



UNIVERSIDADE FEDERAL DO TRIANGULO MINEIRO - UFTM
CURSO DE LICENCIATURA EM MATEMÁTICA

Análise de ativos financeiros por meio de Rede Neural Recorrente do tipo
LSTM

Vinícios Antônio Passos Balduino

Uberaba - MG
Março de 2023

Análise de ativos financeiros por meio de Rede Neural Recorrente do tipo LSTM

Vinícios Antônio Passos Balduino

Trabalho de conclusão de curso apresentado ao Curso de Licenciatura em Matemática da UFTM, como requisito parcial para obter o título de Licenciado em Matemática, sob a orientação do professor Dr. Osmar Aléssio e a co-orientação do professor Dr. Flávio Luiz de Moraes Barboza.

Orientador: Prof. Dr. Osmar Aléssio

Co-orientador: Prof. Dr. Flávio Luiz de Moraes Barboza .

Uberaba - MG
Março de 2023

Análise de ativos financeiros por meio de Rede Neural Recorrente do tipo LSTM

Vinícios Antônio Passos Balduino

Trabalho de conclusão de curso apresentado ao
Curso de Licenciatura em Matemática da UFTM,
como requisito parcial para obter o título de Licen-
ciado em Matemática, sob a orientação do professor
Dr. Osmar Aléssio e a co-orientação do professor
Dr. Flávio Luiz de Moraes Barboza..

____ de _____ 2023.

Prof. Dr. Osmar Aléssio
Orientador
UFTM

Prof. Dr. Flávio Luiz de Moraes Barboza
Co-orientador
UFU

Prof. Dr. Leandro Cruvinel Lemes
UFTM

M.e Alan Lopes Melo
UFTM

Resumo

Devido ao desenvolvimento tecnológico ao qual a sociedade está submersa a possibilidade da criação de ferramentas antes impensáveis ou impossíveis se torna realidade. Com a quantidade absurda de dados ao qual se pode ter acesso, analisar e tirar conclusões assertivas sem o auxílio de uma boa ferramenta torna o trabalho excessivamente exaustivo. Este trabalho se dedica a explorar a criação de um modelo de rede neural recorrente do tipo LSTM, o submetendo diferentes testes de parâmetros e comparando seus resultados a modelos teóricos como o passeio aleatório e média móvel. Por fim constatamos que o melhor desempenho do modelo foi semelhante ao passeio aleatório, enquanto que o modelo com parâmetro *stateful=True* apresentou o pior desempenho nos testes.

Palavras-chave: Rede neural recorrente; Predição do preço de ativo; LSTM; Stateful.

Abstract

Due to the technological development that society is submerged in, the possibility of creating tools that were previously unthinkable or impossible becomes a reality. With the absurd amount of data that can be accessed, analyzing and drawing assertive conclusions without the aid of a good tool becomes excessively exhausting work. This work is dedicated to exploring the creation of a recurrent neural network model of the LSTM type, subjecting it to different parameter tests and comparing its results to theoretical models such as random walk and moving average. Finally, we found that the model's best performance was similar to random walk, while the model with the parameter "stateful=True" showed the worst performance in the tests.

Keywords: Recurrent Neural Network; Asset price prediction; LSTM; Stateful.

Sumário

1	INTRODUÇÃO	10
1.1	Justificativa	10
1.2	Objetivos	11
1.2.1	Objetivo Geral	11
1.2.2	Objetivos Específicos	11
1.3	Organização do Trabalho	11
2	DESENVOLVIMENTO	12
2.1	Fundamentação teórica	12
2.1.1	Introdução ao Mercado financeiro	12
2.1.1.1	Estratégias básicas de investimento	13
2.1.1.2	Hipóteses de mercado.	14
2.1.2	Aprendizado de Máquina	15
2.1.2.1	O que são as redes neurais artificiais.	16
2.1.2.1.1	Rede neural simplificada	16
2.1.2.1.2	Redes neurais complexas	18
2.1.2.2	O funcionamento da rede neural	22
2.1.3	Série temporal	23
2.2	Metodologia	25
2.2.1	Séries analisadas e descrição dos dados	25
2.2.1.1	Ambev	25
2.2.1.2	Banco do Brasil	26
2.2.1.3	Banco Bradesco	26
2.2.1.4	Taes	26
2.2.1.5	Cemig	27
2.2.2	Pré processamento dos dados.	27
2.2.2.1	Métricas de desempenho	29
2.2.2.1.1	Erro médio absoluto - MAE	30
2.2.2.1.2	Erro percentual médio - MAPE	30
2.2.2.1.3	Média do erro quadrado - MSE	30
2.2.2.1.4	Raiz quadrada da média de erro quadrado - RMSE	30
2.2.2.1.5	R quadrado - R^2	31
2.3	Resultados e discussões	32
3	CONCLUSÃO	40

REFERÊNCIAS 41

Lista de ilustrações

Figura 1 – Rede neural artificial do tipo <i>Multilayer Perceptron</i>	18
Figura 2 – Rede neural recorrente	18
Figura 3 – Arquitetura de célula do tipo LSTM.	19
Figura 4 – Estado da memória da célula LSTM.	20
Figura 5 – Entrada de dados e portão de esquecimento.	21
Figura 6 – Portão de entrada.	21
Figura 7 – Portão de saída.	22
Figura 8 – Influencia do parâmetro <i>stateful</i>	29
Figura 9 – Comparativo do preço real e previsto pela rede para a ação ABEV3.SA	32
Figura 10 – Comparativo do preço real e previsto pela rede para a ação BBAS3.SA	33
Figura 11 – Comparativo do preço real e previsto pela rede para a ação BBDC4.SA	34
Figura 12 – Comparativo do preço real e previsto pela rede para a ação TAEE11.SA	35
Figura 13 – Comparativo do preço real e previsto pela rede para a ação CMIG4.SA	36
Figura 14 – Perdas de treinamento das séries	37

Lista de tabelas

Tabela 1 – Subconjunto da série ABEV3.SA com os 5 primeiros registros	26
Tabela 2 – Subconjunto da série BBAS3.SA com os 5 primeiros registros	26
Tabela 3 – Subconjunto da série BBDC4.SA com os 5 primeiros registros	26
Tabela 4 – Subconjunto da série TAEE11.SA com os 5 primeiros registros	27
Tabela 5 – Subconjunto da série CMIG4.SA com os 5 primeiros registros	27
Tabela 6 – Métricas do conjunto de teste da série ABEV3.SA	32
Tabela 7 – Métricas do conjunto de teste da série BBAS3.SA	33
Tabela 8 – Métricas do conjunto de teste da série BBDC4.SA	34
Tabela 9 – Métricas do conjunto de teste da série TAEE11.SA	35
Tabela 10 – Métricas do conjunto de teste da série CMIG4.SA	36
Tabela 11 – Métricas dos modelos para o conjunto de teste da série ABEV3.SA . .	38
Tabela 12 – Métricas dos modelos para o conjunto de teste da série BBAS3.SA . .	38
Tabela 13 – Métricas dos modelos para o conjunto de teste da série BBDC4.SA . .	38
Tabela 14 – Métricas dos modelos para o conjunto de teste da série TAEE11.SA . .	39
Tabela 15 – Métricas dos modelos para o conjunto de teste da série CMIG4.SA . .	39

1 Introdução

O mercado de ações é parte importante da economia de um país e quanto mais ativo, maior é o impacto positivo economicamente. Não é de se surpreender a quantidade de estudos acerca deste objeto, inclusive de áreas indiretamente relacionadas.

Teorias econômicas recentes abordam a economia como comportamental, tratam o mercado como possivelmente previsível (LO, 2004). Isso quer dizer que, segundo estas teorias, seria possível prever a movimentação do mercado em algum determinado momento. Caso o mercado seja de fato previsível significa que lucros extraordinários podem ser feitos a baixíssimo risco para o investidor.

A evolução do preço de uma ação ao longo do tempo pode ser interpretada como uma série temporal. É um objeto estatístico onde um dos principais assuntos discutidos é sua previsibilidade, que pode ser medida de acordo com alguns testes como (LJUNG; BOX, 1978).

Em um sociedade altamente tecnológica como a atual, a possibilidade do desenvolvimento de ferramentas antes impensáveis ou impossíveis de serem construídas, acaba se tornando realidade. Dentre os principais métodos estatísticos pelos quais é possível estudar séries temporais, uma abordagem que vem se destacando ao longo dos anos são as redes neurais artificiais (ANGELO et al., 2011), (BULÍN; ŠMÍDL; ŠVEC, 2019).

Neste trabalho são discutidos os conceitos do funcionamento das redes neurais, com foco no modelo de rede neural recorrente do tipo LSTM. Este modelo é objeto de estudo de outros autores da atualidade, e vem ganhando destaque no meio científico devido ao, já esperado, bom desempenho.

1.1 Justificativa

Uma vez que o mercado de ações pode ser previsível, explorar esta possibilidade é o caminho lógico a ser seguido, por pessoas interessadas nesta possibilidade. Para que explorar esta possível previsibilidade, faz sentido analisar a evolução do preço da ação sob um olhar matemático. Devido à quantidade de dados que se pode ter acesso, análises computacionais tem se mostrado promissoras. Ainda assim, existem possibilidades de análises em aberto para diferentes modelos computacionais fazendo com que seja necessário uma grande quantidade de estudos acerca deste assunto.

Por mais que não seja um dos objetivos deste trabalho, ele é escrito sempre procurando uma forma didática de apresentar os conceitos. Assuntos que não são necessários para o desenvolvimento do trabalho não serão discutidos de maneira abrangente, dessa forma é possível desenvolver de um trabalho didático e bem estruturado.

Diante disso, este trabalho tem como sua principal justificativa contribuir com pesquisas a serem desenvolvidas na área de análise de dados financeiros com redes neurais artificiais. Serão discutidas todas as configurações da rede e seus resultados buscando sempre uma maneira didática de apresentação. Muito do desenvolvimento científico nesta área se faz de testes empíricos, então, quando possível serão explorados os conceitos matemáticos que embasam a configuração da rede.

1.2 Objetivos

Os objetivos deste trabalho podem ser divididos entre objetivo geral e objetivos específicos.

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é o desenvolvimento de uma rede neural recorrente do modelo *Long Short Term Memory* (LSTM) para a predição de série temporal financeira, comparando o resultado da rede a modelos teóricos. Serão testados também alterações nos parâmetros da rede procurando explorar o impacto destas alterações no seu desempenho.

1.2.2 Objetivos Específicos

Os objetivos específicos podem ser desenvolvidos da seguinte maneira:

- Estudar e desenvolver modelos de redes neurais recorrentes.
- Estudar teorias do mercado de ações, como hipótese do mercado adaptativo e hipótese do mercado eficiente.
- Entender os processos matemáticos por trás das predições da rede neural.

1.3 Organização do Trabalho

Este trabalho está dividido em Introdução, que foi o tópico que acompanhamos agora, em seguida temos a Fundamentação teórica, que visa nos explicar a base para a execução do trabalho. Metodologia nos ajuda a entender como o trabalho foi desenvolvido, e por fim as Conclusões que nada mais são do que os relatos da execução do trabalho.

2 Desenvolvimento

2.1 Fundamentação teórica

Este capítulo apresenta de uma maneira geral os conceitos fundamentais a serem trabalhados durante a dissertação. Partindo do pressuposto de que, nem sempre tais conceitos são do conhecimento do público em geral, o objetivo desta seção é apresentar e esclarecer as teorias financeiras bem como computacionais que embasam este trabalho.

2.1.1 Introdução ao Mercado financeiro

O mercado financeiro de um país tem por objetivo a negociação de produtos financeiros, sejam eles, ações, commodities, títulos, e demais bens com algum valor financeiro. Em geral este mercado tem um papel importante na economia de seu país, e através deste, pessoas investidoras "emprestam" seu capital a pessoas tomadoras, por meio da compra destes produtos financeiros. O investidor que compra determinado produto, também "compra" a sua promessa de rendimento e busca sempre a melhor forma de se obter isso, balanceando relação entre risco e rentabilidade. O tomador que vende o produto, busca se financiar, talvez pelo fato de para ele a ação já tenha se valorizado o suficiente e não faça mais sentido manter aquele produto, ou no caso do tomador ser a própria empresa buscando crédito fora do mercado bancário.

Este trabalho se desenvolve com base no mercado de ações brasileiro. No Brasil existe atualmente um mercado de ações, que é a B3 S.A. (Bolsa, Brasil e Balcão), as negociações nestes mercados são feitas durante os ditos pregões. Na B3 cada pregão acontece das 10:00h às 18:00h de dias úteis. Uma ação segundo (citar alguém Fulano) pode ser definida como a menor parte de uma empresa de capital aberto. Através dos dados das negociações diárias, o preço de uma determinada ação apresenta alguns indicadores:

- Preço de abertura da ação, como o próprio nome diz é o preço da ação no começo do pregão. Não necessariamente este é igual ao preço de fechamento do dia anterior, por conta de negociações que podem ocorrer depois do pregão, esse período de negociação é chamado de *after market* e acontece para que conciliar os horários de negociação entre a bolsa brasileira e as norte americanas (FINKELSZTAIN, 2022).
- Preço de alta da ação, é o maior valor que a ação tomou durante o pregão daquele dia. Como o preço da ação é variável, não necessariamente o preço de alta será o mesmo do fechamento. Este indicador pode trazer boas análises de tendência do preço da ação.

- Preço de baixa, ao contrário do preço de alta, a baixa é o menor valor que a ação atingiu ao longo do dia. Possui a mesma importância do indicador anterior. Assim como o preço de alta, o de baixa pode ou não ser igual ao de fechamento ou abertura.
- O último indicador de preço de uma ação é o fechamento, este é o último preço que uma ação atingiu durante o pregão.
- Outro indicador diário que uma ação tem é o volume de negociações. Juntamente com os indicadores de alta e baixa este conjunto é importante para analisar a movimentação do mercado e quanto o mercado se interessou na movimentação de preço de determinada ação.

Os primeiros 4 indicadores juntos possibilitam análises de OHLC (*Open, High, Low e Close*), Abertura, Alta, Baixa e Fechamento respectivamente. O fato dos preços das ações serem interpretados como indicadores baseados no tempo, seja, diariamente, semanalmente, ou até mesmo com diferença de minutos, o que possibilita a realização de estudos como objetos de séries temporais, o que será apresentado no decorrer do trabalho.

2.1.1.1 Estratégias básicas de investimento

Uma estratégia de investimento, como o próprio nome sugere, consiste de uma maneira sistematizada de se posicionar no mercado, para a elaboração de uma boa estratégia de investimento, é necessário levar em consideração, a sua própria tolerância a riscos, seus objetivos, quantidade de capital que pretende investir, e o que mais se julgar necessário.

O investidor que opta por comprar e segurar uma ação por tempo indeterminado, segue uma estratégia conhecida no mercado como *buy and hold*, esta estratégia favorece a compra de ativos que o investidor acredita serem perpétuos, pensando na sua valorização a longo prazo. É comum também em estratégias assim, a compra de ações pagadoras de dividendos, geralmente essas ações são de empresas já consolidadas, que em vez de aplicarem o lucro no crescimento da empresa, distribuem aos seus acionistas.

Já o investidor que segue uma estratégia de compra e venda rápida de ações busca o lucro na oscilação de valor da ação, esta estratégia recebe o nome de *trade*. Operações de *trade* podem durar de minutos a alguns dias, um ativo não pode passar muito mais do que isso na carteira de quem opera por esta estratégia.

Para qualquer estratégia de investimento o momento certo de comprar uma determinada ação é crucial. Na estratégia de *buy and hold*, pode parecer que não, afinal ela consiste basicamente em comprar a ação acreditando na sua valorização a longo prazo, mas não é assim. Existem indicadores que possibilitam aos investidores optarem por investir em uma empresa, e no caso desta estratégia, estes indicadores geralmente são de fundamentos das empresas, como balanço, lucro histórico, valor patrimonial, dentre outros, eles são chamados de indicadores fundamentalistas, e é por meio de uma análise destes indicadores

que um investidor desta estratégia define se uma empresa está em um bom momento de compra ou não.

Analisando a estratégia de *trade*, além do momento de compra, o movimento da ação logo a seguir é o que irá determinar se a pessoa investidora terá lucro ou não e para auxiliar na tomada de decisão de movimento do mercado, existem análises e indicadores que auxiliam o investidor.

Levando em consideração a importância do momento de compra e do seguinte, cabe o questionamento se existe a possibilidade de prever, qual o melhor momento e qual a tendência de movimentação do mercado a partir deste momento, ou seja, se o mercado pode ser previsto.

2.1.1.2 Hipóteses de mercado.

Se o mercado for previsível e uma vez devidamente previsto, o investidor poderá obter sempre o maior retorno possível, isso seria a utopia do mercado financeiro. Abordando esta possibilidade, se faz necessário discussão duas teorias principais, a Hipótese de Mercado Eficiente (HME) e a Hipótese de Mercado Adaptativo (HMA).

Segundo Mussa (MUSSA; YANG E.AND TROVÃO; FAMÁ, 2008) a HME surge em meados de 1970 publicada por Eugene F. Fama no *Journal of Finance*, Cambridge. Tal teoria é significativa para o mercado até os dias atuais. A sua proposta é descrever o mercado como eficiente, sendo assim, tendo sempre refletido em seu preço todas as informações referentes a ele.

De acordo com Gomes (GOMES; SOUZA; SILVA, 2019) a teoria da HME trata o ser humano como “*Homo Economics*” e com isso, este tomaria sempre as melhores decisões frente ao mercado financeiro. O que na prática pode nem sempre ser verdade, pois essa forma de pensar ignora no ser humano suas decisões com base na emoção. Em contrapartida (LO, 2004), intersecta a psicologia humana à economia, e propõe a HMA, que tem por objetivo verificar a previsibilidade do mercado por ciclos, ou seja, entre eventos semelhantes, poderia haver a possibilidade do mercado repetir a iniciativa tomada anteriormente, o que garantiria uma certa previsibilidade.

De acordo com (DOURADO G. DE A.; TABAK, 2014) países com a economia ainda em desenvolvimento tendem a possuir um mercado menos eficiente, além disso afirma que apesar de não concluir que o mercado brasileiro segue a tendência de mercado adaptativo, não é possível descartar essa possibilidade.

Até o momento foi possível compreender que o mercado, pode apresentar comportamento previsível e que os dados obtidos pela movimentação deste, podem ser interpretados como séries temporais. Dentre vários estudos estatísticos que envolvem séries temporais, a análise por meio de redes neurais tem tomado espaço dentre os diversos temas de pesquisas atuais.

2.1.2 Aprendizado de Máquina

A capacidade lógica do ser humano é uma das características que o distingue em sociedade. A possibilidade de transmitir essa capacidade cognitiva para outra pessoa já é praticada a muito tempo por meio do sistema educacional, não como o conhecemos hoje, mas de certa forma ainda um sistema educacional. Porém, a possibilidade de algo não humano adquirir tal capacidade cognitiva só existe devido ao desenvolvimento computacional que encontramos atualmente. É possível interpretar o aprendizado de máquina como, o processo de fazer com que a máquina possua a capacidade de tomar decisões para as quais ela não estava previamente programada para o fazer.

O principal campo de estudo do aprendizado de máquina são justamente as formas e processos fazer com que a máquina aprenda. Existem alguns paradigmas que embasam o aprendizado de máquinas, estes paradigmas estão relacionados aos métodos e problemas computacionais que os envolvem. Dentre eles podemos destacar três.

- **Aprendizado supervisionado:** Aqui os dados são passados para a máquina como pares (x_i, y_i) , onde x_i é o dado de entrada e y_i o dado de saída, ou a resposta esperada.
- **Aprendizado não supervisionado:** Já aqui o modelo recebe somente os dados de entrada, aprendendo os padrões do problema e chegando às próprias conclusões.
- **Aprendizado por reforço:** Este paradigma se assemelha um pouco ao anterior, porém ao invés de receber os dados de entrada com as saídas respectivas, ele recebe os dados de entrada, uma saída qualquer e uma medida de precisão quanto àquela saída.

Como dito anteriormente, para problemas computacionais diferentes existem paradigmas diferentes a serem explorados. Os objetivos pretendidos com o modelo e sua forma de construção, dependem destes problemas. E para modelos diferentes existem, algoritmos, indicadores de desempenho e até tratamento de dados diferentes. Sendo assim é importante identificar qual problema computacional está sendo estudado para tratar os dados, bem como avaliar o desempenho do modelo da maneira correta.

Dentre os principais problemas computacionais, cabe destacar:

- **Problemas de regressão:** Basicamente buscam prever um valor quase que “exato”, essa exatidão possui estreita ligação com o quão eficiente o modelo é, por exemplo a previsão de quantos anos um sujeito irá viver com base em dados colhidos ao longo de sua vida.
- **Problemas de classificação:** Como o próprio nome sugere problemas deste tipo buscam classificar uma determinada entrada dentre várias possíveis. Um exemplo são o diagnostico de doenças, que podem indicar tanto positivo quanto negativo.

- Problemas de agrupamento: Problemas deste tipo geralmente estão ligados às semelhanças ou até mesmo à tendência de eventos acontecerem juntos. O exemplo mais comum deste tipo de problema remete aos prateleiras de supermercados, onde com base no que sabem dos hábitos dos consumidores, os donos de mercado organizam as prateleiras para deixarem próximos produtos que sempre são comprados juntos, ou até uma distância proposital afim de expor o consumidor a mais produtos na trajetória de compras dele.

Pode se observar que o aprendizado de máquina auxilia o ser humano para tomada de decisões mais assertivas. Atualmente é imensa a quantidade de dados que é possível manipular, tornando assim difícil para qualquer pessoa analisar estes dados manualmente e tomar boas decisões com base neles. Com isso é importante que pesquisas e estudos se desenvolvam nessa área. O aprendizado de máquina vem de uma necessidade do ser humano e se desenvolve pela interação com este.

O processo de aprendizado do ser humano acontece através das sinapses do cérebro e, intuitivamente, é possível perceber que o computador não tem um. O aprendizado de máquina utiliza-se de uma estrutura conhecida como rede neural artificial (RNA).

2.1.2.1 O que são as redes neurais artificiais.

Uma RNA tem seu funcionamento baseado nos neurônios do cérebro, em outras palavras, o conceito que define a RNA seria a interpretação matemática do funcionamento do cérebro e por meio dela é possível a aplicação do aprendizado de máquina.

Assim como o cérebro humano a rede neural possui neurônios e sinapses e formas semelhantes de transmitir informação. Estas semelhanças são próximas sob uma interpretação prática. Existem diferentes modelos de redes neurais, essa variedade se justifica tanto pelos vários problemas computacionais quanto pela evolução destas. Não cabe ao escopo deste trabalho um detalhamento de diferentes modelos de redes neurais ou da evolução histórica destes estudos, o que se segue é uma apresentação didática do funcionamento básico de uma rede neural com os modelos primitivos e a discriminação do modelo de rede neural *Long Short Term Memory* (LSTM).

2.1.2.1.1 Rede neural simplificada

O modelo simplificado de uma RNA é o *perceptron*, sua arquitetura consiste em uma camada de neurônios de entrada, responsável pelo recebimento de cada informação. A quantidade de neurônios desta camada é equivalente com a quantidade de informações que serão passadas à rede por vez. Além desta, uma camada de neurônios de saída, responsável pela conclusão da rede, esta camada pode conter desde um, a vários neurônios, dependendo do problema computacional que a rede foi desenvolvida para resolver. Todos

os neurônios da camada de entrada são ligados a cada neurônio da camada de saída. Estas ligações recebem o nome de sinapses. O funcionamento desta rede acontece justamente pela capacidade das sinapses atribuírem diferentes pesos para cada ligação, e através do resultado destas atribuições com a soma de um determinado número b , chamado bias, tirarem conclusões a cerca de cada informação. Um conceito matemático semelhante a este processo de atribuição de pesos é a multiplicação de matrizes.

A camada de entrada pode ser interpretada como uma matriz linha $I = (e_{1j})_{1 \times N}$ e os pesos uma matriz coluna $P = (p_{i1})_{N \times 1}$ (cujos dados nós desconhecemos), resultando em uma matriz $R = (r_{11})_{1 \times 1}$.

$$\begin{bmatrix} i_{11} & i_{12} & . & . & i_{1n} \end{bmatrix} \times \begin{bmatrix} p_{11} \\ p_{21} \\ . \\ . \\ p_{n1} \end{bmatrix} + b = \begin{bmatrix} r_{11} \end{bmatrix} \quad (2.1)$$

E de acordo com a multiplicação de matrizes, definido na forma convencional.

$$r_{11} = i_{11} \times p_{11} + i_{12} \times p_{21} + \dots + i_{1n} \times p_{n1} + b. \quad (2.2)$$

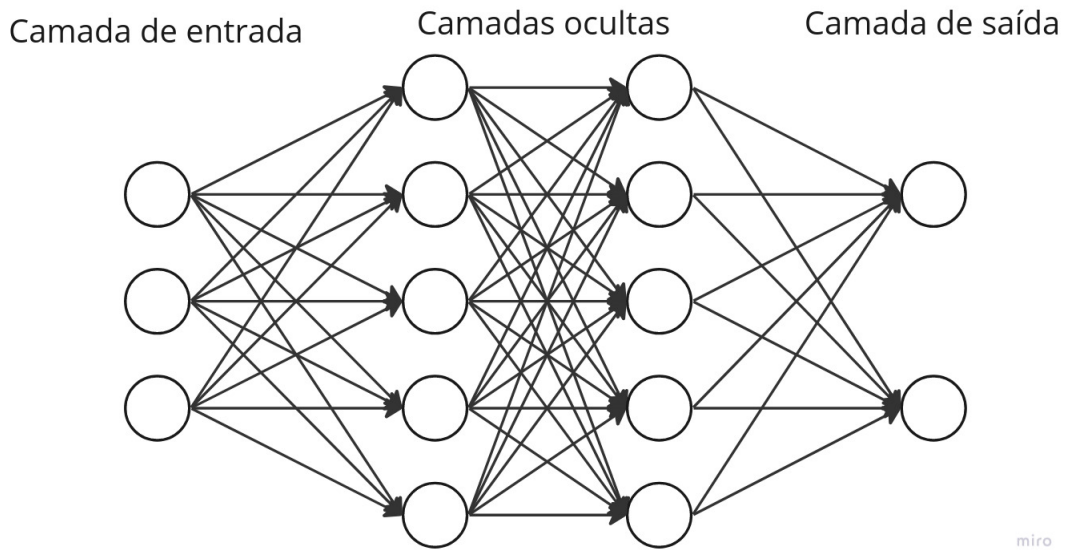
A atribuição de pesos acontece justamente na multiplicação entre um matriz e outra, porém, não é possível constatar que essa operação de fato seja uma multiplicação, uma vez que é desconhecida a forma como a rede atribui seus pesos.

Um exemplo da aplicação deste modelo seria na previsão de preço de um determinado imóvel. Existem fatores que influenciam o preço dos imóveis, como o aquecimento da economia para o setor de construção, a localização, a metragem, dentre vários outros. Para fins didáticos, tome que o valor do imóvel (V) seja influenciado somente pela metragem (M), cor do imóvel (C) e localização (L). Aplicando este exemplo ao modelo teórico discutido até o momento teríamos que:

$$V = M \times p_{11} + C \times p_{21} + L \times p_{31} + b. \quad (2.3)$$

Desta forma a rede poderia encontrar um valor justo para o imóvel sem que uma avaliação dele seja feita.

Caso seja necessário, há a possibilidade de construir a rede com camadas intermediárias, muitas vezes estas camadas recebem o nome de camadas ocultas. As camadas ocultas não possuem um número específico de neurônios, sendo assim podem ser manipulado para melhor desempenho da rede. A função destas camadas é a de auxiliar a distribuição de pesos para cada informação da entrada. A figura 1 apresenta a forma primitiva desta estrutura da rede conhecida como *Multilayer Perceptron*.

Figura 1 – Rede neural artificial do tipo *Multilayer Perceptron*

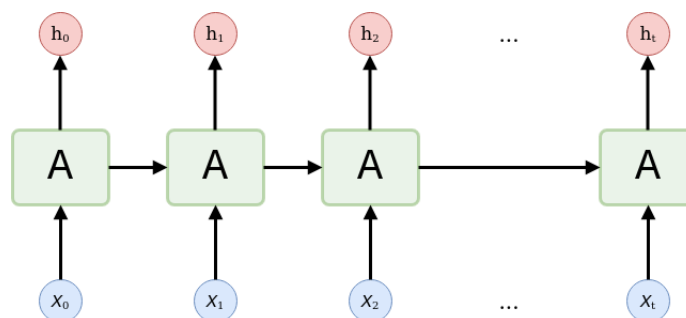
Fonte: autor

Com a evolução dos estudos de redes neurais outros modelos puderam ser desenvolvidas, o que permitiu a criação de redes neurais mais eficazes.

2.1.2.1.2 Redes neurais complexas

Com o desenvolvimento de novas arquiteturas de redes um novo conceito também surgiu, as redes neurais recorrentes RNR. Enquanto no exemplo de rede anterior a informação era tratada linearmente, entrando pela camada de entrada e seguindo as várias camadas ocultas que a rede pode apresentar até a camada de saída. Em uma RNR a informação não necessariamente é tratada de maneira linear, os neurônios da camada de entrada podem receber informação da camada de saída, bem como os neurônios das camadas ocultas podem trocar informação entre si. Arquiteturas assim expõe a rede a mais dados, o que pode implicar na melhor distribuição de pesos para as informações.

Figura 2 – Rede neural recorrente



Fonte: autor

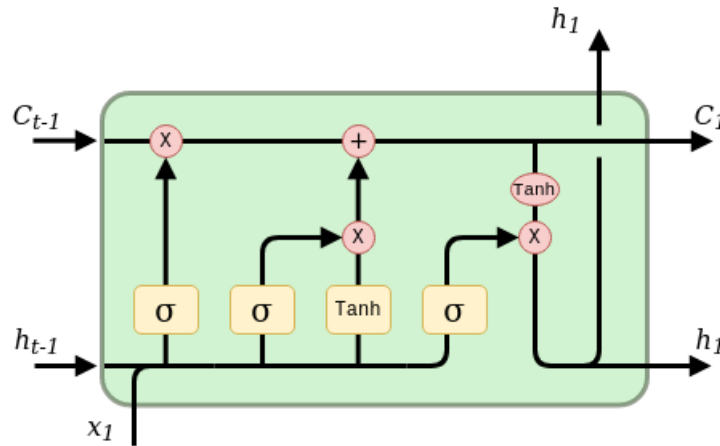
Na RNR a cada interação a informação da camada anterior influencia tanto quanto a informação que entra pela camada de entrada. Devido a isso, a informação de camadas

distantes acaba sendo diluída no decorrer das interações e pode ser que a informação que entra tenha sua relevância diminuída.

Uma RNR que tem sido objeto de estudos atualmente é a rede do tipo memória de curto e longo prazo, popularmente conhecidas como *Long Short-Term Memory* (LSTM) (CAUX; BERNARDINI; VITERBO, 2020), (BULÍN; ŠMÍDL; ŠVEC, 2019), (GARRUTI et al., 2021), (CHENG et al., 2017), (COSTA M.G.; BASSO, 2021), (ZHANG; CHU; SHEN, 2021).

Graças à sua arquitetura, a rede LSTM se diferencia bastante das redes neurais convencionais, não possuindo camadas de neurônios e sim, células de memória. Estas células possuem sinapses assim como os neurônios das redes anteriores, porém tais sinapses não possuem a tarefa de fazer atribuição de pesos e sim transmitir a informação. A distribuição de pesos pela rede LSTM acontece dentro das próprias células de memória, assim como mostra a figura 3.

Figura 3 – Arquitetura de célula do tipo LSTM.



Fonte: adaptado de (ZHANG; CHU; SHEN, 2021)

Pode se observar na figura 3 que a célula de memória LSTM possui elementos dentro dela. Estes elementos possibilitam que ocorram manipulações das informações dentro da célula.

- As setas possuem a função de carregar a informação de forma numeral e concatenando-as aos pesos.
- Os retângulos amarelos são as funções de ativação, possuem o papel de transformar os valores anteriores. Nesta arquitetura é representada as funções sigmoid 2.4 que sempre gera resultados entre 0 e 1 e tangente hiperbólica 2.5 gerando resultados entre -1 e 1 .

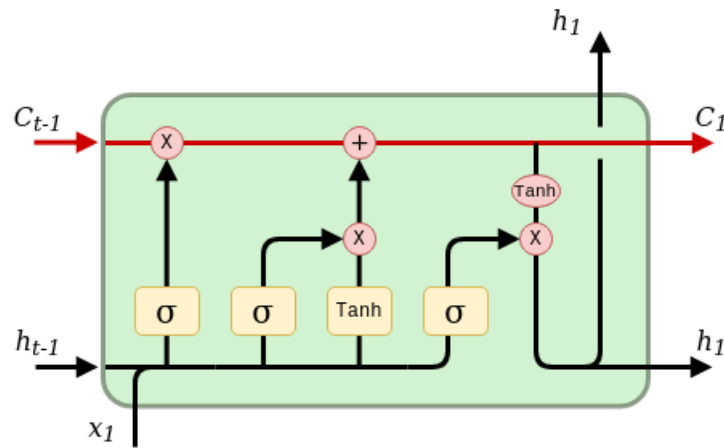
$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2.4)$$

$$\text{tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.5)$$

- As elipses rosas dentro da célula de memória são responsáveis pela concatenação das informações.

Cada célula de memória desta está conectada com a sua anterior e posterior, isso faz com que a informação seja transmitida entre elas. O caminho superior na célula, destacado na figura 4 é o estado da memória, ele é responsável por armazenar todas as conclusões que a rede tira a partir dos dados. O estado da memória de uma célula pode ser passado para a seguinte, se passado isso faz com que a célula atual utilize das conclusões da célula anterior para chegar nas próprias. Podemos observar que o estado da memória está conectado com todos os demais processos da célula, logo cada informação que a célula recebe e devolve influencia e é influenciada por ele.

Figura 4 – Estado da memória da célula LSTM.



Fonte: autor

Na parte inferior esquerda da arquitetura existe entrada de dados, tanto dos novos, quanto os dados de saída da célula anterior e são concatenados. A primeira interação dos dados com a célula acontece no portão de esquecimento, destacado na figura 5, para este portão é atribuído a letra f de *forget*. O portão de esquecimento determina o quanto de informação será perdida do estado da memória anterior, para que assim possa receber novas informações.

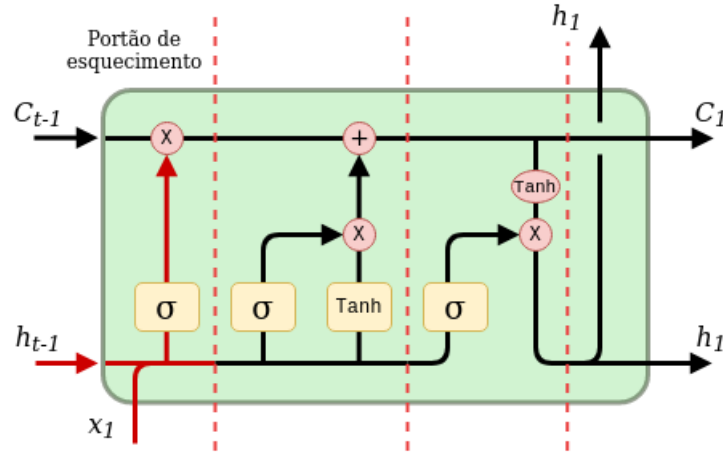
A equação 2.6 descreve o funcionamento do primeiro portão sobre os dados.

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (2.6)$$

W_e é a matriz de pesos do portão de esquecimento, b_e é bias do portão de esquecimento, ele é somado ao resultado a fim de auxiliar no balanceamento. A saída desta equação é um vetor com números entre 0 e 1, este vetor logo é multiplicado pelo estado de memória anterior, fazendo com que a informação seja retirada do estado da memória.

Após o portão de esquecimento tem o de entrada apresentado na figura 6. Este portão possui duas partes, para a primeira é aplicado uma função sigmoid nos dados, esta parte é

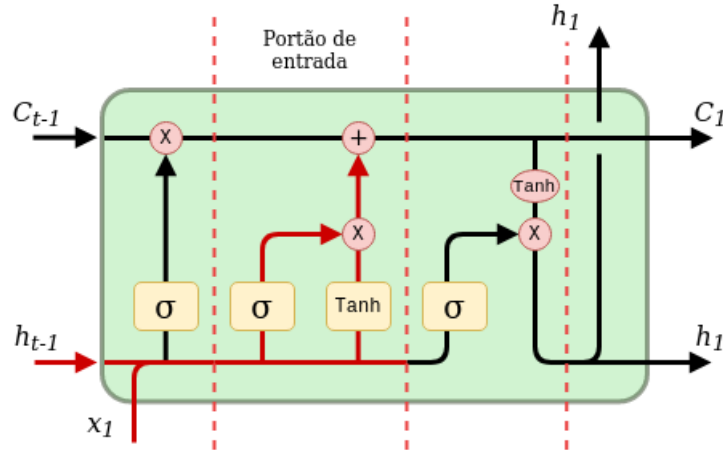
Figura 5 – Entrada de dados e portão de esquecimento.



Fonte: autor

representada pela letra i de *input*, equação 2.10. Na segunda parte é aplicado uma função tangente hiperbólica nos mesmos dados esta parte é representada pela letra \hat{C} por se tratar de um estado de memória parcial, equação 2.8. Neste portão são selecionadas as informações que serão armazenadas no estado da memória e o quanto elas influenciarão o estado.

Figura 6 – Portão de entrada.



Fonte: autor

O funcionamento deste portão é descrito pelas duas equações seguintes.

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (2.7)$$

$$\hat{C}_t = \tanh(W_{\hat{C}} \times [h_{t-1}, x_t] + b_{\hat{C}}) \quad (2.8)$$

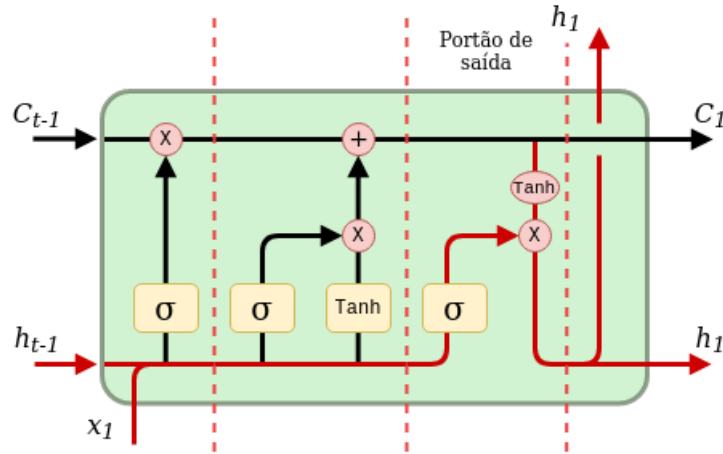
Em ambos os casos W e b representam respectivamente, a matriz de pesos e o bias. As saídas das equações são multiplicada e o vetor resultante deste portão possui números

entre 0 e 1, e em seguida este vetor é concatenado com o estado de memória anterior, formando assim o estado atual da memória, equação 2.9.

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t \quad (2.9)$$

Para concluir a interação com a célula, os dados precisam passar pelo portão de saída, apresentado na figura 7. Neste momento a informação que entrou no começo da célula sofre todas as distribuições de peso conforme o estado da célula atual, gerando assim novos dados a serem passados para a próxima célula ou a conclusão do processo, caso seja a ultima célula.

Figura 7 – Portão de saída.



Fonte: autor

O funcionamento deste portão é descrito pelas duas equações seguintes.

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (2.10)$$

2.1.2.2 O funcionamento da rede neural

Quando a rede neural é iniciada ela precisa passar por um processo de treinamento, visto que os pesos ideais para cada informação são desconhecidos. A princípio é necessário que os dados sejam divididos entre o conjunto de treino e o conjunto de testes. Geralmente entre 60% e 70% são separados para o conjunto de treino, estes dados são expostos à rede para que seja feito o processo de balanceamento dos pesos. Os demais dados são separados para o conjunto de teste, eles são expostos à rede para avaliação do seu desempenho.

A atribuição de pesos da rede neural por meio das suas interações com o conjunto de dados de treino. A parte responsável dentro da RNA por controlar os pesos da rede, é a chamada função de perda, ou *loss*. Um dos métodos mais usados é o erro quadrático médio (*Mean Squared Error* - MSE), equação 2.11.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.11)$$

No caso do método de MSE, a rede faz o somatório do quadrado da diferença entre o cada resultado real y e sua respectiva previsão \hat{y} , e com isso ela conclui uma interação. Após este processo, a rede altera alguns valores de pesos e bias e repete o processo anterior, o ideal é que o valor do MSE convirja para 0, indicando assim que a rede está conseguindo fazer uma melhor distribuição de valores a cada interação.

A quantidade de interações que a rede fará afim de obter a menor taxa de perda é programável, no entanto, uma quantidade excessiva de interações não implica na excelente configuração da rede. Em determinado momento a ela tomará a melhor decisão possível com base no conjunto de dados fornecido, logo, o ajuste de pesos se tornará ineficiente, tornando assim as demais interações um gasto excessivo da capacidade computacional, podendo inclusive prejudicar o aprendizado da rede.

Após estabilização de pesos é preciso avaliar o desempenho da rede, fazendo as previsões e comparando com os dados de teste. Ao final do processo existem três possíveis diagnósticos para o aprendizado da rede, para a conclusão acerca do aprendizado, é necessário observar o decaimento do valor gerado pela função de perda, o valor demonstre o decaimento constante significa que a rede aprendeu com a amostra trabalhada, e quando ela for submetida a novos dados para previsões reais, ela o fará sem qualquer viés.

Já no caso da rede não atingir um bom resultado é possível classificar o problema de duas maneiras, sendo de sub ajustamento (*underfitting*) ou sobre ajustamento (*overfitting*).

- **Underfitting:** Neste caso o modelo não conseguiu aprender com os dados, logo, teremos uma grande taxa de erro tanto no conjunto de treino como no conjunto de teste
- **Overfitting:** Neste caso o modelo viciou nos dados do conjunto de treino, isso quer dizer que, para este conjunto o modelo apresenta excelente predição, porém quando inseridos os dados de teste ou ainda novos dados, não teremos previsão satisfatória.

2.1.3 Série temporal

Uma série temporal o conjunto de observações sobre uma variável ordenados no tempo. Séries temporais estão presentes em vários acontecimentos do mundo real, por exemplo, a quantidade de passageiros de uma determinada companhia aérea medido através dos dias, a venda diária de uma determinada loja e neste caso a variação do preço de ações no mercado de capitais.

Séries temporais possuem algumas características, como, tendência, ciclo e sazonalidade:

- A tendência de uma série indica sua movimentação no longo prazo.

- Os ciclos das séries são os movimentos de alta e de queda durante a sua movimentação de tenência.
- Já a sazonalidade da série está relacionada a ciclos que ocorrem com uma determinada frequência, seja na mesma época do ano, ou em determinado dia.

Existem maneiras estatísticas de analisar séries temporais como mostra (CHU; ZHANG, 2003), porém, há também a possibilidade da predição destas séries por meio das redes neurais como mostra (ANGELO et al., 2011), inclusive estas são objetos de estudos muitos trabalhos atuais e costumam apresentar um bom desempenho em comparação a testes estatísticos.

2.2 Metodologia

Para o desenvolvimento do trabalho utilizou-se da metodologias de pesquisa quantitativa, fazendo uso de coleta de dados, análise e tratamento, bem como a análise dos resultados. De acordo com (NASCIMENTO; CAVALCANTE, 2018) e (MANZATO; SANTOS, 2012) as pesquisas quantitativas usam de linguagem matemática para a descrição e quantificação de fenômenos. Possuem características como a possibilidade de diferentes pesquisadores obterem os mesmos resultados em observações distintas, descrições numéricas que em muitos casos consideradas mais ricas do que descrições verbais e manipulação de dados estatísticos.

Toda a parte de código deste trabalho foi desenvolvida na linguagem de programação Python3 ¹ por meio da ferramenta Colab do Google ². A utilização desta ferramenta permite que o código fique armazenado e execute em nuvem, trazendo maior praticidade.

Nesta seção encontramos o que mais tange à manipulação dos dados analisados bem como as demais ferramentas utilizadas.

2.2.1 Séries analisadas e descrição dos dados

Os conjuntos de dados analisados neste trabalho consiste de dados reais da Bolsa de Valores de São Paulo (B3). Foram extraídos do site Yahoo Finance ³, por meio da biblioteca para linguagem python, yfinance ⁴, as séries temporais referentes aos preços de abertura, alta, baixa, fechamento, fechamento ajustado e volume de negociações das ações de cinco empresas brasileiras. Uma do setor alimentício, Ambev (ABEV3.SA). Duas do setor bancário, Banco do Brasil (BBAS3.SA) e Banco Bradesco (BBDC4.SA). Por fim duas do setor elétrico, Taesa (TAEE11.SA) e Companhia Elétrica de Minas Gerais (CMIG4.SA). O período analisado consiste da janela entre os dias 01 de janeiro de 2008 a 31 de Dezembro de 2022, totalizando assim 15 anos.

A escolha destes ativos se faz pela possibilidade de comparar os resultados com diversos trabalhos desenvolvidos, como (GARRUTI et al., 2022), (COSTA M.G.; BASSO, 2021).

2.2.1.1 Ambev

A tabela 1 apresenta os cinco primeiros dados referentes à evolução do preço desta ação.

¹ <https://www.python.org/>

² <https://colab.research.google.com/>

³ <https://finance.yahoo.com/>

⁴ <https://pypi.org/project/yfinance/>

Tabela 1 – Subconjunto da série ABEV3.SA com os 5 primeiros registros

Data	Abertura	Alta	Baixa	Fechamento	F. Ajustado	Volume
2008-01-02	4.994394	4.994394	4.788625	4.914484	2.891679	387934
2008-01-03	4.914484	4.934461	4.914484	4.932464	2.902258	958574
2008-01-04	4.914484	4.916881	4.787826	4.869734	2.865348	1241391
2008-01-07	4.795018	4.901298	4.754663	4.754663	2.797640	1399068
2008-01-08	4.794618	4.866537	4.735085	4.830178	2.842074	670751

Fonte: autor

2.2.1.2 Banco do Brasil

Na tabela 2 pode ser observado os cinco primeiros dados referentes à evolução do preço desta ação.

Tabela 2 – Subconjunto da série BBAS3.SA com os 5 primeiros registros

Data	Abertura	Alta	Baixa	Fechamento	F. Ajustado	Volume
2008-01-02	30.10	30.820000	29.25	29.799999	12.188092	4728300
2008-01-03	29.51	29.840000	28.85	29.350000	12.004047	5213700
2008-01-04	29.35	29.740000	28.51	29.010000	11.864990	4230800
2008-01-07	29.00	30.230000	28.67	29.700001	12.147195	2734200
2008-01-08	30.00	30.629999	29.60	30.309999	12.396684	3126600

Fonte: autor

2.2.1.3 Banco Bradesco

A tabela 3 apresenta os cinco primeiros dados referentes à evolução do preço desta ação.

Tabela 3 – Subconjunto da série BBDC4.SA com os 5 primeiros registros

Data	Abertura	Alta	Baixa	Fechamento	F. Ajustado	Volume
2008-01-02	23.535587	23.635349	21.989271	22.030840	12.152112	0
2008-01-03	22.113974	22.317656	21.490459	21.577751	11.904346	0
2008-01-04	21.781433	21.968487	21.137135	21.324188	11.764462	5694095
2008-01-07	21.428108	21.993427	21.415638	21.719082	11.982324	0
2008-01-08	22.113974	22.525494	21.931076	22.051622	12.165777	0

Fonte: autor

O volume de negociações não foi passado como variável de interesse, logo não fez diferença para o modelo ele estar zerado em alguns momentos.

2.2.1.4 Taesa

A tabela 4 mostra os cinco primeiros dados referentes à evolução do preço desta ação.

Tabela 4 – Subconjunto da série TAEE11.SA com os 5 primeiros registros

Data	Abertura	Alta	Baixa	Fechamento	F. Ajustado	Volume
2008-01-02	10.333333	10.333333	9.513333	9.513333	2.456057	5700
2008-01-03	10.283333	10.333333	9.866666	10.283333	2.654848	47400
2008-01-04	10.306666	10.666666	10.306666	10.366666	2.676362	9900
2008-01-07	10.333333	10.333333	10.333333	10.333333	2.667757	1800
2008-01-08	10.330000	10.333333	10.330000	10.333333	2.667757	39300

Fonte: autor

2.2.1.5 Cemig

A tabela 5 apresenta os cinco primeiros dados referentes à evolução do preço desta ação.

Tabela 5 – Subconjunto da série CMIG4.SA com os 5 primeiros registros

Data	Abertura	Alta	Baixa	Fechamento	F. Ajustado	Volume
2008-01-02	8.379058	8.379058	8.379058	8.379058	2.635544	0
2008-01-03	8.292983	8.353732	8.227169	8.328439	2.619622	4516910
2008-01-04	8.353732	8.439807	7.989207	8.151225	2.563881	8688236
2008-01-07	8.161355	8.252494	8.052520	8.199343	2.579015	5577984
2008-01-08	8.199343	8.199343	8.199343	8.199343	2.579015	0

Fonte: autor

Todas as amostras contam com um total de 3722 dados não nulos. A autocorrelação das séries objetos foram averiguadas de acordo com o teste de (LJUNG; BOX, 1978), constatando assim a possível previsibilidade da série. Para este teste foi usado a biblioteca Statsmodel ⁵ também para a linguagem python.

2.2.2 Pré processamento dos dados.

O objetivo é a previsão de alta do preço da ação no dia seguinte, para isso foram utilizadas as variáveis exógenas de abertura, baixa e fechamento além da própria série de alta.

Foi estipulada uma janela temporal de três dias para a previsão da alta do quarto dia. Prever a alta, alinhada com uma estratégia de investimento, pode ser mais lucrativo do que prever o fechamento, (GARRUTI et al., 2022) segue esta mesma estratégia.

Foram separados 70% dos dados para treinamento dos modelos, o restante ficou reservado para teste do desempenho dos mesmos. Devido ao modelo ser desenvolvido seguindo uma arquitetura LSTM, é necessário que os dados estejam escalonados. O método utilizado foi o MinMaxScaler e as métricas de avaliação também são da biblioteca sickit

⁵ <https://www.statsmodels.org/stable/index.html>

learn ⁶, o escalonamento retorna sempre um valor entre 0 e 1 e funciona da seguinte maneira, equação 2.12.

$$Escalonado(x_i) = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (2.12)$$

Para a arquitetura dos modelos foi utilizada a biblioteca Tensor Flow ⁷ por meio da interface da biblioteca Keras ⁸.

Ambos os modelos foram construídos utilizando o método *Sequential*, este método permite que arquitetura de empilhamento de redes possam ser construídas, o que permite trabalhar com o conceito de aprendizado profundo, conhecido também como *deep learning*.

Para o primeiro modelo foi construída uma camada de rede neural do tipo LSTM com cem células de memória, seguida por outra camada do tipo LSTM com mais cem células de memória. Por conta desta arquitetura é necessário indicar na construção da primeira camada que retorne toda a sequência de saída não somente a última saída. Após as duas camadas LSTM, foi construída uma camada densa com vinte e cinco neurônios todos conectados entre si e por fim uma camada densa de saída, como o problema computacional trabalhado aqui é de regressão a saída deve conter um único neurônio. A arquitetura construída possui um total de 124.951 parâmetros ajustáveis.

O conjunto de testes para o primeiro modelo contou com um total de 2.602 amostras, cada amostra com três passos no tempo e quatro variáveis.

Para os testes do primeiro modelo, foram utilizadas trezentas épocas de treinamento e amostragem de cem dados, isso quer dizer que o modelo interagiu com os dados de treino trezentas vezes e a cada interação ele recebeu cem dos 2.602 dados. Foram realizadas as previsões e comparadas com dois modelos teóricos, o de passeio aleatório *random walk* (FAMA, 1995), onde a melhor estimativa para o preço de uma ação em um momento t , é o preço da ação no momento $t - 1$ e a média móvel de 7 dias, considerando que o preço da ação no momento t é a média do preço dos 7 dias anteriores a ele.

Para o segundo modelo foi utilizada a mesma arquitetura, porém, o objetivo deste teste é avaliar uma grande dependência temporal da série, isso fez com que alguns ajustes fossem necessários. Primeiramente na arquitetura da rede foram configuradas nas duas camadas LSTM o parâmetro *stateful=True*, isso faz com que as interações anteriores do modelo com a série influenciem as interações atuais e a rede foi configurada para que não faça embaralhamento de amostras, sendo assim, a rede as receberá de maneira linear da primeira para a última. O comprimento da amostra passada para o treinamento do modelo deve ser um número que divide tanto a quantidade de elementos na amostra de treino quanto na de teste, sendo assim o foi definido como dois, no outro modelo era cem.

⁶ <https://scikit-learn.org/stable/index.html>

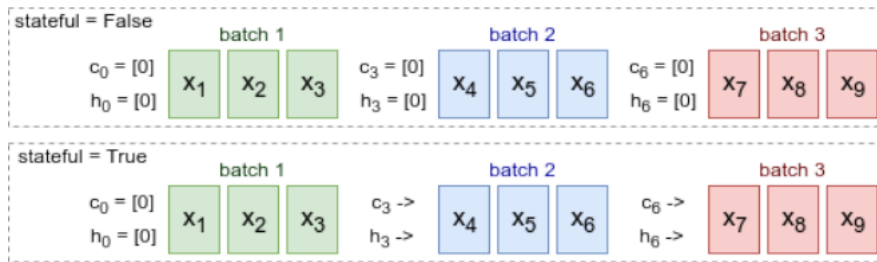
⁷ <https://www.tensorflow.org/?hl=pt-br>

⁸ <https://keras.io/>

Este teste foi adotado como de grande dependência temporal, devido ao fato de que, quando foi definido a janela temporal de três dias, foi considerado que o valor da alta do quarto dia é dependente as quatro variáveis dos três dias anteriores, o que já traz uma certa dependência temporal. Porém, da forma como foram feitos os ajustes para o segundo modelo, entende-se que a dependência temporal da rede acontece desde a primeira interação até a ultima.

A figura 8 exemplifica como o parâmetro *stateful* age sobre a rede.

Figura 8 – Influencia do parâmetro *stateful*.



Fonte: (BULÍN; ŠMÍDL; ŠVEC, 2019)

Devido a necessidade da alteração no comprimento da amostra, foi testado também o desempenho do primeiro modelo com o mesmo tamanho de amostra do segundo e comparados. Abordagens semelhantes podem ser encontradas na literatura (BULÍN; ŠMÍDL; ŠVEC, 2019), acerca da propriedade *stateful*, como neste trabalho, bem como (CHENG et al., 2017), porém aqui o estudo é em torno da técnica de *dropout*. Ao todo para cada série foi possível recolher três resultados de modelos de redes neurais e dois de modelos teóricos.

As funções de ativação das camadas LSTM para todos os modelos foram a tangente hiperbólica, enquanto da camada densa foi linear. Para a otimização do modelo foi utilizado o algoritmo ADAM (KINGMA; BA, 2014) e a função de perda MSE. A quantidade de amostras passada ao modelo simultaneamente, como descrito anteriormente, varia conforme o teste e a quantidade de repetições para adequação dos pesos foram trezentas em todos os testes.

2.2.2.1 Métricas de desempenho

As métricas para avaliação das previsões da rede foram a erro médio absoluto *Mean Absolute Error* (MAE); o erro percentual médio *Mean Absolute Percentage Error* (MAPE); a média do erro quadrado *Mean Squared Error* (MSE), é o mesmo erro avaliado pela função de perda, no entanto aqui ele não possui a função de ajudar na configuração da rede, somente auxiliar na avaliação do desempenho desta; a raiz quadrada do quadrado da média de erro *Root Mean Squared Error* (RMSE); e por fim o erro R quadrado R^2 .

2.2.2.1.1 Erro médio absoluto - MAE

O calculado é dado pela seguinte fórmula:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.13)$$

É utilizado o modulo dos erros, pois há casos em que o erro negativo pode zerar o positivo fazendo com que o valor deste erro seja menor que o real, passando uma ideia de que o modelo é mais preciso do que de fato é. Esta métrica aponta o quão distante está a previsão do valor real, logo, quanto mais próximo de 0 o resultado for, melhor é o modelo, porém, não se pode afirmar a sua eficácia com base em somente uma métrica.

2.2.2.1.2 Erro percentual médio - MAPE

Este cálculo é feito pela seguinte fórmula:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (2.14)$$

Assim como a maioria das métricas quanto mais próximo de 0 significa que melhor foi o aprendizado do modelo. Enquanto a métrica anterior trazia a média da distancia de previsão em valores absolutos, aqui essa distância é representada em valores percentuais.

2.2.2.1.3 Média do erro quadrado - MSE

Já a média do erro quadrado a fórmula:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.15)$$

Semelhante ao MAE 2.13, porem por haver a exponenciação do erro, dá um maior peso aos maiores erros. O uso desta métrica deixa o modelo sensível a valores discrepantes, *outliers*.

2.2.2.1.4 Raiz quadrada da média de erro quadrado - RMSE

A raiz quadrada da média do erro quadrado possui a formula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.16)$$

Assim como o erro anterior MSE, este erro também é sensível a *outliers*, porém a diferença deste com o MSE é que neste o valor retornado está na mesma unidade de medida que o valor estudado, assim como o MAE.

2.2.2.1.5 R quadrado - R^2

A métrica R quadrado é representada equação:

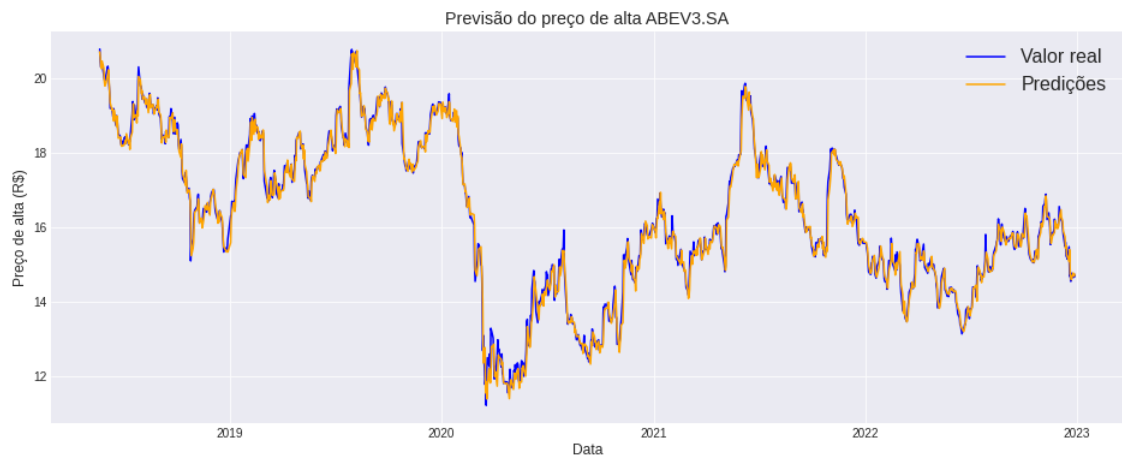
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.17)$$

Nesta fórmula \bar{y} representa a média dos valores reais. Esta métrica descreve o quanto o modelo é confiante e consegue explicar os dados, o resultado geralmente varia entre 0 e 1, porém, há a possibilidade desta métrica retornar um valor negativo.

2.3 Resultados e discussões

Conforme as próximas figuras e quadros apresentados, podemos observar a performance do modelo em relação ao preço real, bem como a comparação às estratégias de passeio aleatório e média móvel para as cinco diferentes ações.

Figura 9 – Comparativo do preço real e previsto pela rede para a ação ABEV3.SA



Fonte: Autor

A figura 9 apresenta a comparação entre os resultados estimados e o valor real do preço de alta para a ação ABEV3.SA, durante o período dos dias 03 de julho de 2018 ao dia 29 de dezembro de 2022, representando 30% dos dados da série temporal desta ação. As previsões são traçadas em amarelo, enquanto o valor real em azul. Podemos observar que a rede conseguiu acompanhar a movimentação da série. A tabela 6 traz as métricas do modelo e nos auxiliará a tirar mais conclusões sobre seu desempenho neste teste.

Tabela 6 – Métricas do conjunto de teste da série ABEV3.SA

Métrica	Modelo	Passeio aleatório	Média móvel 7 dias
MAE:	0.19	0.20	0.30
MAPE:	1.21%	1.24%	1.25%
MSE:	0.08	0.08	0.17
RMSE:	0.28	0.28	0.41
R^2 :	0.98	0.98	0.96

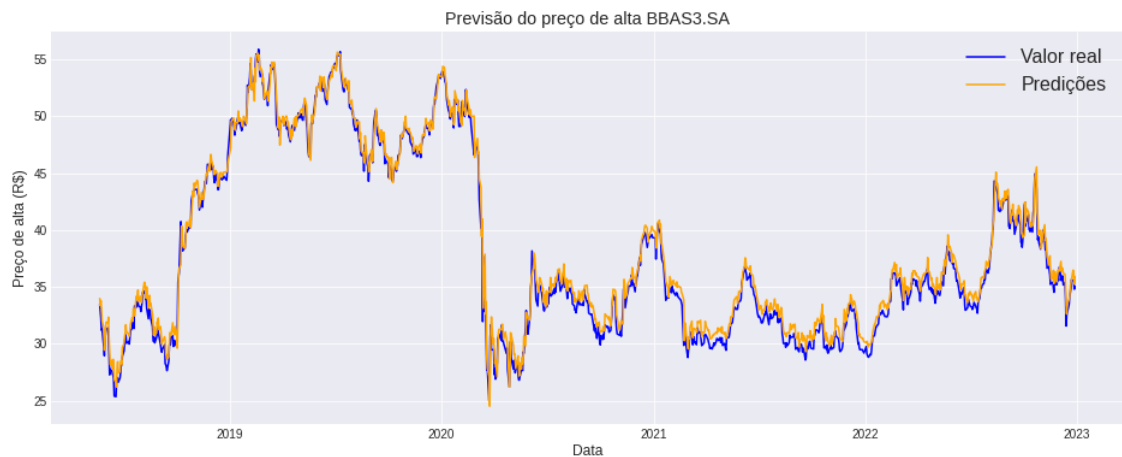
A métrica MAE indica quanto em média o modelo pode errar, considerando valores absolutos, para o modelo esta métrica tem valor de 19 centavos, se mostrando superior aos demais métodos estatísticos. O MAPE também indica a média de erro, porém, em valores percentuais, a superioridade do modelo já era esperada, uma vez que segundo a métrica anterior ele apresentou menor erro. O valor apontado pela MSE não possui um

significado prático como os anteriores, mas, devido ao fato desta métrica elevar a diferença entre os valores reais e previstos ao quadrado faz com que ela seja muito sensível ao erro, a tornando muito importante para análise do desempenho do modelo. Já a RMSE, assim como a MAE, nos indica um valor na mesma unidade do preço da ação, porém, a RMSE é sensível a valores discrepantes. Por ultimo, podemos constatar pela métrica R^2 , que o modelo consegue explicar 98% das previsões.

De acordo com o que observamos, podemos constatar que para esta série o modelo conseguiu melhores resultados que a técnica de passeio passeio aleatório e a média móvel em 7 dias.

Para a próxima análise temos a figura 10 que apresenta a comparação entre os resultados estimados e o valor real do preço de alta para a ação BBAS3.SA. Assim como o anterior o período descrito no gráfico representa 30% dos dados da série e descrevem entre os dias 03 de julho de 2018 e 29 de dezembro de 2022. O gráfico segue a mesma configuração do anterior, sendo que as previsões são traçadas em amarelo, enquanto o valor real em azul. Mais uma vez a rede acompanhou as tendências, porém, podemos observar que gráfico dos valores previstos não sobrepõe o valor real em muitos pontos.

Figura 10 – Comparativo do preço real e previsto pela rede para a ação BBAS3.SA



Fonte: Autor

Tabela 7 – Métricas do conjunto de teste da série BBAS3.SA

Métrica	Modelo	Passeio aleatório	Média móvel 7 dias
MAE:	0.72	0.56	0.86
MAPE:	2.05%	1.50%	2.28%
MSE:	0.86	0.70	1.44
RMSE:	0.91	0.83	1.20
R^2 :	0.99	0.99	0.98

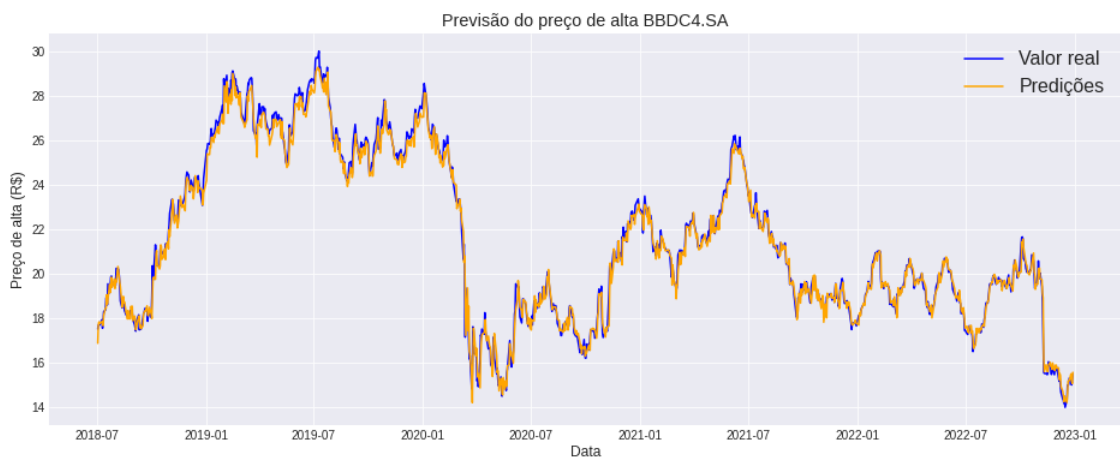
Pela tabela 7 podemos confirmar a suposição que tivemos quando analisamos o gráfico 10, apesar da rede explicar bem o que acontece nas previsões, estas foram distantes dos valores reais. Inclusive para esta série o modelo foi superior a estratégia de média móvel, porém, apresentou resultados inferiores à estratégia de passeio aleatório.

GARRUTI et al. (2021) ao analisar esta mesma ação entre os anos 2013 a 2017 conseguiu melhores resultados com o modelo de LSTM do que os apresentados neste trabalho. Quando observamos seus resultados do modelo MLP, os aqui apresentados se mostram promissores.

O modelo descrito neste trabalho roda com os mesmos hiper-parâmetros para todas as séries. Para conseguirmos melhores resultados, seria preciso que o modelo fosse ajustado para cada uma delas da maneira que melhor as descreva, para isso vários testes de construção de modelo poderiam ser feitos, o que não é o caso deste trabalho.

A figura 11 a apresenta a comparação entre os resultados estimados e o valor real do preço de alta para a ação BBDC4.SA. O período de tempo corresponde novamente a 30% dos dados da série, sendo representados aqui o período entre os dias 03 julho de 2018 a 29 de dezembro de 2022. Para esta série o modelo apresentou maior aproximação entre os valores previstos e reais que para a anterior.

Figura 11 – Comparativo do preço real e previsto pela rede para a ação BBDC4.SA



Fonte: Autor

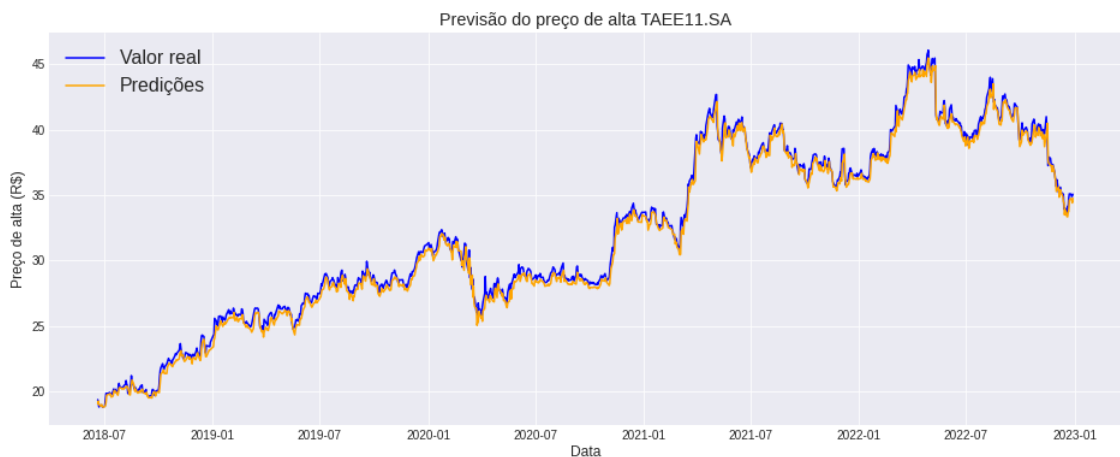
Tabela 8 – Métricas do conjunto de teste da série BBDC4.SA

Métrica	Modelo	Passeio aleatório	Média móvel 7 dias
MAE:	0.26	0.30	0.61
MAPE:	1.27%	1.43%	6.30%
MSE:	0.14	0.17	0.48
RMSE:	0.37	0.42	0.69
R^2 :	0.99	0.99	0.97

De acordo com o que trás a tabela 8, é possível confirmarmos a aproximação entre os valores citados. Ela também nos indica que, o modelo apresentou resultado superiores aos testes.

Observando a figura 12 temos a comparação entre os resultados estimados e o valor real do preço de alta para a ação TAEE11.SA. O período de tempo é o mesmo, sendo entre os dias 3 de julho de 2018 até 29 de dezembro de 2022, correspondendo a 30% dos dados da série. O modelo consegue acompanhar a movimentação da série, no entanto, são poucos os momentos que o valor previsto e o real se coincidem.

Figura 12 – Comparativo do preço real e previsto pela rede para a ação TAEE11.SA



Fonte: Autor

Tabela 9 – Métricas do conjunto de teste da série TAEE11.SA

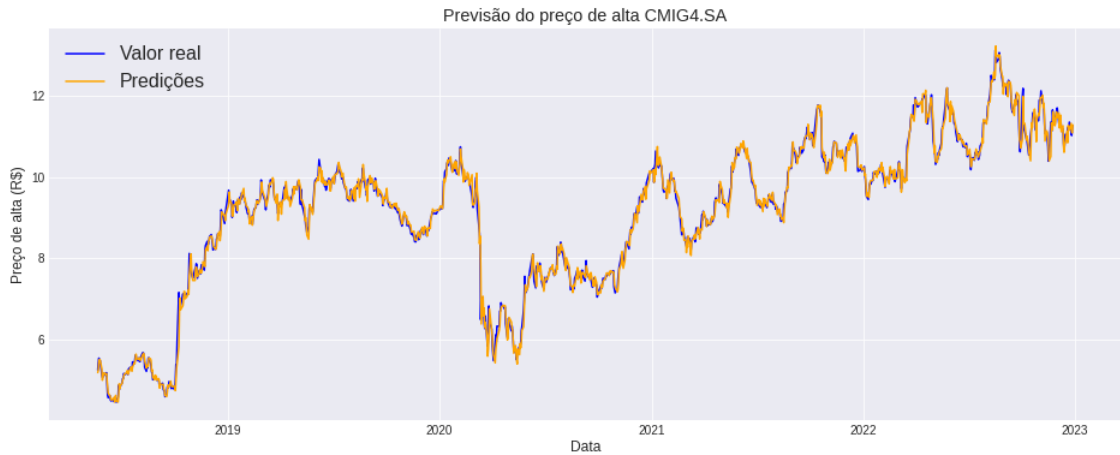
Métrica	Modelo	Passeio aleatório	Média móvel 7 dias
MAE:	0.40	0.30	0.43
MAPE:	1.24%	0.93%	1.32%
MSE:	0.26	0.19	0.35
RMSE:	0.51	0.43	0.59
R^2 :	0.99	1.00	0.99

A tabela 9, traz resultados muito superiores para a estratégia de passeio aleatório, enquanto que o modelo e a estratégia de média móvel apresentam resultados mais próximos.

A figura 13 expressa a comparação entre os resultados estimados e o valor real do preço de alta para a ação CMIG4.SA, o período de tempo é o mesmo das outras séries, 03 de julho de 2018 e 29 de dezembro de 2022, representando 30% dos dados, os períodos coincidem pois todas as séries possuem a mesma quantidade de dados. Mais uma vez o modelo acompanha bem a tendência dos dados e em determinados momentos o gráfico de

valores previstos sobrepõe o de valores reais, indicando assim que neste momento o modelo previu exatamente o valor.

Figura 13 – Comparativo do preço real e previsto pela rede para a ação CMIG4.SA



Fonte: Autor

Tabela 10 – Métricas do conjunto de teste da série CMIG4.SA

Métrica	Modelo	Passeio aleatório	Média móvel 7 dias
MAE:	0.11	0.13	0.20
MAPE:	1.24%	1.49%	2.28%
MSE:	0.03	0.04	0.08
RMSE:	0.17	0.19	0.28
R^2 :	0.99	0.99	0.97

De acordo com as métricas apresentadas na tabela 9, o modelo apresenta resultados melhores que as estratégias para esta série. Comparando estes resultados com (GARRUTI et al., 2022), temos que o modelo deste trabalho apresenta resultados positivos para dois dos modelos construídos no trabalho dele, enquanto que para o terceiro o nosso modelo precisaria de ajustes.

Os resultados discutidos até o momento nos permite criar uma noção do desempenho do modelo. Podemos observar que são raros os momentos onde o modelo acerta exatamente o valor real, criando assim uma distância entre os valores reais e previstos. Por se tratar de um problema de regressão essa é uma situação normal, e são exatamente os indicadores discutidos aqui que possibilitam a elaboração de uma estratégia de investimentos com base neste modelo.

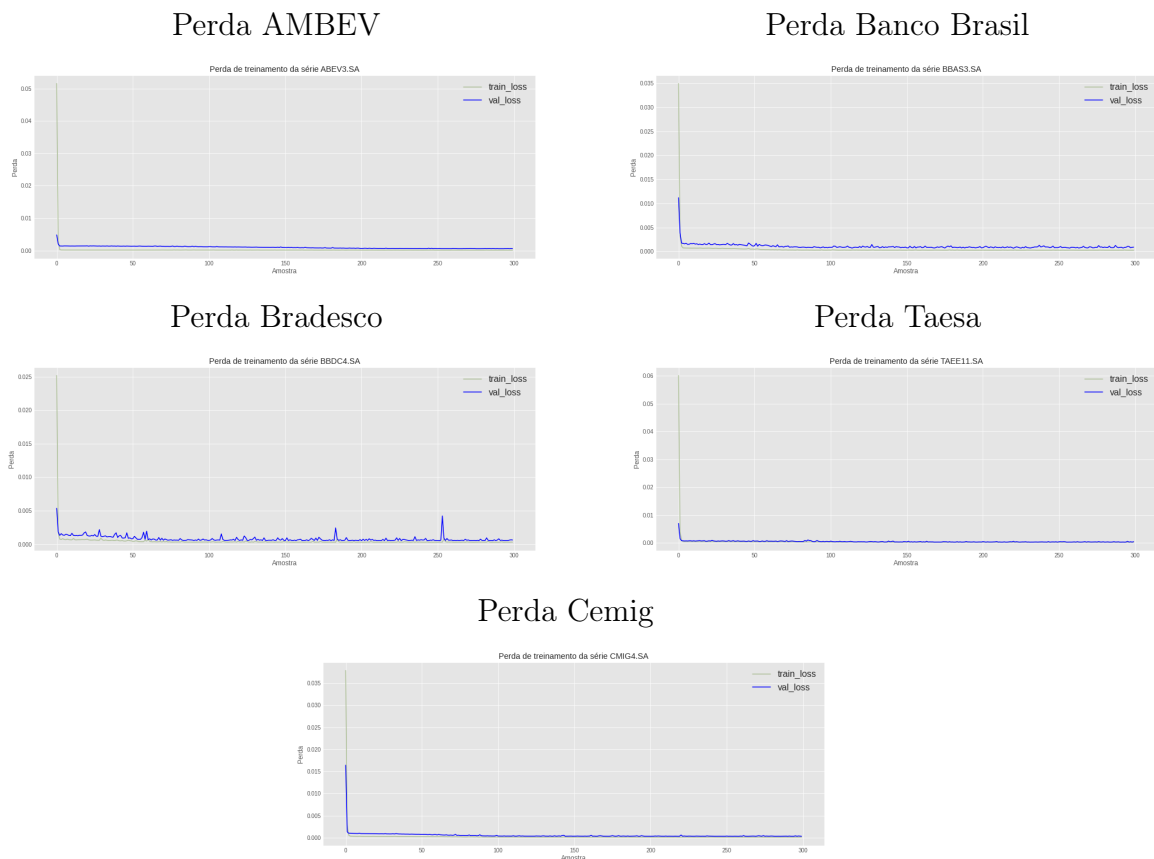
Para melhorar as previsões seria necessário que estuar cada série individualmente e construir um modelo que melhor a descreva, o que claramente não acontece no trabalho,

onde a partir de um único modelo sem nenhuma alteração de hiper parâmetros ou arquitetura, fizemos a previsão de cinco séries diferentes.

Apesar da diferença de resultados citada anteriormente o modelo apresentou um desempenho satisfatório em comparação a muitos outros que vemos na literatura como os modelos apresentados em (COSTA M.G.; BASSO, 2021).

Em nenhum momento nestes testes o modelo apresentou problemas de *underfitting* ou *overfitting*, as imagens a seguir apresentam as curvas de perda do modelo para cada ação.

Figura 14 – Perdas de treinamento das séries



Fonte: Autor

Pelas curvas pode-se notar que o modelo começou apresentando bastante perda, no decorrer das interações entre modelo e amostra, ele se ajustou, indicando que aprendeu com as interações. Quando o modelo aprende não quer dizer que ele seja eficiente, quer dizer que ele é confiante, quando passados novos dados ele trará previsões dentro das métricas apresentadas, a eficiência do modelo são as métricas que estimam.

Quanto ao teste de forte dependência temporal, temos nas tabelas a seguir a comparação entre os modelos desenvolvidos neste trabalho. Sendo o modelo 1 o mesmo dos testes anteriores. O modelo 1-2, o mesmo modelo porém com o tamanho da amostra definido em 2 e o modelo 2, com forte dependência temporal e quantidade de amostras por interação definido em 2.

Para a primeira série temos o resultado das métricas pela tabela 11.

Tabela 11 – Métricas dos modelos para o conjunto de teste da série ABEV3.SA

Métrica	Modelo 1	Modelo 1-2	Modelo 2
MAE:	0.19	0.17	0.81
MAPE:	1.21%	1.09%	5.35%
MSE:	0.08	0.06	0.87
RMSE:	0.28	0.25	0.93
R^2 :	0.98	0.98	0.79

Fonte: Autor

O modelo 1-2 apresentou resultados melhores que o modelo 1, porém como a quantidade de amostragem nas interações foram reduzidas, isso faz com que a rede regule o peso com base na informação dada naquele momento, sendo assim, pode ser que a rede acabe apresentando *overfitting*. Para esta série, o modelo 2 apresentou o pior desempenho, sendo superado pelas duas estratégias apresentadas.

O desempenho dos modelos sobre a série BBAS3.SA é descrita pela tabela de métricas abaixo 12.

Tabela 12 – Métricas dos modelos para o conjunto de teste da série BBAS3.SA

Métrica	Modelo 1	Modelo 1-2	Modelo 2
MAE:	0.72	0.74	1.63
MAPE:	2.05%	2.03%	4.42%
MSE:	0.86	0.83	3.80
RMSE:	0.91	0.91	1.95
R^2 :	0.99	0.99	0.94

Fonte: Autor

Mais uma vez a dependência temporal não ajuda na previsão da série. A diminuição da amostragem por interação parece não impactar de maneira significativa o desempenho do modelo para esta série.

A tabela 13 mostra as métricas de desempenho de todos os modelos para a série BBDC4.SA.

Tabela 13 – Métricas dos modelos para o conjunto de teste da série BBDC4.SA

Métrica	Modelo 1	Modelo 1-2	Modelo 3
MAE:	0.26	0.31	0.80
MAPE:	1.27%	1.57%	3.46%
MSE:	0.14	0.17	1.03
RMSE:	0.37	0.42	1.01
R^2 :	0.99	0.99	0.93

Fonte: Autor

Mais uma vez o primeiro teste trás resultados superiores aos demais.

A tabela 14 nos auxilia a comparar o desempenho dos modelos pelas métricas para a série TAEE11.SA.

Tabela 14 – Métricas dos modelos para o conjunto de teste da série TAEE11.SA

Métrica	Modelo 1	Modelo 1-2	Modelo 3
MAE:	0.40	0.48	1.99
MAPE:	1.24%	1.49%	7.46%
MSE:	0.26	0.34	6.43
RMSE:	0.51	0.58	2.54
R^2 :	0.99	0.99	0.86

Fonte: Autor

Aqui o modelo com grande dependencia temporal tem seu pior desempenho dentre todas as séries.

Por fim temos a tabela 15 apresentando as métricas dos modelos para a série CMIG4.SA.

Tabela 15 – Métricas dos modelos para o conjunto de teste da série CMIG4.SA

Métrica	Modelo 1	Modelo 1-2	Modelo 3
MAE:	0.11	0.12	0.81
MAPE:	1.24%	1.35%	5.35%
MSE:	0.03	0.03	0.87
RMSE:	0.17	0.17	0.93
R^2 :	0.99	0.99	0.79

Fonte: Autor

Diferente do que trás (BULÍN; ŠMÍDL; ŠVEC, 2019) o modelo com o parâmetro *stateful=True* não demonstrou resultados superiores ao modelo com dependência baixa dependência temporal.

3 Conclusão

Neste estudo foi investigado a qualidade preditiva de uma arquitetura de rede neural baseada em células de memória LSTM, sob diferentes parâmetros em suas camadas e na sua interação com os dados, além de comparadas as previsões a modelos teóricos.

Para as séries analisadas, os resultados mostram que a forma como a rede foi estruturada, em geral não apresentou capacidade preditiva superior ao modelo teórico de passeio aleatório, sendo superada por este em dois dos cinco experimentos. Porém quando analisado o desempenho das técnicas de estrutura do modelo, o com baixa dependência temporal se mostrou muito superior.

Existem outros parâmetros a serem analisados, como a utilização da técnica de *droupout*, que faz com que o modelo descarte informações de tempos em tempos, faz sentido estudar a combinação destas técnicas, pois como o modelo foi submetido a uma quantidade grande de dados com informações dependentes a longo prazo, pode ser que nem toda informação seja relevante a fim de ser carregada durante todo o modelo.

No caso do segundo experimento, como a quantidade de dados passados à série por interação eram poucos, o modelo poderia ser ajustado para parar as interações antes de completar as 300, não se pode saber o impacto desta alteração por meio deste trabalho.

Referências

- ANGELO, C. F. de et al. **Séries temporais e redes neurais: uma análise comparativa de técnicas na previsão de vendas do varejo brasileiro.** *BBR-Brazilian Business Review*, FUCAPE Business School, v. 8, n. 2, p. 1–21, 2011.
- BULÍN, M.; ŠMÍDL, L.; ŠVEC, J. **On using stateful LSTM networks for key-phrase detection.** In: SPRINGER. *Text, Speech, and Dialogue: 22nd International Conference, TSD 2019, Ljubljana, Slovenia, September 11–13, 2019, Proceedings 22*. [S.l.], 2019. p. 287–298.
- CAUX, M. de; BERNARDINI, F.; VITERBO, J. **Short-term forecasting in bitcoin time series using LSTM and GRU RNNs.** In: SBC. *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*. [S.l.], 2020. p. 97–104.
- CHENG, G. et al. **An Exploration of Dropout with LSTMs.** In: *Interspeech*. [S.l.: s.n.], 2017. p. 1586–1590.
- CHU, C.-W.; ZHANG, G. P. **A comparative study of linear and nonlinear models for aggregate retail sales forecasting.** *International Journal of production economics*, Elsevier, v. 86, n. 3, p. 217–231, 2003.
- COSTA M.G.; BASSO, L. F. C. **Previsão da Taxa de Câmbio com Rede Neural Recorrente LSTM e Indicadores Técnicos.** In: *XLV Encontro da ANPAD - EnANPAD 2021*. [S.l.: s.n.], 2021.
- DOURADO G. DE A.; TABAK, B. M. **Teste da Hipótese de Mercados Adaptativos para o Brasil.** *Revista Brasileira de Finanças*, v. 12, n. 4, p. 517, 2014. Acessado em: 01/08/2022. Disponível em: <<http://www.spell.org.br/documentos/ver/39858/teste-da-hipotese-de-mercados-adaptativos-para-o-brasil>>.
- FAMA, E. F. **Random walks in stock market prices.** *Financial analysts journal*, Taylor & Francis, v. 51, n. 1, p. 75–80, 1995.
- FINKELSZTAIN, G. *Ofício-Circular*. [S.l.], 2022.
- GARRUTI, D. V. T. et al. **Aplicação de LSTM e MLP para prever preço de ações no Brasil.** In: *EGEN - Encontro de Gestão e Negócios - 2021*. [S.l.: s.n.], 2021.
- GARRUTI, D. V. T. et al. **Qualidade de previsões dos preços máximos e mínimos: uma análise aplicada a ações do setor elétrico.** In: *XXIX SIMPEP*. [S.l.: s.n.], 2022.
- GOMES, L. F. d. A.; SOUZA, P. V. S. d.; SILVA, C. A. T. **Hipótese dos mercados adaptativos: uma análise da eficiência de títulos do mercado acionário brasileiro.** *Revista de Contabilidade do Mestrado em Ciências Contábeis da UERJ*, v. 24, n. 3, p. 25, 2019. Acessado em: 01/08/2022. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/rcmccuerj/article/view/52255/34106>>.
- KINGMA, D. P.; BA, J. Adam: **A method for stochastic optimization.** *arXiv preprint arXiv:1412.6980*, 2014.

- LJUNG, G. M.; BOX, G. E. **On a measure of lack of fit in time series models.** *Biometrika*, Oxford University Press, v. 65, n. 2, p. 297–303, 1978.
- LO, A. W. **The adaptive markets hypothesis.** *The Journal of Portfolio Management*, Institutional Investor Journals Umbrella, v. 30, n. 5, p. 15–29, 2004.
- MANZATO, A. J.; SANTOS, A. B. **A elaboração de questionários na pesquisa quantitativa.** *Departamento de Ciência de Computação e Estatística–IBILCE–UNESP*, v. 17, 2012.
- MUSSA, A.; YANG E.AND TROVÃO, R.; FAMÁ, R. **Hipótese de mercados eficientes e finanças comportamentais - as discussões persistem.** *Revista FACEF pesquisa*, v. 11, n. 1, 2008. Acessado em: 01/08/2022. Disponível em: <<https://periodicos.unifacef.com.br/index.php/facefpesquisa/article/view/109>>.
- NASCIMENTO, L. F. d.; CAVALCANTE, M. M. D. **Abordagem quantitativa na pesquisa em educaçãoInvestigações no cotidiano escolar.** *Revista Tempos e Espaços em Educação*, v. 11, n. 25, 2018. Acessado em: 28/02/2023. Disponível em: <<https://dialnet.unirioja.es/servlet/articulo?codigo=8287842>>.
- ZHANG, Y.; CHU, G.; SHEN, D. **The role of investor attention in predicting stock prices: The long short-term memory networks perspective.** *Finance Research Letters*, Elsevier, v. 38, p. 101484, 2021.