vários períodos. Na previsão *single-step*, foram testadas janelas temporais (variáveis preditoras) de 3, 5, 7, 9, 11 13 e 15 dias; na previsão *multi-step*, foram utilizadas janelas de 30, 60 e 90 dias.

A previsão de vários períodos à frente foi feita de modo direto, ou seja, cada predição considerou exclusivamente dados constantes na janela de entrada. Os dados foram normalizados no intervalo zero a um. Os dados de treino se estenderam de 2002 a 2017. A arquitetura LSTM contemplou 2 camadas ocultas com 256 e 128 neurônios. A métrica utilizada foi *Mean-Squared-Error* (MSE). Observou-se uma degradação das previsões conforme o horizonte preditivo se estendia no tempo. Uma vez que se utilizou o preço puro de ações, concluiu-se que a utilização de outras variáveis como fatores micro e macroeconômicos pode ser útil para aperfeiçoar a performance da utilização de redes neurais em tarefa de previsão *multi-step*.

Thu, Thanh e Le Quy (2022) utilizaram modelo baseado na arquitetura LSTM para prever o preço de ações. Sua abordagem utilizou ações semelhantes entre si como variáveis preditoras, além de valer-se da aplicação de uma janela dinâmica deslizante na fase de predição baseada na análise de alteração significativa no valor do ativo.

Lawi, Mesra e Amir (2022) propuseram oito modelos de aprendizagem profunda do tipo LSTM e GRU para classificar o movimento do preço de ações em três grupos. Nesse estudo, a acurácia dos modelos foi avaliada pela utilização do Mean Absolute Percentage Error (MAPE), Root Mean Squared Percentage Error (RMSPE) e do Rooted Mean Dimensional Percentage Error (RMDPE). Seus resultados mostraram que, na tarefa proposta, ambas as arquiteturas seriam acuradas, com modelos lastreados na arquitetura GRU exibindo a melhor acurácia nos dados de teste. Isso não obstante, modelos LSTM se revelaram mais consistentes no que diz respeito a convergência.

Kumar e Haider (2021) trataram da otimização de hiperparâmetros de modelos RNN-LSTM aplicados à previsão do preço de ações com base em dados intraday. Eles combinaram RNN-LSTM com o algoritmo de polinização de flores e RNN-LSTM com o método de otimização *particle swarm*. Os autores relataram que obtiveram uma redução da taxa de erro e uma melhoria na acurácia.

Ding e Qin (2020) propuseram uma associação de redes neurais com camadas LSTM de maneira a prever, simultânea e respectivamente, os preços de abertura, máximo e mínimo de ações, de maneira que todas elas recebam os dados de entrada, mas a segunda recebia além disso a previsão da primeira e a terceira as previsões da primeira e da segunda. Os resultados descritos sinalizam a obtenção de melhores resultados com essa abordagem do que com redes neurais LSTM que não se valeram dessa associação.

Hargreaves e Chen (2020) utilizaram LSTM para prever o movimento do preço de ações na Australian Stock Market e para identificar quais comprar com vistas a compor um *portfolio* rentável. Os resultados mostraram que a RNN LSTM superou as abordagens baseadas em árvore de decisão (CART) e ARIMA. Enfatiza a habilidade da RNN LSTM em detectar relações não lineares nos dados.

Firouzjaee e Khaliliyan (2023) buscaram compreender o impacto de *features* correlacionadas na interpretabilidade de modelos LSTM aplicados à previsão do preço de ações de companhias de petróleo. Os resultados, contudo, demonstraram que adicionar mais variáveis não melhoram sua interpretabilidade, sublinhando, ainda, a importância de cautela ao confiar exclusivamente no modelo para fazer previsões, vez que não seria fácil identificar os fatores-chave que estariam influenciando os movimentos do preço do ativo.

Nota-se, por conseguinte, que vários esforços têm sido empreendidos para aperfeiçoar a utilização de modelos de RNA lastreados na arquitetura LSTM, seja no aspecto da qualidade das estimativas, seja no da sua interpretabilidade. Nesse sentido, cabe mencionar ainda as contribuições de Vohra e Tanna (2022, 2023), Guo (2023), Kadadekar *et al.* (2023), Alasiri e Qahmash (2023), Dhokane e Agarwal (2023), Mengqin e Ye (2023), Zhang, Sjarif e Ibrahim (2023), Murthy *et al.* (2023), Abiodun e Omoware (2023), Lv e Cui (2023) e Chen (2023).

Assim, o presente trabalho busca contribuir com esta literatura ao trazer evidências dos reflexos da escolha de variáveis preditoras comumente utilizadas na análise técnica e da sua defasagem, na qualidade das previsões produzidas por redes neurais artificiais do tipo LSTM quando aplicadas ao mercado de ações brasileiro em um período pós-pandemia.



3

3

METODOLOGIA

3.1 Dados

Os dados utilizados consistem na cotação diária de ações que compõe o IBOVESPA relativas a Vale S.A. (VALE3), Petróleo Brasileiro S.A. (PETR3 e PETR4), Itaú Unibanco Holding S.A. (ITUB4) e Banco Bradesco S.A. (BBDC4), estendendo-se de 2012 a 2022. Esses papéis estão entre os mais importantes na composição do IBOVESPA, com participação percentual de 15%, 4,3%, 7,3%, 6,8% e 3,8%, aproximada e respectivamente¹.

As variáveis utilizadas consistem nos preços de abertura, máxima, mínima e de fechamento, como também no volume de ações negociadas. No que tange a indicadores técnicos, são utilizadas bandas de Bollinger e o indicador de força relativa (*Relative Strength Indicator* ou RSI). A variável resposta foi o preço de fechamento.

As observações são divididas em dados de treinamento/validação (2012 a 2021) e testes (2022). Não se utiliza a divisão estanque do *dataset* em três, visto que a partição de validação recairia no ano de 2021 que pode ser considerado um período atípico, mas fundamental para o aprendizado do modelo a ser utilizado no ano seguinte. Isso não obstante, durante o treinamento, reserva-se, aleatoriamente, em cada época, aproximadamente 12% dos dados para compor uma partição de validação. A performance do modelo, por sua vez, é avaliada uma única vez, após a conclusão do treinamento, utilizando-se o *dataset* de teste (2022).

Os dados são escalonados dividindo-se cada métrica pelo maior valor observado na partição de treinamento. O menor valor observado em cada métrica não é utilizado para a normalização dos dados. Uma razão para isso é que não há preço negativo de ação, mas é possível em algum momento observem-se cotações menores do que o valor mínimo utilizado para escalonar os dados.

Além disso, uma vez que a métrica utilizada é MAPE, sua interpretação pode ser mais direta quando aplicável a dados não

¹ Composição em 4 de dezembro de 2023.

escalonados. Ou seja, caso os valores observados e as previsões do modelo sejam multiplicados pelo valor máximo utilizado no escalonamento, o valor do MAPE não se altera. Isso, contudo, não aconteceria com dados normalizados entre zero e um.

3.2 Da arquitetura LSTM

Na linha das redes neurais recorrentes, a arquitetura LSTM tem o condão de processar os dados de entrada de modo sequencial. Ou seja, os dados não são processados ao mesmo tempo, mas apresentados de modo sequencial à camada recorrente. Para cumprir essa tarefa, a camada neural recorrente possui um estado oculto (*hidden state*) que é atualizado à medida em que se avança na leitura sequencial do *input*. Em cada passo, a camada recebe a próxima observação e o *hidden state* anterior, que são utilizados para atualizar esse mesmo *hidden state*. O *output* da camada é produzido ou a cada passo em que se lê um *input* ou ao final após a conclusão da leitura de toda a sequência. Neste caso, o *hidden state* final é que é utilizado para produzir o *output* da camada.

Matematicamente, tem-se que

$$h_t = X_t W_{th} + h_{t-1} W_{hh} + bias \quad (1)$$

$$o_t = f(h_t) \text{ ou } o = f(h_T) \quad (2)$$

em que X_t é um vetor de *features* no tempo t , h_{t-1} é o estado interno anterior da camada, e W_{th} e W_{hh} são matrizes que multiplicam esses vetores atualizando, assim, o *hidden state*.

O que as equações acima querem mostrar é que uma camada do tipo recorrente atualiza constantemente seu *hidden state* (h) em função do *input* atual (X_t) e do seu *hidden state* anterior (h_{t-1}) e que seu *output* é produzido em função desse *hidden state*, seja sequencialmente, seja ao final, com espeque no derradeiro estado interno (h_T).

Um dos benefícios dessa abordagem é que ela economiza parâmetros a serem estimados no tratamento de sequências de diferentes extensões. À guisa de ilustração, em uma camada densa, a quantidade de parâmetros é uma função linear da extensão da defasagem das variáveis preditoras e da sua dimensionalidade,

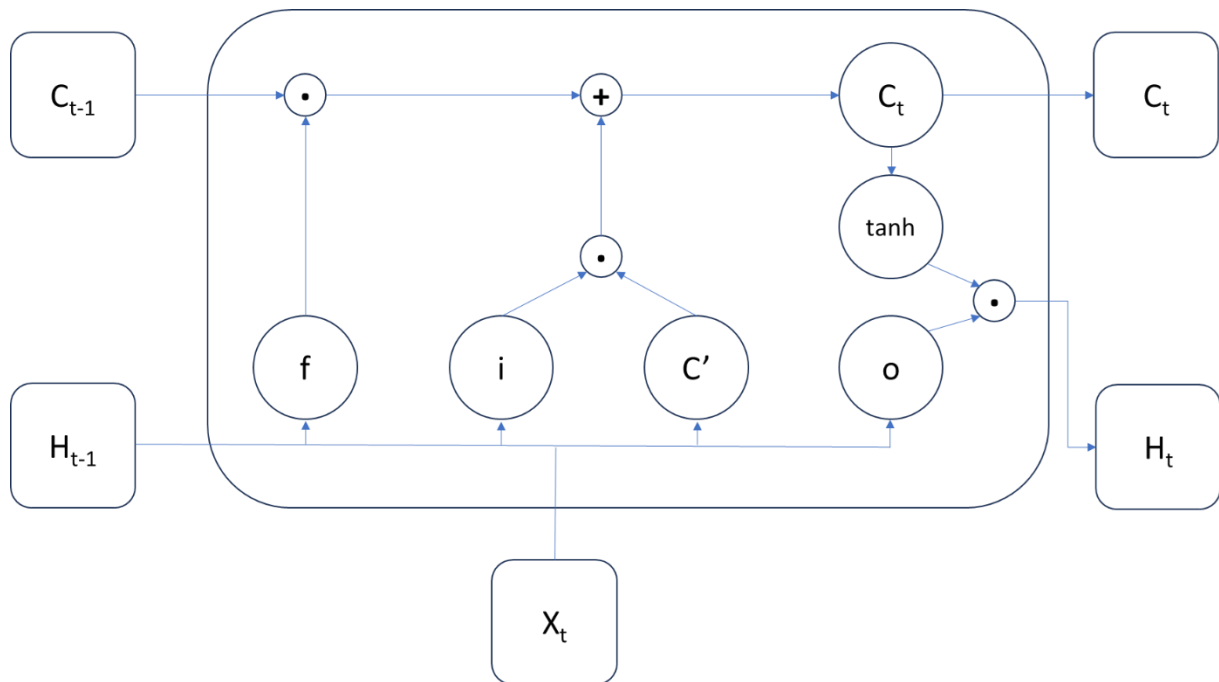
enquanto, em se tratando de redes neurais recorrentes, o aumento da defasagem não impacta na quantidade de parâmetros, visto que os dados defasados são lidos sequencialmente. Ou seja, $N_{dense} \sim T \cdot D$, mas $N_{RNN} \sim D$.

Todavia, o fato de a estimação dos parâmetros depender da apresentação ordenada dos dados faz com que essa abordagem não seja muito beneficiada pela possibilidade de paralelismo. Outro desafio associado ao treinamento de RNN consiste na implementação do algoritmo de correção dos pesos treináveis do modelo (*Backpropagation Through Time* ou BPTT). É que à medida em que a sequência se torna mais extensa, os sinais de erros obtidos no cotejo do *output* produzido e o valor esperado podem tender a desaparecer ao longo do seu iter até as primeiras camadas da rede neural (*vanishing gradient*) ou, eventualmente, aumentar de modo explosivo (*exploding gradient*).

Além disso, persiste o desafio da rede neural aprender relações de dependência entre itens mais afastados dentro da sequência de *input*, o que se torna mais evidente em se tratando da utilização de RNN no processamento de linguagem natural e a identificação de relações de dependência entre termos de uma mesma frase ou oração.

Com o intuito de obviar esses desafios, a LSTM apresenta algumas diferenças em relação a um modelo recorrente simples (*vanilla RNN*), notadamente no que tange à introdução dos conceitos de célula de memória e *gates*. A LSTM vai apresentar uma célula de memória denominada C que é atualizada a cada *time step*. Essa célula vai ser utilizada para a produção do *output* e para a atualização do *hidden state*. A Figura 1 descreve seus principais componentes.

Figura 1: Arquitetura LSTM



Fonte: Elaborada pelo autor.

Os componentes da Figura 1 podem ser úteis para descrever a arquitetura de uma camada LSTM e destacar parâmetros objeto da estimação. Mencionem-se, desde logo, a existência de 3 componentes denominados *forget gate*, *input gate* e *output gate*, bem como a existência de célula de estado (*state cell*).

Em cada *time step*, a camada recebe o vetor de variáveis preditoras no tempo t e o *hidden state* anterior ($t-1$). Os portões *forget gate*, *input gate* e *output gate* recebem como *input* o resultado da combinação linear desses tensores (mais um termo de *bias*) e lhes aplica uma função de sigmoide, de maneira que os valores resultantes podem variar entre 0 e 1. O *forget gate* é aplicado *component wise* à célula de estado, de maneira a decidir quais informações serão esquecidas (ou visto de outra maneira, preservadas).

O componente C' é o resultado da aplicação da função tangente hiperbólica aos *inputs* recebidos e contém novos valores que vão atualizar a célula de estado após sua multiplicação elemento a elemento pelo *input gate*. Com efeito, a soma, de C_{t-1} com C' resulta em C_t . O novo valor da célula de estado passa, então, por uma transformação não linear (tangente hiperbólica) *element wise*, dando origem a um vetor que passa, também *element wise*, por uma multiplicação pelo resultado do *output gate*, decidindo, assim, quanto da informação constante na nova célula de estado passa para h_t .

As operações podem ser assim resumidas:

$$g_{forget} = \sigma(W_{fx}X_t + W_{fh}h_{t-1} + b_f) \quad (3)$$

$$g_{input} = \sigma(W_{ix}X_t + W_{ih}h_{t-1} + b_i) \quad (4)$$

$$C' = \tanh(W_{tx}X_t + W_{th}h_{t-1} + b_t) \quad (5)$$

$$g_{output} = \sigma(W_{ox}X_t + W_{oh}h_{t-1} + b_o) \quad (6)$$

Sendo que a cada *time step*:

$$C_t = g_{forget} \odot C_{t-1} + g_{input} \odot C'_t \quad (7)$$

$$h_t = g_{output} \odot \tanh(C_t) \quad (8)$$

onde σ é a função sigmoide, W , uma matriz, b , um vetor, \tanh é a função tangente hiperbólica e \odot significa multiplicação elemento a elemento (*element wise*).

Dada esta configuração, os parâmetros treináveis em uma camada LSTM são as matrizes W_{fx} , W_{fh} , W_{ix} , W_{ih} , W_{tx} , W_{th} , W_{ox} , W_{oh} e os vetores b_f , b_i , b_t , e b_o . Assim, e.g., quando o *input* de uma camada LSTM com 40 células (dimensão do *hidden state*) tem dimensões 15x5 (i.e. 15 *time steps* de 5 variáveis), a camada LSTM terá 7360 parâmetros treináveis $[(40 * 5 + 40 * 40 + 40) * 4]$ parâmetros para treinar. Mister observar que a quantidade de parâmetros não depende de T , ou seja, da extensão da defasagem das variáveis preditoras, mas tão somente da quantidade de *features* e da dimensão do *hidden state*.

3.3 Dos modelos

No que tange às variáveis preditoras, são utilizadas diferentes defasagens, nomeadamente, de 5, 15, 30, 60 e 90 pregões, e diferentes combinações de dados relacionados ao histórico de cotação do ativo. Os grupos consistiram em (a) somente o próprio preço de fechamento, (b) preços de abertura, máximo, mínimo e de fechamento (OHLC), (c) OHLC e volume e (d) OHLC, volume e RSI, BB inferior e superior do preço de fechamento.

Uma vez que a utilização de valores históricos encontra paralelo na análise grafista, busca-se testar também o impacto de alguns indicadores técnicos na atividade preditiva da rede neural. Esses grupos são identificados nas figuras e tabelas da seção seguinte como grupos 1, 4, 5 e 8, respectivamente, fazendo-se, assim, alusão à dimensionalidade das variáveis preditoras utilizadas em cada grupo.

A variável resposta utilizada em cada modelo consiste no valor de fechamento da ação nos 5 dias seguintes à observação mais recente (leiam-se pregões). É especificado um modelo para cada combinação de janela preditora, variáveis predictoras e ativo de referência.

A arquitetura da rede neural utilizada em todas as configurações de variáveis predictoras consiste em duas camadas LSTM de 40 unidades, seguidas por duas camadas densas, estas com respectivamente de 40 e 5 unidades e funções de ativação *RELU* e linear.

O treinamento se estende por até 100 épocas e utiliza *batches* de 16 observações. O otimizador utilizado é o otimizador Adam e a sessão de treinamento faz uso de *early stopping* com 10 épocas de paciência (relativamente à métrica MAPE da partição de validação), mas com opção de restaurar os pesos do melhor modelo até então obtido, obviando, assim, *overfitting*.

A avaliação de cada modelo, por sua vez, é feita à luz da partição de testes, obtendo-se as previsões para os horizontes $t+1$ a $t+5$ e cotejando-se esses valores com os de fato observados. As previsões são obtidas ao mesmo tempo para cada horizonte preditivo, ou seja, de modo independente.

3.4 Dos valores Shapley

Com o intuito de aferir a importância de cada variável utilizada pelo modelo para realizar suas previsões, utilizam-se valores Shapley. A contribuição das variáveis predictoras são feitas em 3 perspectivas, nomeadamente, da *feature* utilizada, sua defasagem e a consideração simultânea de ambas.

Os valores Shapley se caracterizam pelo atributo da aditividade. Isso significa que, para cada previsão feita pelo modelo, um valor é atribuído para cada variável utilizada (leiam-se *feature* e defasagem) de maneira que a soma desses valores seja igual ao valor previsto pelo modelo. No cálculo, é estipulado um valor base de referência, válido para todas as previsões, e a esse valor são somados os valores Shapley de cada variável utilizada, para cada uma das suas defasagens. Os valores que são adicionados ao valor de referência podem ser positivos ou negativos e seu módulo pode ser interpretado como a contribuição da *feature*/defasagem para a composição do resultado final.

O conceito de valores Shapley deita suas raízes na teoria dos jogos cooperativos e busca resolver o problema da atribuição do valor da participação de cada jogador em uma coalisão. Mais formalmente, a estimação do valor Shapley observa a seguinte formulação (Mas-Collel, Whinston, Green, 1995):

$$Sh_i(I, v) = \sum_{T \subseteq I} \frac{(|I| - |T|)! (|T| - 1)!}{|I|!} (v(T) - v(T \setminus \{i\})) \quad (9)$$

Essa expressão estipula que o valor Shapley do jogador i no grupo de jogadores I equivale à soma das contribuições marginais correspondentes à diferença entre a utilidade das coalisões formadas pelos integrantes de T (i incluso) e $T \setminus \{i\}$ (ou seja, i excluído) ponderada (a diferença) por $\frac{(|I| - |T|)! (|T| - 1)!}{|I|!}$. O valor da coalisão é definido pela função v . Além disso, convém anotar que o somatório ocorre por todos os subconjuntos T de I compostos por $|T|$ elementos em que i é o t -ésimo elemento desse subconjunto.

Aplicando-se essa ideia à seara de *machine learning*, revela-se o valor Shapley como um instrumento de interpretação da contribuição de cada variável na formação de cada previsão feita pelo modelo. Para isso, procede-se à soma dos valores Shapley em módulo de cada variável e sua defasagem relativamente a cada previsão levada a efeito no *dataset* de testes. Os valores assim obtidos são agregados e normalizados, de maneira que o resultado final possa ser interpretado como a contribuição global de cada atributo/defasagem para o modelo.

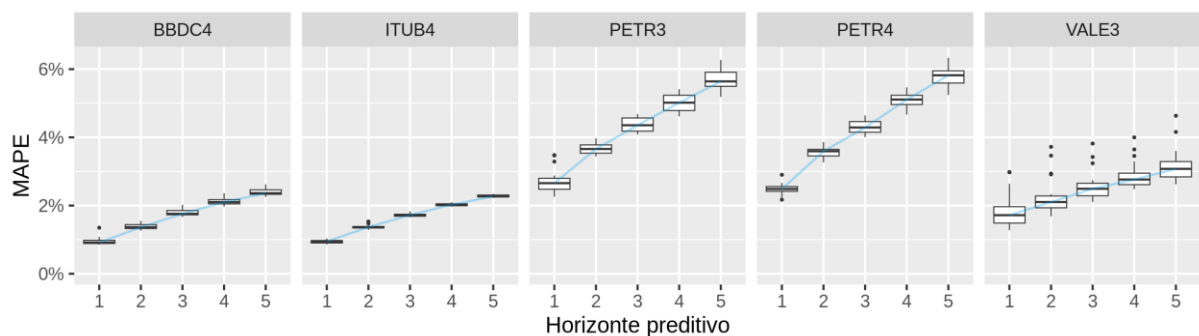


4 RESULTADOS

As figuras nesta seção colacionadas buscam descrever a performance dos modelos testados no que tange ao MAPE em diferentes perspectivas.

A Figura 2 e a Tabela 1 descrevem a distribuição do MAPE em função do horizonte preditivo, i.e., D+1 a D+5 pregões contados da observação mais recente. A linha em azul liga a mediana de cada grupo de observações. Nota-se, com efeito, um aumento progressivo do erro percentual médio conforme o horizonte preditivo se estende de D1 a D5.

Figura 2: Distribuição MAPE por horizonte preditivo



Fonte: Elaborada pelo autor.

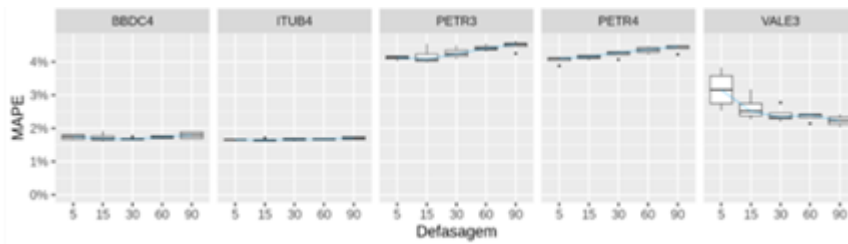
Tabela 1: Mediana do MAPE por ação e horizonte preditivo

Ação	t+1	t+2	t+3	t+4	t+5
BBDC4	0.92%	1.36%	1.76%	2.10%	2.36%
ITUB4	0.95%	1.37%	1.72%	2.02%	2.27%
PETR3	2.66%	3.66%	4.36%	5.02%	5.64%
PETR4	2.49%	3.59%	4.29%	5.11%	5.82%
VALE3	1.72%	2.10%	2.49%	2.76%	3.08%

Fonte: Elaborada pelo autor.

A Figura 3 e a Tabela 2 descrevem a distribuição do MAPE em função da defasagem das variáveis predictoras.

Figura 1: Distribuição do MAPE em função da defasagem das variáveis predictoras



Fonte: Elaborada pelo autor.

Tabela 2: Mediana do MAPE por ação e defasagem das variáveis predictoras

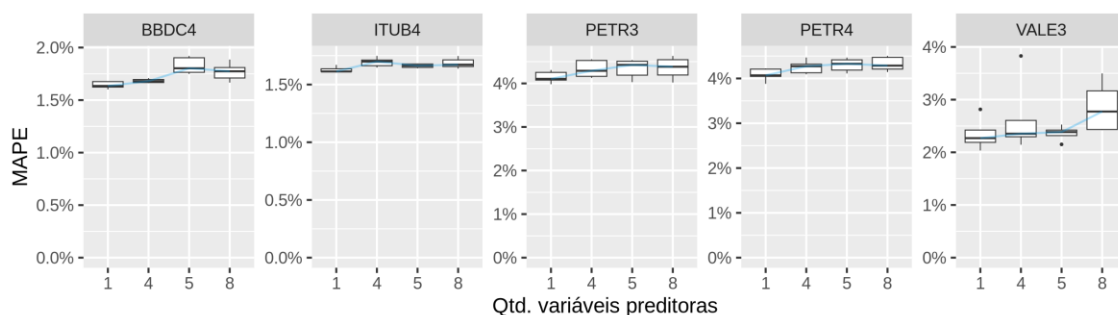
Ação	5	15	30	60	90
BBDC4	1.71%	1.68%	1.67%	1.77%	1.73%
ITUB4	1.66%	1.64%	1.64%	1.69%	1.73%
PETR3	4.04%	4.18%	4.23%	4.40%	4.52%
PETR4	4.10%	4.18%	4.32%	4.37%	4.45%
VALE3	2.79%	2.86%	2.74%	2.43%	2.31%

Fonte: elaborada pelo autor.

A linha que liga as medianas sugere que o aumento da defasagem das variáveis predictoras não tem o mesmo efeito em todas as ações consideradas. VALE3 parece se beneficiar por seu aumento, BBDC4 e ITUB4, indiferentes e PETR3 e PETR4 manifestamente não se beneficiam por utilizar observações mais antigas.

Na Figura 4, consta a distribuição do MAPE em função da quantidade de variáveis predictoras utilizadas. Os números de 1, 4, 5, e 8 fazem referência aos modelos que utilizaram como variáveis predictoras, incrementalmente, (i) preço de fechamento, (ii) OHLC, (iii) OHCL e volume e (iv) OHLC, volume e indicadores técnicos. A Tabela 3, por sua vez, traz o valor da mediana do MAPE de cada grupo.

Figura 4: Distribuição do MAPE em função da quantidade de variáveis predictoras



Fonte: Elaborada pelo autor.

Tabela 3: Mediana do MAPE por ação e quantidade de variáveis preditoras

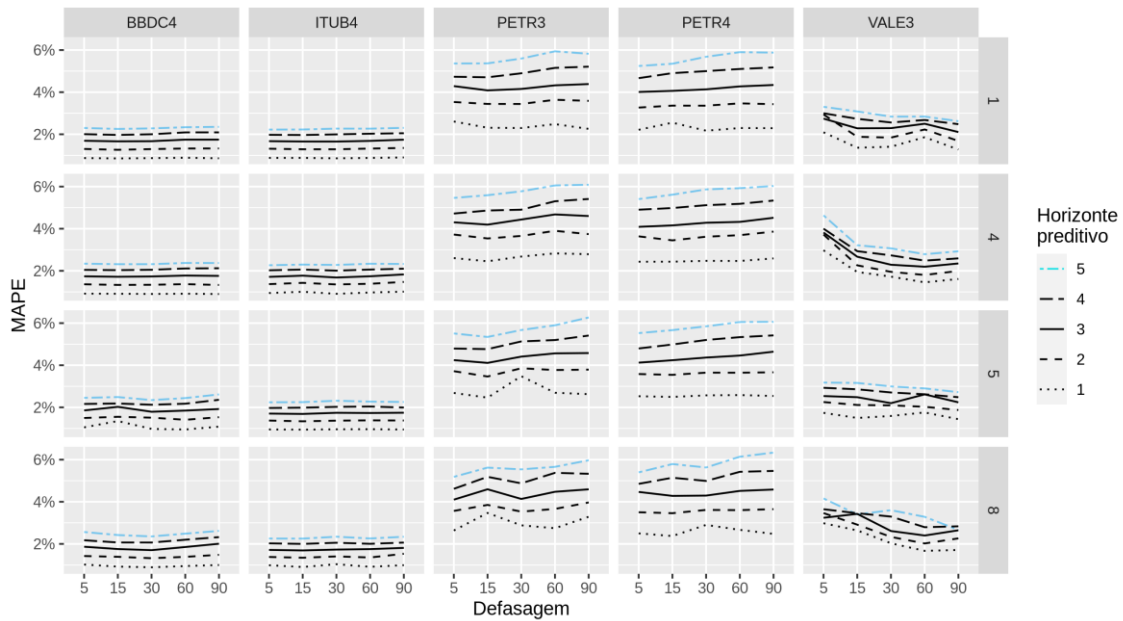
Ação	1	4	5	8
BBDC4	1.63%	1.68%	1.80%	1.77%
ITUB4	1.61%	1.70%	1.66%	1.67%
PETR3	4.10%	4.29%	4.43%	4.38%
PETR4	4.07%	4.27%	4.33%	4.29%
VALE3	2.27%	2.35%	2.38%	2.77%

Fonte: Elaborada pelo autor.

Os dados sugerem que o aumento da quantidade de variáveis preditoras não conduz necessariamente a uma melhoria do desempenho dos modelos de RNN testados. Chama a atenção, contudo, que, para algumas ações, a julgar pela mediana, a inclusão de indicadores técnicos de alguma forma reverteu a tendência de aumento do MAPE (e.g. BBDC4, PETR3 e PETR4). VALE3, por sua vez, parece ter sido muito beneficiada pela inclusão da variável relativa ao volume, notadamente porque a dispersão do MAPE ao longo dos vários modelos testados foi bastante reduzida.

A Figura 5 descreve a distribuição do MAPE simultaneamente em função do horizonte preditivo e da defasagem das variáveis preditoras, segmentando as observações em grupos conforme a quantidade de variáveis preditoras utilizadas. Nele é possível observar que a diferença do MAPE entre os horizontes temporais é maior em algumas ações (PETR3 e PETR4) dos que as demais.

Figura 2: Distribuição do MAPE em função da defasagem, segmentando-se por horizonte preditivo e quantidade de variáveis predictoras

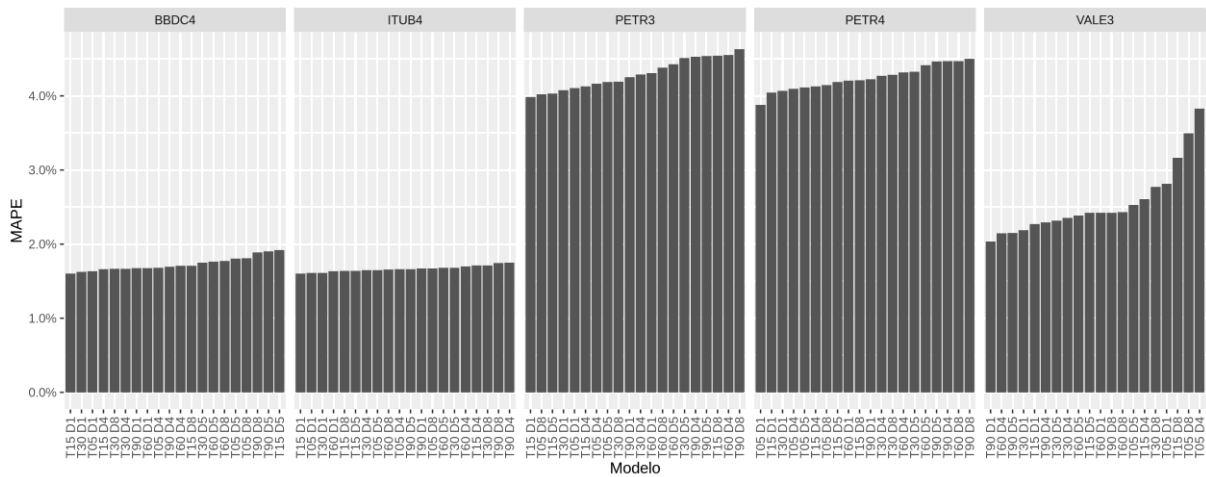


Fonte: Elaborada pelo autor.

Com efeito, observa-se uma maior dispersão do erro em função do horizonte preditivo em relação a PETR3 e PETR4 ao longo de cada defasagem considerada e de grupo de variáveis utilizado. Outrossim, o gráfico sugere que VALE3, em todas as combinações de variáveis predictoras testadas, tendeu a se beneficiar pelo aumento da sua defasagem. BBDC4 e ITUB4, por sua vez, parecem algo indiferentes às variações de defasagem e de grupo de variáveis utilizado.

Por derradeiro, a Figura 6 descreve a performance de cada modelo individualmente considerado, ou seja, por ativo, defasagem e grupo de variáveis, agregando-se, contudo, o MAPE em relação a todos os horizontes preditivos utilizados (D1 a D5). O MAPE dos modelos com melhor performance pode ser visto na Tabela 4.

Figura 3: Distribuição do MAPE por ação e modelo utilizado



Fonte: Elaborada pelo autor.

Tabela 4: MAPE por ação, defasagem e quantidade de variáveis predictoras

Ação	Defasagem	Qtd. Variáveis	MAPE
BBDC4	15	1	1.60%
ITUB4	15	1	1.60%
PETR3	15	1	3.98%
PETR4	5	1	3.88%
VALE3	90	1	2.03%

Fonte: Elaborada pelo autor.

A Figura 6 revela que o melhor desempenho da RNN foi em relação aos ativos BBDC4 e ITUB4, os quais se revelaram aparentemente mais insensíveis à defasagem e ao grupo de variáveis predictoras utilizado. No outro extremo, nota-se uma maior variabilidade do MAPE de VALE3 quando se alteram a defasagem e o grupo de variáveis predictoras. Uma variabilidade algo mais leve pode ser identificada em relação a PETR3 e PETR4, cujo MAPE, contudo, é mais elevado do que as demais.

Isso não obstante, observa-se que, para cada ativo, o melhor modelo utilizou a configuração com somente uma variável preditora i.e. o próprio valor de fechamento da ação. VALE3 foi a que mais se beneficiou com o aumento defasagem (modelo T90 D1), enquanto PETR4 alcançou seu melhor resultado como modelo mais simples dentre todos i.e. 5 pregões de defasagem e somente 1 variável preditora (T05 D1). BBDC4, ITUB3 e PETR3, por sua vez, parece que colheram

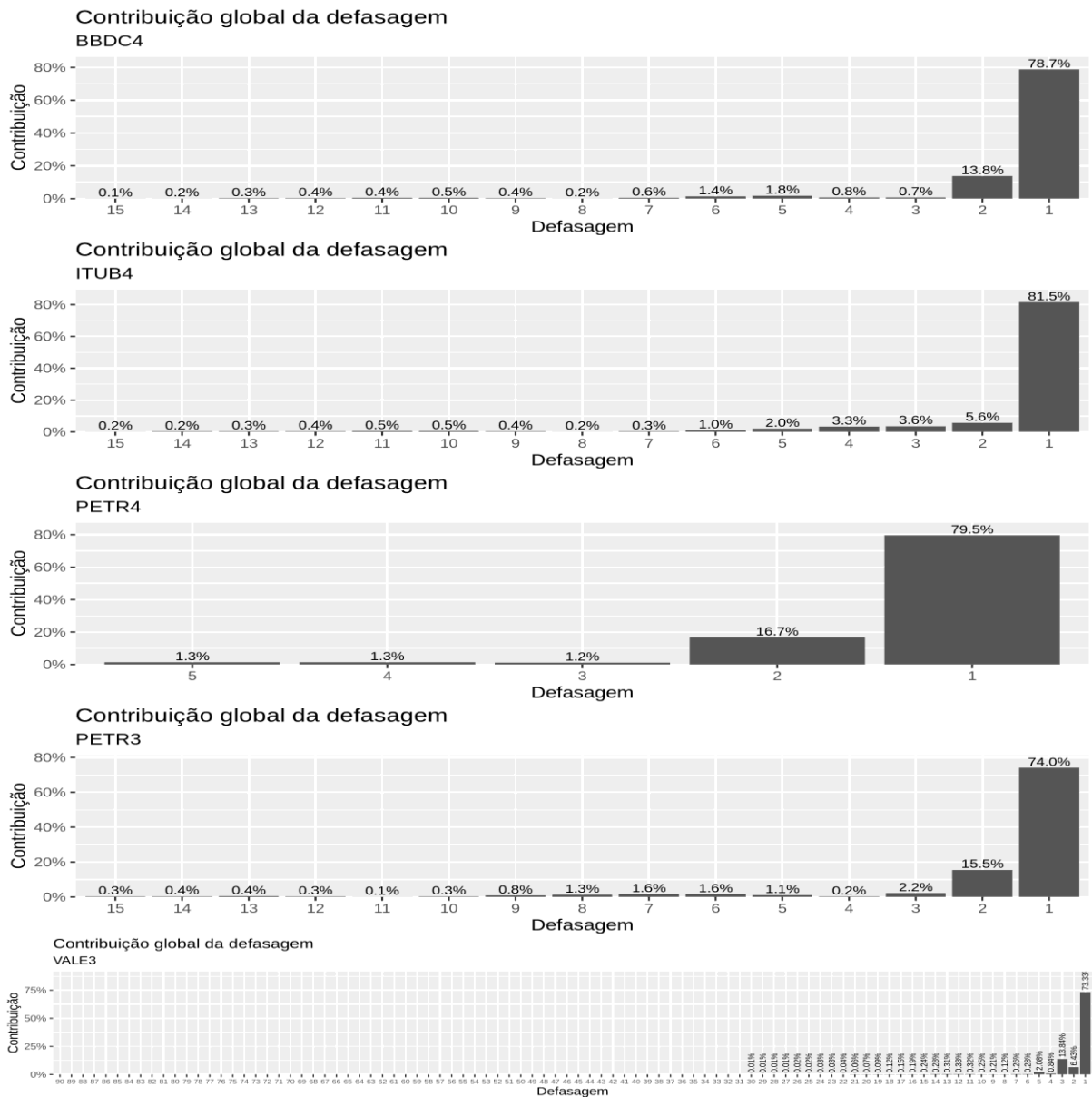
algum benefício pelo aumento da defasagem da variável preditora até 15 pregões.

Os dados, assim, permitem concluir que nem todos os ativos se beneficiam da mesma maneira com o aumento da complexidade do modelo via aumento da defasagem. Todos, contudo, alcançaram melhores resultados com a configuração univariada, ou seja, tendo tão somente o próprio preço de fechamento como variável preditora.

Até aqui foram consideradas defasagens e grupos de variáveis com o intuito de identificar a melhor combinação na perspectiva da performance da RNN-LSTM para cada ativo. Mas isso não revela como e em que medida cada fator contribuiu para alcançar os resultados obtidos. Sendo assim, a presente seção faz uso de valores Shapley com o intuito de descortinar como cada defasagem e variável contribuiu de maneira global para a formação das previsões da RNA.

Uma vez que vários modelos foram testados, aborda-se a seguir a contribuição global da defasagem do melhor modelo de cada ativo. Os painéis da Figura 7 ilustram, de modo global, a importância de cada defasagem na formação da estimativa produzida por cada modelo univariado. Nota-se, invariavelmente, a importância da observação mais recente em todas as ações no modelo univariado. Nos modelos que utilizaram 15 defasagens, observa-se que até a sexta observação mais recente as contribuições são maiores do que 1%.

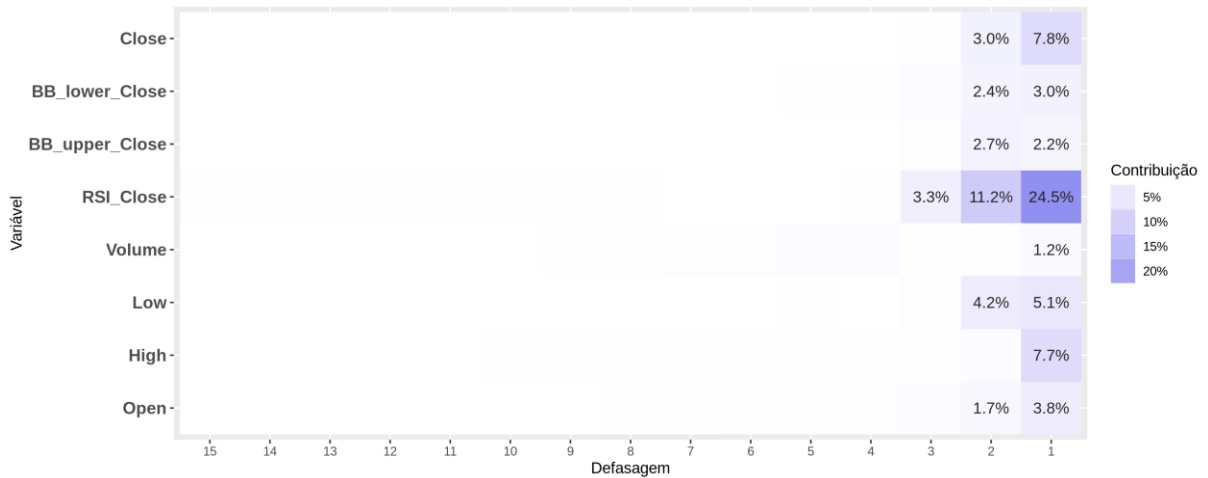
Figura 4: Contribuição global das defasagens nos melhores modelos de cada ação – percentual de valores Shapley absolutos



Fonte: Elaborada pelo autor.

A Figura 8 traz a contribuição global de defasagens e variáveis preditoras nas estimativas produzidas modelo ITUB4 T15 D8 que é o melhor modelo para essa ação que utiliza todas as variáveis preditoras testadas. Observa-se um predomínio da influência das observações mais recentes e, de modo inusitado, do indicador RSI (preço de fechamento) para a composição das estimativas relativas ao ano de 2022. Isso pode ser um sinal da utilidade da combinação de indicadores técnicos com redes neurais do tipo LSTM de maneira a aperfeiçoar sua capacidade preditiva.

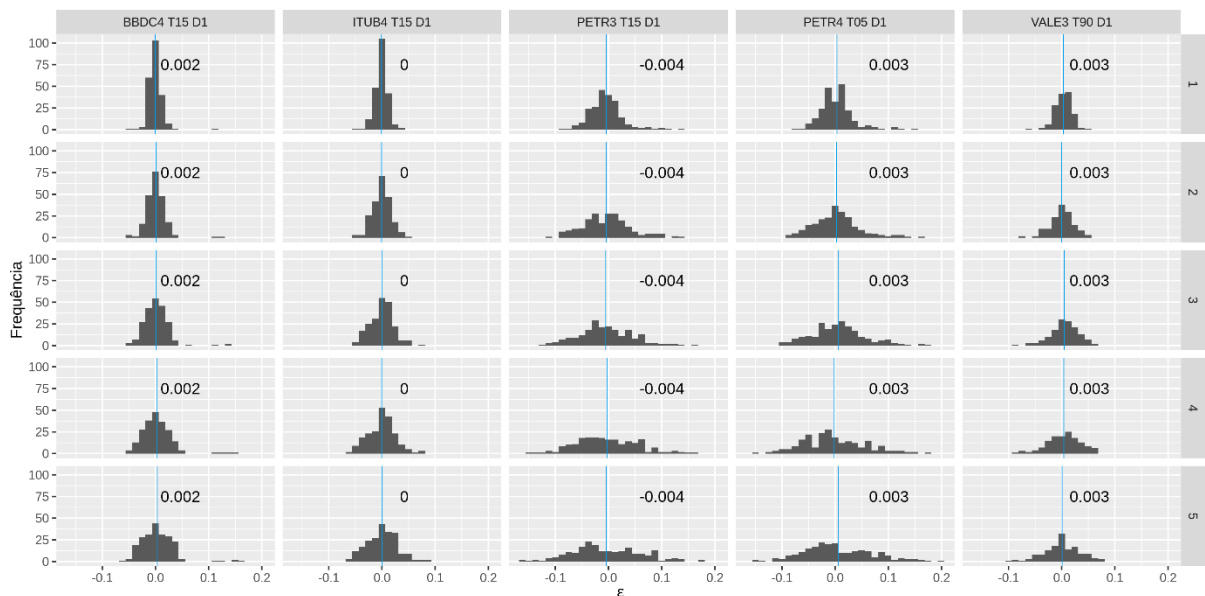
Figura 5: Contribuição global de valores Shapley absolutos para variáveis preditoras e suas defasagens relativamente ao modelo ITUB4 T15 D8



Fonte: Elaborada pelo autor.

Uma vez descrita a influência das variáveis preditoras e suas defasagens no MAPE e na formação das previsões, cumpre considerar, a seguir, aspectos relacionados aos resíduos. A Figura 9 contém histogramas da distribuição dos erros dos modelos com melhor MAPE para cada ativo em relação a cada horizonte preditivo considerado. A linha vertical azul diz respeito à média dos resíduos e são todas elas praticamente iguais a zero.

Figura 6: Distribuição dos resíduos por modelo e por horizonte preditivo



Fonte: Elaborada pelo autor.

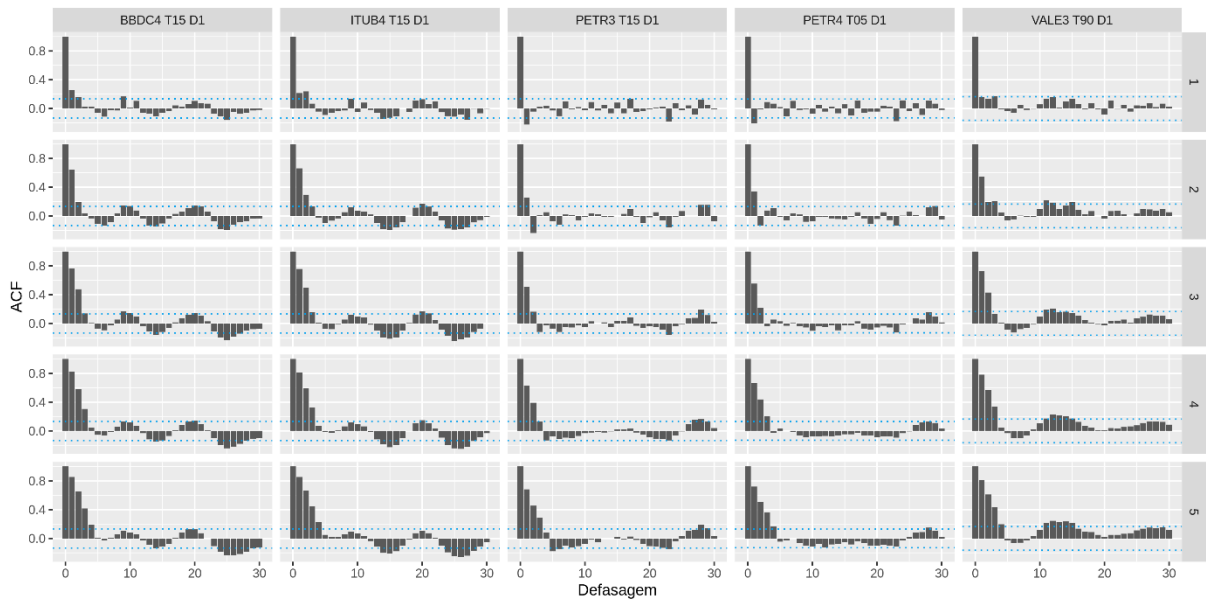
A Tabela 5 contempla o resultado do teste de Shapiro-Wilk da normalidade dos resíduos descritos na Figura 9. Na tabela constam a estatística de teste e respectivo p-valor, sendo a hipótese nula pela normalidade da distribuição. Convém destacar, que a um nível de significância de 5%, a normalidade dos resíduos somente é rejeitada no horizonte preditivo t+1 em relação ao modelo ITUB4 com 15 defasagens e uma variável preditora (preço de fechamento). A um nível de significância de 1%, não se rejeita a hipótese nula de que os resíduos de VALE3 com 90 defasagens e uma variável preditora (preço de fechamento) são normalmente distribuídos.

Tabela 5: Estatística e p-valor do teste Shapiro-Wilk dos resíduos por horizonte preditivo

Modelo	t+1	t+2	t+3	t+4	t+5
BBDC4 T15 D1	W = 0.82 p = 0.00	W = 0.84 p = 0.00	W = 0.87 p = 0.00	W = 0.88 p = 0.00	W = 0.88 p = 0.00
ITUB4 T15 D1	W = 0.99 p = 0.03	W = 0.99 p = 0.50	W = 0.99 p = 0.08	W = 0.99; p = 0.10	W = 0.99 p = 0.07
PETR3 T15 D1	W = 0.93 p = 0.00	W = 0.98 p = 0.01	W = 0.99 p = 0.24	W = 0.99 p = 0.29	W = 0.99 p = 0.18
PETR4 T05 D1	W = 0.91 p = 0.00	W = 0.97 p = 0.00	W = 0.98 p = 0.01	W = 0.99 p = 0.04	W = 0.99 p = 0.04
VALE3 T90 D1	W = 0.98 p = 0.04	W = 0.98 p = 0.04	W = 0.98 p = 0.07	W = 0.99 p = 0.16	W = 0.99 p = 0.41

Fonte: Elaborada pelo autor.

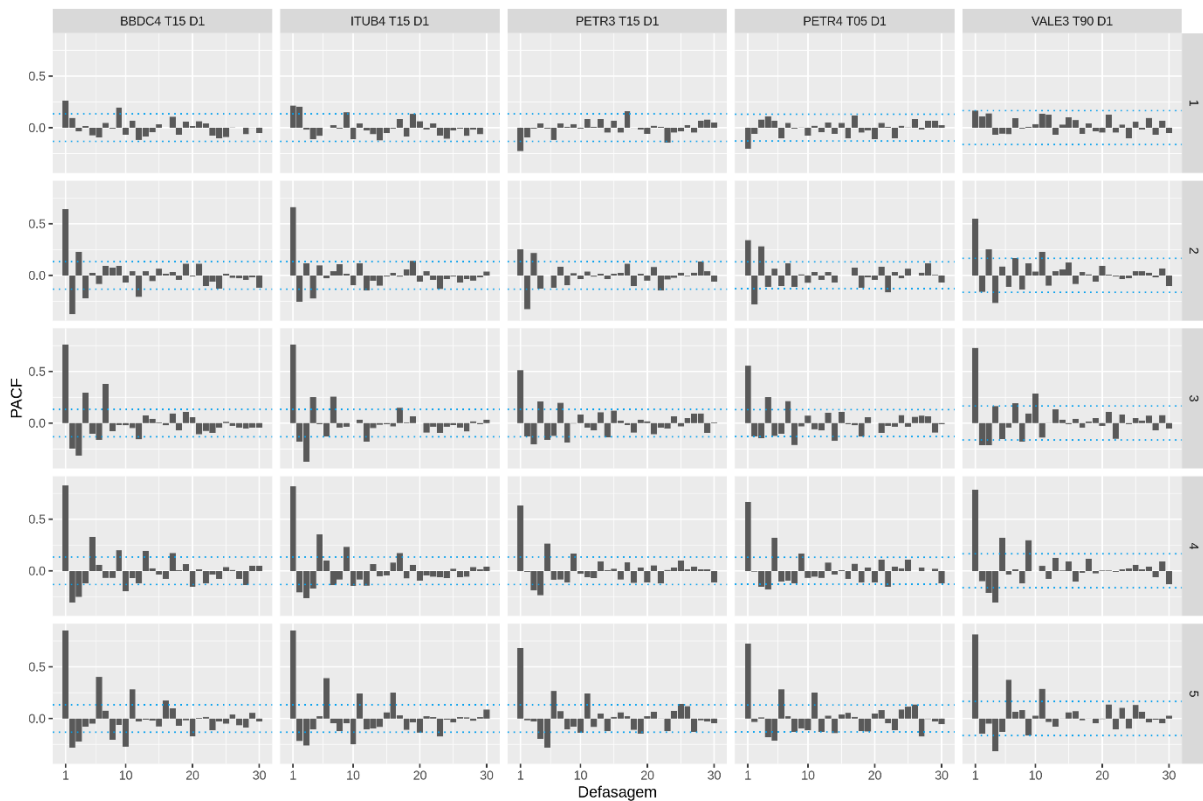
A Figura 10 descreve a função de autocorrelação (FAC) dos desses resíduos em relação a cada horizonte temporal. O decaimento lento do valor da FAC sugere que os resíduos exibem ainda certo grau de autocorrelação, o qual parece aumentar com a protração no tempo do horizonte preditivo (1 a 5).

Figura 7: Função de autocorrelação dos resíduos

Fonte: Elaborada pelo autor.

A Figura 11, a seu turno, traz valores relativos à função de autocorrelação parcial (FACP) dos resíduos em função do modelo utilizado e do horizonte preditivo. Observa-se que, à exceção do horizonte preditivo imediato (i.e. $t+1$), os demais resíduos, independentemente do modelo, sugerem um fator de autocorrelação muito forte ainda com a defasagem imediata e, de modo mais ameno, mas ainda significativo, com defasagens maiores.

Figura 8: Função de autocorrelação parcial dos resíduos



Fonte: Elaborada pelo autor.

Convém sublinhar que a FACP não decai de modo suave com a defasagem, havendo valores significativos separados entre si por intervalos regulares e.g. de 4 pregões. Observa-se ainda que todos os modelos considerados, em se tratando do horizonte preditivo $t+5$, exibem valores de autocorrelação parcial significativos separados por 4 defasagens até a defasagem 11. Isso pode ser sugestivo de algum padrão relacionado ao dia da semana em que a ação foi negociada.

Esses fatos sugerem que a arquitetura utilizada para testar a influência das variáveis predictoras e as suas defasagens nas previsões feitas com a utilização RNN do tipo LSTM possivelmente não exauriram o potencial dos dados. Assim, a ilação que parece emergir dos dados é a de que a utilização de arquiteturas mais profundas, com mais camadas LSTM empilhadas, e com maior dimensionalidade podem, eventualmente, ser úteis para exaurir padrões de autocorrelação que parecem persistir nos resíduos.

Os achados trazidos neste trabalho buscaram, assim, colocar em evidência a influência das variáveis predictoras na qualidade das previsões levadas a efeito por redes neurais artificiais que fazem uso de camadas LSTM. Ao utilizar a mesma arquitetura ao longo de diferentes combinações de ativos (ações), séries predictoras e defasagens, pode-se

colher evidências de que o aumento irrefletido de defasagens e a inserção automática de mais variáveis não conduz necessariamente a um modelo com melhor performance. Isso pode ser importante para uma melhor compreensão das vantagens e limitações de camadas neurais do tipo recorrente na aplicação aqui considerada.



5

5 CONCLUSÃO

O presente trabalho buscou examinar, de maneira experimental, a efetividade da configuração de variáveis preditoras nas previsões produzidas por redes neurais artificiais recorrentes em relação ao preço de fechamento de ações negociadas na bolsa de valores B3. Os resultados indicam que previsões relativamente precisas podem ser produzidas com uma combinação ótima de defasagens e dimensionalidade das variáveis preditoras. A escolha parece ser fortemente dirigida a dados, dependendo, em essência, do ativo considerado.

Além disso, não parece haver evidência de que a utilização de indicadores técnicos conduza necessariamente a ganhos na qualidade das previsões. Isso soa coerente na medida em que se estima que a rede neural possa, durante a fase de treinamento, identificar padrões refletidos nesses indicadores. Todavia, é fato que, em relação a alguns modelos, sua utilização coincidiu com a melhoria do MAPE. De modo surpreendente, para a ação ITUB4, utilizando-se 15 defasagens, o indicador RSI do preço de fechamento se revelou assaz útil, globalmente, para a formação das predições do modelo.

Os resultados obtidos, assim, sugerem a existência de fundamentos para a abordagem da análise técnica no sentido de que a série histórica parece conter informação útil à previsão do preço de ações. O trabalho não tratou da otimização da arquitetura da rede neural nem utilizou heurísticas como *drop out*. Buscou tão somente colher achados que possam nortear a utilização prática desse tipo de rede neural no que tange à configuração dos *datasets* utilizados.

Estudos futuros podem considerar o aprofundamento de algumas linhas de pesquisa. Mencione-se, à guisa de ilustração, a utilização de redes neurais recorrentes treinadas especificamente para a previsão da direção da alteração do preço da ação ou, ainda, a utilização de notícias, juntamente com cotações históricas, para aperfeiçoar a capacidade preditiva dos modelos. Além disso, a utilização de indicadores da análise grafista parece ser uma linha promissora para aperfeiçoar a capacidade preditiva de modelos RNN do tipo LSTM.



REFERÊNCIAS

REFERÊNCIAS

REFERÊNCIAS

ABIODUN, O. J.; OMOWARE, J. M. A comparative analysis of stock series prediction of apple and google datasets using deep learning. **FUW Trends in Science and Technology Journal**. p. 366–374, 13 ago. 2023. Disponível em: <http://www.ftstjournal.com/uploads/docs/82%20Article%2058%20pp%20366-374.pdf>. Acesso em: 15 jan. 2024.

ALASIRI, R. A.; QAHMASH, A. Analysis and Forecasting of Saudi Stock Market Using Time Series Algorithms. Em: **2023 3RD International Conference on Computing and Information Technology (ICCIT)**. Tabuk, Saudi Arabia: IEEE, 13 set. 2023. Resumo disponível em: <https://ieeexplore.ieee.org/document/10273971/>. Acesso em: 15 jan. 2024.

BATHLA, G.; RANI, R.; AGGARWAL, H. Stocks of year 2020: prediction of high variations in stock prices using LSTM. **Multimedia Tools and Applications**, v. 82, p. 1–17, 25 fev. 2022. Disponível em: https://www.researchgate.net/publication/358872443_Stocks_of_year_2020_prediction_of_high_variations_in_stock_prices_using_LSTM. Acesso em: 15 jan. 2024.

BUDIHARTO, W. Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM). **Journal of Big Data**, v. 8, n. 1, p. 47, 11 mar. 2021. Disponível em: <https://journalofbigdata.springeropen.com/counter/pdf/10.1186/s40537-021-00430-0>. Acesso em: 15 jan. 2024.

CHANDOLA, D. *et al.* Forecasting directional movement of stock prices using deep learning. **Annals of Data Science**, v. 10, n. 5, p. 1361–1378, 1 out. 2023. Disponível em: <https://link.springer.com/content/pdf/10.1007/s40745-022-00432-6.pdf>. Acesso em: 15 jan. 2024.

CHEN, S. High-frequency stock return prediction using state-of-the-art deep learning models. **International Journal of Financial Engineering**, v. 10, n. 03, p. 2350023, set. 2023. Resumo disponível em: <https://www.worldscientific.com/doi/abs/10.1142/S2424786323500238>. Acesso em: 15 jan. 2024.

DHOKANE, R.; AGARWAL, S. **Stock market prediction using the LSTM algorithm in association with the Relative Strength Index (RSI) and Exponential Moving Average (EMA) indicators.** Disponível em: https://www.researchgate.net/publication/374469075_Stock_market_prediction_using_the_LSTM_algorithm_in_association_with_the_Relative_Strength_Index_RSI_and_Exponential_Moving_Average_EMA_indicators/. Acesso em: 15 jan. 2024.

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, jun. 2020. Disponível em: <http://link.springer.com/10.1007/s13042-019-01041-1>. Acesso em: 15 jan. 2024.

DIXON, M. F.; HALPERIN, I.; BILOKON, P. **Machine learning in finance: from theory to practice.** Cham, Switzerland: Springer, 2020.

FIROUZJAEI, J. T.; KHALILIYAN, P. **The interpretability of LSTM models for predicting oil company stocks: impact of correlated features.** arXiv, 20 dez. 2023. Disponível em: <http://arxiv.org/abs/2201.00350>. Acesso em: 15 jan. 2024.

GERS, F.; ECK, D.; SCHMIDHUBER, J. Applying LSTM to Time Series Predictable through Time-Window Approaches. Em: **Artificial Neural Networks - ICANN 2001, International Conference Vienna**, Austria, August 21-25, 2001. DOI: 10.1007/3-540-44668-0_93. Disponível em: https://www.researchgate.net/publication/221079584_Applying_LSTM_to_Time_Series_Predictable_through_Time-Window_Approaches. Acesso em: 15 jan. 2024.

GUO, S. Prediction of China Stock returns under the COVID-19 Pandemic based on the LSTM Model. **Highlights in Science, Engineering and Technology**, v. 68, p. 182–188, 9 out. 2023. Disponível em: https://www.researchgate.net/publication/374705396_Prediction_of_China_Stock_returns_under_the_COVID-19_Pandemic_based_on_the_LSTM_Model. Acesso em: 15 jan. 2024.

HARGREAVES, C.; CHEN, L. **Stock prediction using deep learning with long-short-term-memory networks.** 11 dez. 2020. Disponível em: https://www.researchgate.net/publication/347515776_Stock_Prediction_Using_Deep_Learning_with_Long-Short-Term-Memory_Networks. Acesso em: 15 jan. 2024.

KADADEKAR, S. *et al.* Digital Currency Price Prediction using Machine Learning. **International Journal for Research in Applied Science and Engineering Technology**, v. 11, p. 338–355, 29 set. 2023. Disponível em: https://www.researchgate.net/publication/374738343_Digital_Currency_Price_Prediction_using_Machine_Learning. Acesso em: 15 jan. 2024.

KAUSHIK, R. *et al.* Performance evaluation of deep neural networks for forecasting time-series with multiple structural breaks and high volatility. **CAAI Transactions on Intelligence Technology**, v. 6, n. 3, p. 265–280, set. 2021. Disponível em: <https://arxiv.org/abs/1911.06704>. Acesso em: 15 jan. 2024.

KUMAR, K.; HAIDER, MD. T. U. Enhanced prediction of intra-day stock market using metaheuristic optimization on RNN–LSTM network. **New Generation Computing**, v. 39, n. 1, p. 231–272, abr. 2021. Resumo disponível em: <https://link.springer.com/article/10.1007/s00354-020-00104-0>. Acesso em: 15 jan. 2024.

LAWI, A.; MESRA, H.; AMIR, S. Implementation of Long Short-Term Memory and Gated Recurrent Units on grouped time-series data to predict stock prices accurately. **Journal of Big Data**, v. 9, 7 jul. 2022. Resumo disponível em: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00597-0>. Acesso em: 15 jan. 2024.

LV, G.; CUI, J. **Merging public opinion information and stock numerical data for stock trend prediction based on deep learning**. Em: INTERNATIONAL CONFERENCE ON IMAGES, SIGNALS, AND COMPUTING (ICISC 2023). SPIE, 21 ago. 2023. Disponível em: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12783/127830J/Merging-public-opinion-information-and-stock-numerical-data-for-stock/10.1117/12.2691661.full>. Acesso em: 15 jan. 2024

MENGQIN, Z.; YE, P. **Research on Machine Learning System of Stock Prediction Based on Computer LSTM Technology**. Em: 2023 IEEE INTERNATIONAL CONFERENCE ON SENSORS, ELECTRONICS AND COMPUTER ENGINEERING (ICSECE). Jinzhou, China: IEEE, 18 ago. 2023. Resumo disponível em: <https://ieeexplore.ieee.org/document/10263553/>. Acesso em: 15 jan. 2024

MINAEE, S. *et al.* **Deep learning based text classification: a comprehensive review**, 4 jan. 2021. Disponível em: <http://arxiv.org/abs/2004.03705>. Acesso em: 15 jan. 2024

MURTHY, G. K. *et al.* Stock price prediction using LSTM, Support Vector Machine, Random Forest and GRU. **Industrial Engineering Journal**. v. 52, n. 6, jun. 2023. Disponível em: http://www.journal-iiie-india.com/1_june_23/5.6_6.pdf. Acesso em: 15 jan. 2024.

NEUSSER, K. **Time Series Econometrics**. Springer, 2016.

PINHEIRO, T. J. **Modelos de deep learning e previsão de preço de ações: estudos de casos da bolsa brasileira**. 30 mar. 2022. Disponível em:

<https://tede2.uepg.br/jspui/bitstream/prefix/3672/1/Thiago%20Jos%C3%A9%20Pinheiro.pdf>. Acesso em: 15 jan. 2024.

THU, H.; THANH, T.; LE QUY, T. Dynamic Sliding Window and Neighborhood LSTM-Based Model for Stock Price Prediction. **SN Computer Science**, v. 3, 1 maio 2022. Resumo disponível em: <https://link.springer.com/article/10.1007/s42979-022-01158-1>. Acesso em: 15 jan. 2024.

VOHRA, A. A.; TANNA, P. Impact of News Sentiments on Stock Price Prediction. **International Conference on Science, Engineering and Technology (ICSET 2022)**, 2022. Disponível em: https://soe.rku.ac.in/conferences/data/43_4234_ICSET%202022.pdf. Acesso em: 15 jan. 2024.

VOHRA, A. A.; TANNA, P. J. Evaluation of factors involved in predicting Indian stock price using machine learning algorithms. **International Journal of Business Intelligence and Data Mining**, v. 23, n. 3, p. 201–263, 1 jan. 2023. Resumo disponível em: <https://doi.org/10.1504/ijbidm.2023.133147>. Acesso em 15 jan. 2024.

ZHANG, C.; SJARIF, N. N. A.; IBRAHIM, R. Deep learning models for price forecasting of financial time series: a review of recent advancements: 2020-2022. **WIREs Data Mining and Knowledge Discovery**, p. e1519, 28 set. 2023. Disponível em: <https://arxiv.org/abs/2305.04811>. Acesso em: 15 jan. 2024.



idp

Bo
pro
cit
ref
Noss
são e

idp

A ESCOLHA QUE
TRANSFORMA
O SEU CONHECIMENTO