UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DOIS VIZINHOS
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

MURILO ROSSATO FERNANDES

COMPARAÇÃO ENTRE MODELOS LSTM PARA A PREDIÇÃO DO VALOR DAS AÇÕES DA PETROBRAS

TRABALHO DE CONCLUSÃO DE CURSO DE ESPECIALIZAÇÃO

DOIS VIZINHOS
2022

MURILO ROSSATO FERNANDES

COMPARAÇÃO ENTRE MODELOS LSTM PARA A PREDIÇÃO DO VALOR DAS AÇÕES DA PETROBRAS

COMPARISON BETWEEN LSTM MODELS FOR PREDICTION OF PETROBRAS STOCKS

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Me. Francisco Pereira Junior

DOIS VIZINHOS
2022

MURILO ROSSATO FERNANDES

COMPARAÇÃO ENTRE MODELOS LSTM PARA A PREDIÇÃO DO VALOR DAS AÇÕES DA PETROBRAS

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de aprovação: 9 de setembro de 2022

Francisco Pereira Junior
Mestrado
Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

Rosangela de Fátima Pereira Marquesone
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

Henrique Yoshikazu Shishido
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Cornélio Procópio

DOIS VIZINHOS
2022

RESUMO

As predições temporais são utilizadas em diversas áreas, mas em uma delas a acurácia do modelo tem um valor agregado financeiramente direto: o mercado de ações. Este estudo compara dois modelos de LSTM para a predição de valores futuros das ações da Petrobras – um modelo levando em conta apenas a cotação ao longo do tempo (univariável) e outro modelo considerando outras cotações que tem uma correlação forte com a cotação das ações da empresa (modelo multivariável). Os resultados mostraram o modelo univariável tem uma performance melhor, levando em consideração a ação e as características selecionadas (cotação do petróleo, taxa Selic, volume de negociação, preço de abertura e fechamento, máxima e mínima do dia).

Palavras-chave: LSTM; Ações; Petrobras; Mercado Financeiro.

ABSTRACT

Time series predictions are used in several areas, but in one of them the accuracy of the model has a direct financial added value: the stock market. This study compares two LSTM models for predicting future values of Petrobras shares - a model taking into account only the price over time (univariate) and another model considering other quotations that have a strong correlation with the price of Petrobras shares (multivariate model). The results showed that the univariate model has a better performance, taking into account the stock and the selected characteristics (oil price, Selic rate, trading volume, opening and closing price, high and low of the day).

Keywords: LSTM; Stocks; Petrobras; Stock Marketing.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LSTM	*Long Short Term Memory*
Mult	Multiváriavel
RNN	Rede Neural Recorrente
Uni	Univariável

SUMÁRIO

1 INTRODUÇÃO

Quando se trata em prever resultados futuros poucas áreas recebem tanta atenção quanto o mercado financeiro. Responsáveis por movimentar altas somas monetárias, quase todos os seus participantes seguem a lógica básica: acertar o máximo e errar o mínimo.

Se no passado as informações eram restritas e chegavam de forma lenta aos acionistas, hoje a transmissão de informação é instantânea, permitindo tomadas de decisões rápidas e possibilitando análises distintas: análise técnica (com base nos gráficos, médias móveis e etc.), fundamentalista (com base no resultado financeiro da empresa) dentre outras.

Paralelamente à evolução da transmissão de dados do mercado financeiro ocorreu também o desenvolvimento de algoritmos de processamento de grandes volumes de dados com o objetivo de predição de séries temporais, dentre eles o LSTM *(Long Short Term Memory)*.

O LSTM é uma rede neural recorrente muito utilizada para a predição em séries temporais, dentre outros motivos por armazenar valores em intervalos arbitrários. Com base nessas características, o LSTM é usualmente utilizado para predição de séries no mercado financeiro.

A intenção do trabalho não é a recomendação de compra e venda de ações, apenas um comparativo entre os modelos e análise de suas performances frente a um problema de predição de uma ação em bolsa de valores.

1.1 Objetivos

O objetivo geral do trabalho é desenvolver um algoritmo de LSTM para avaliar seu desempenho de predição das ações da Petrobras (PETR4) na B3.

1.1.1 Objetivos Específicos

O objetivo específico é comparar a performance de dois modelos LSTM para a predição das ações da Petrobras: LSTM Univariável (levando em conta valores da cotação apenas) e LSTM Multivariável (considerando valor da cotação, preço do petróleo, juros, Ibovespa e volume).

1.2 Justificativa

As ações da PETR4 deveriam ter forte correlação com todos os atributos selecionados no modelo multivariável. Desta maneira, a ideia é avaliar se o modelo realmente fica mais robusto e com um resultado de performance melhor.

1.3 Materiais e Métodos

Para a realização deste trabalho foram utilizados três datasets (Ibovespa, Brent e Selic) encontrados no site https://www.kaggle.com/. Os dados utilizados são compostos por valores do mercado real.

1 – Dataset Ibovespa – replica os dados do Ibovespa, e é composto por data, *ticker* da ação, preço de abertura, preço de fechamento, ponto mais alto, ponto mais baixo e volume de ações negociadas no dia.

2 – Dataset Selic - composto pela data e valor da taxa Selic.

3 – Dataset Brent – composto pela data e valor do petróleo Brent.

A ferramenta utilizada para manipular os dados foi o Google Colab e a linguagem padrão é o python 3.7.

Além disso, diversas bibliotecas foram importadas, como psycopg2, datetime, pandas, Numpy, math, matplotlib, random, statistics, scikit-learn e collections.

1.4 Organização do Trabalho

A primeira etapa do trabalho (capítulo 2) se dedicará à revisão de literatura, onde serão apresentados elementos que trarão sustentação teóricas ao estudo.

A etapa subsequente é a metodologia (capítulo 3), onde foram seguidas as seis etapas do CRISP-DM que embasaram a execução do estudo.

Por último os resultados foram avaliados em uma conclusão.

2 REVISÃO DE LITERATURA

Esta etapa do trabalho foi dedicada ao entendimento do negócio através da fundamentação teórica dos principais temas que compõem o estudo. Para o aprofundamento teórico necessário, foi feito um detalhamento sobre o mercado financeiro, redes neurais recorrentes, LSTM e métricas de avaliação de um modelo.

2.1 Mercado financeiro

A Bolsa de Valores é uma instituição administradora do mercado onde é permitida a compra e venda de ações. O objetivo principal desta instituição é configurar um ambiente seguro e organizado para essas negociações. Dessa forma, garante que os investidores recebam as ações compradas de maneira eficiente e segura e que as transações sejam rápidas e práticas (TOROINVESTIMENTOS, 2022).

A Petrobras tem suas ações negociadas na Bolsa de Valores brasileira, sendo a segunda maior empresa que compõe o índice do Ibovespa (carteira teórica de ações, com os papéis com maior volume financeiro da Bolsa), tendo um peso de 10% de sua composição (TOROINVESTIMENTOS, 2022). O volume de ações da Petrobras aumento significativamente em 2010, quando a empresa lançou mais 3,75 bilhões de ações no mercado para captar recursos financeiros necessários para a execução do pré-sal (VERSIGNASSI, 2019).

O valor de uma empresa, ou seja, da sua cotação na Bolsa, decorre de expectativas futuras de desempenho financeiro da mesma. A criação de valor por uma empresa é formada pela combinação de diversos fatores e estratégias adotadas pela empresa, como giro dos investimentos, planejamento tributário, margens de lucro e desempenho operacional. De maneira geral, quando há uma tendência de melhores resultados financeiros, o preço da ação tende a subir (NETO, 2020).

A cotação das ações oscila diariamente e a flutuação exata dos preços é imprevisível por natureza. As ações da Petrobras, por exemplo, chegaram a perder 85% do valor entre 2008 e 2016 e subiram 500% entre 2016 e 2019 (VERSIGNASSI, 2019).

2.2 Redes Neurais Recorrentes

Redes neurais recorrentes ou RNNs (do inglês *Recurrent Neural Network*) são uma família de redes neurais para processar dados sequenciais. Assim como as redes neurais são especializadas em processar valores como uma imagem, a rede neural recorrente é especializada em processar uma sequencia de valores x(1), . . . , x() (GOODFELLOW; BENGIO; COURVILLE, 2016).

As RNNs são compostas por muitas redes neurais encadeadas, o que lhes permite processar uma série de dados onde uma rede aprende com suas experiências anteriores. As

RNNs podem ser aplicadas para solicionar diversos desafios, desde escrita a reconhecimento de voz (SELIYA, 2021).

Para simplificar, nos referimos as RNNs como operando em uma sequência que contém vetores x(t) com o índice de passo de tempo t variando de 1 a t . Na prática, as redes recorrentes geralmente operam em mini lotes de tais sequências, com um comprimento de sequência diferente para cada membro do mini lote. Além disso, o índice de passo de tempo não precisa se referir literalmente à passagem do tempo no mundo real, mas apenas à posição na sequência (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.3 LSTM

As RNN podem solucionar diversos tipos de desafios, porém são limitadas a considerar atrás no tempo por aproximadamente 10 períodos de tempo. Isso ocorre por que o sinal de retroalimentação desaparece (WERBOS, 1990). Essa questão foi resolvida com o *Long Short Term Memory* (LSTM). Redes LSTM são até certo ponto biologicamente plausíveis e capazes de aprender mais de 1.000 períodos de tempo, dependendo da complexidade da rede construída (MORRIS, 2019).

A arquitetura original de LSTMs contém unidades especiais chamadas blocos de memória na camada oculta recorrente. Os blocos de memória contêm memória células com auto conexões armazenando (lembrando) o estado temporal da rede, além de unidades multiplicativas especiais chamadas portas para controlar o fluxo de informações. Cada bloco de memória contém uma porta de entrada que controla o fluxo de ativações de entrada na célula de memória e uma porta de saída que controla o fluxo de saída da célula ativações no resto da rede (BEAUFAYS, 2014).

Mais tarde, para resolver uma fraqueza de modelos LSTM impedindo-os de processar entrada contínua fluxos que não são segmentados em subsequências – o que permitiria redefinir os estados das células no início das subsequências – um portão de esquecimento foi adicionado ao bloco de memória. Um portão de esquecimento dimensiona o estado interno da célula antes de adicioná-lo como entrada à célula por meio de conexão auto-recorrente da célula, portanto, esquecendo adaptativamente ou redefinindo a memória da célula (BEAUFAYS, 2014).

Os pesos de viés das portas de entrada e saída são inicializados com valores negativos, e os pesos do portão de esquecimento são inicializados com valores positivos. A partir disso, segue-se que no início do treinamento, a ativação do portão de esquecimento será próxima de '1.0'. A célula de memória se comportará como uma célula de memória LSTM padrão sem um portão de esquecimento. Isso evita que a célula de memória LSTM esqueça, antes de realmente aprender alguma coisa (MORRIS, 2019).

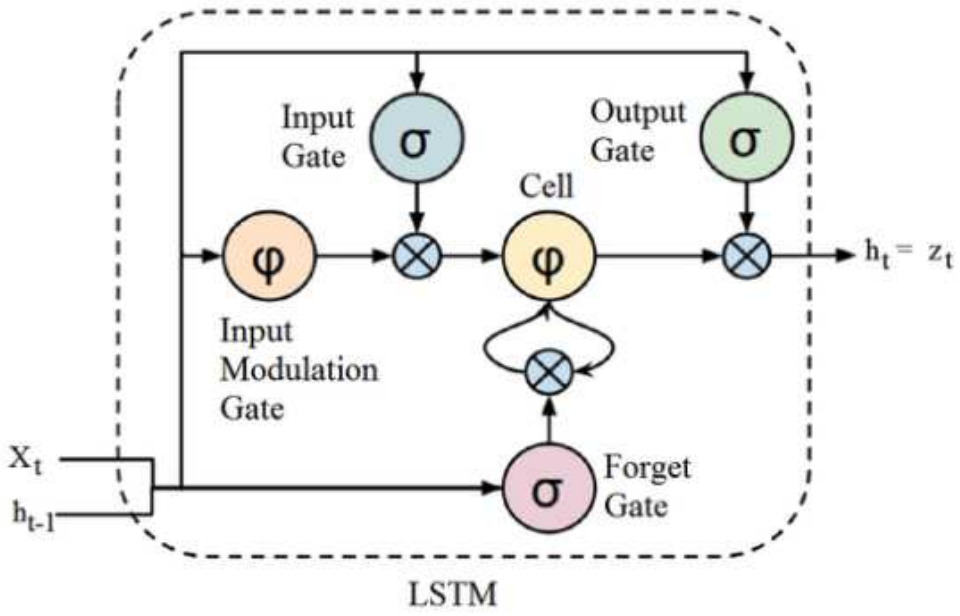
Em resumo do que foi mencionado acima, A LSTM possui uma estrutura em cadeia que contém quatro redes neurais e diferentes blocos de memória chamados células. A informação é retida pelas células e as manipulações de memória são feitas pelos portões (*gates*). Existem três portões:

Portão de esquecimento: remove as informações que não são mais úteis. Duas entradas: x-t (entrada no momento específico) e h-t-1 (saída de célula anterior) são alimentadas ao portão e multiplicadas por matrizes de peso, seguidas pela adição do viés. O resultante é passado por uma função de ativação que fornece uma saída binária. Se para um determinado estado de célula a saída for zero, a informação é esquecida e para a saída um, a informação é retida para uso futuro.

Portão de entrada: a adição de informações úteis ao estado da célula é feita pelo *input gate*. Primeiro, a informação é regulada usando a função sigmoide que filtra os valores a serem lembrados de forma similar ao forget gate usando as entradas h-t-1 e x-t. Então, um vetor é criado usando a função tanh que dá saída de -1 a +1, que contém todos os valores possíveis de h-t-1 e x-t. Os valores do vetor e os valores regulados são multiplicados para obter as informações úteis.

Portão de saída: a tarefa de extrair informações úteis do estado da célula atual para ser apresentadas como uma saída é feita pelo *output gate*. Primeiro, um vetor é gerado aplicando uma função na célula. Então, a informação é regulada usando a função sigmóide que filtra os valores a serem lembrados usando as entradas h-t-1 e x-t. Os valores do vetor e os valores regulados são multiplicados para serem enviados como uma saída e entrada para a próxima célula (DSA, 2022).

Figura 1 – Arquitetura LSTM



Fonte: DSA (2022)

Apesar das vantagens citadas para o LSTM, seu desempenho para problemas de séries temporais nem sempre é satisfatório. Semelhante ao RNN, a arquitetura LSTM pode não representar os recursos complexos de dados sequenciais de forma eficiente, principalmente se

eles forem utilizados para aprender dados de séries temporais de longo intervalo com alta não linearidade (KOTB, 2019).

É essencial saber escolher as características que serão usadas no LSTM para predição do mercado financeiro. No trabalho apresentado por (CHEN, 2015), houve um aumento na acurácia da predição de 14,3% para 27,2% quando o autor selecionou cinco características diferentes de uma ação no mercado chinês, ao invés de apenas uma.

2.4 Métricas de Avaliação

O objetivo da fase de avaliação é estimar os resultados do modelo de forma rigorosa e obter a confiança de que são válidos e confiáveis antes de avançar. Igualmente importante, a fase de avaliação também serve para ajudar a garantir que o modelo satisfaça os objetivos de negócios originais (FAWCETT, 2016).

O Erro Médio Absoluto (MAE, do inglês *Mean Absolut Error*) é uma métrica de avaliação de modelo usada com modelos de regressão. O erro absoluto médio de um modelo em relação a um conjunto de teste é a média dos valores absolutos dos erros de previsão individuais em todas as instâncias do conjunto de teste. Cada erro de previsão é a diferença entre o valor verdadeiro e o valor previsto para a instância (G.I., 2011).

Figura 2 – Fórmula MAE

$$mae = \frac{\sum_{i=1}^{n} abs\left(y_i - \lambda(x_i)\right)}{n}$$

Fonte: G.I. (2011)

O erro quadrático médio de um modelo em relação a um conjunto de teste é a média dos erros quadráticos de previsão em todas as instâncias no conjunto de teste. O erro de previsão é a diferença entre o valor verdadeiro e o valor previsto para uma instância (G.I., 2011).

Figura 3 – Fórmula MSE

$$mae = \frac{\sum_{i=1}^{n} abs\left(y_i - \lambda(x_i)\right)}{n}$$

Fonte: G.I. (2011)

2.5 Métodologia CRISP-DM

A métodologia CRISP-DM foi desenvolvida por meio do esforço de um consórcio composto inicialmente com DaimlerChryrler, SPSS e NCR. CRISP-DM significa *CRoss-Industry Standard Process for Data Mining*. Consiste em um ciclo que compreende seis etapas (CLINTON, 2000):

1 - Entendimento do negócio: esta fase inicial concentra-se em entender os objetivos do projeto e requisitos de uma perspectiva de negócios, convertendo esse conhecimento em uma definição do problema de mineração de dados e um plano preliminar projetado para atingir os objetivos;

2 - Entendimento dos dados: esta etapa começa com a coleta inicial dos dados e familiarização com os mesmos, identificando potenciais problemas de qualidade e adquirindo os primeiros *insights* para formação de hipóteses;

3 - Preparação dos dados: esta etapa cobre toda a preparação dos dados, desde os dados brutos até o dataset final. A preparação dos dados incluí a limpeza dos dados, tratamento de dados faltantes, remoção de informações duplicadas dentre outras atividades;

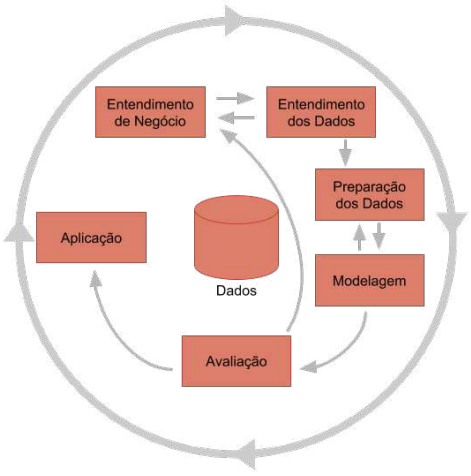
4- Modelagem: técnicas de modelagem podem ser aplicadas para calibrar os parâmetros para valores ótimos;

5 – Avaliação: nesta fase o modelo (ou modelos) obtido são avaliados mais detalhadamente e as etapas executadas para construir o modelo são revisadas para garantir que ele atinja adequadamente o objetivo do negócio;

6 – Implantação: Esta etapa não significa necessariamente o final do projeto. Após essa fase o conhecimento sobre os dados é aumentado e é necessário avaliar se há uma próxima etapa na qual o ciclo se reiniciaria novamente;

A Figura 4 abaixo demonstra de forma visual o ciclo completo da metodoliga CRISP-DM.

Figura 4 – Ciclo CRISP-DM



Fonte: Fonte: https://medium.com/@kvmoura/crisp-dm-79580b0d3ac4

3 DESENVOLVIMENTO E RESULTADOS

Para a execução do trabalho optou-se pela aplicação da metodologia CRISP-DM, que busca a transformação de dados em conhecimento e pode ser aplicada em diferentes contextos de análise de dados.

3.1 Entendimento do negócio

Esta etapa do processo foi cumprida no capítulo 2 do trabalho, durante a revisão de literatura. Durante este processo, pode-se compreender como funciona o mercado acionário, as redes neurais recorrentes, o LSTM, a metodologia CRISP-DM e as métricas de avaliação para algoritmos de predição.

3.2 Entendimento dos dados

Nesta etapa foram analisados os dados que compõe o estudo e o motivo pelo qual são relevantes para a análise.

Dados Ibovespa: o *dataset* do Ibovespa é composto por data (dia do pregão), ticker da ação (abreviação do nome da empresa), preço de abertura, preço de fechamento, cotação mais alta do dia, cotação mais baixa e volume de ações negociadas na data. O *dataset* contém informações de um longo período, porém as análises foram feitas com dados à partir de 2016, totalizando uma amostra de 1709 registros. O mesmo critério foi adotado para os outros *datasets* utilizados. Dentro destes dados foram filtradas apenas ações da Petrobras (PETR4), e o resultado pode ser observado na Figura 5.

Figura 5 – *Dataset* Ibovespa

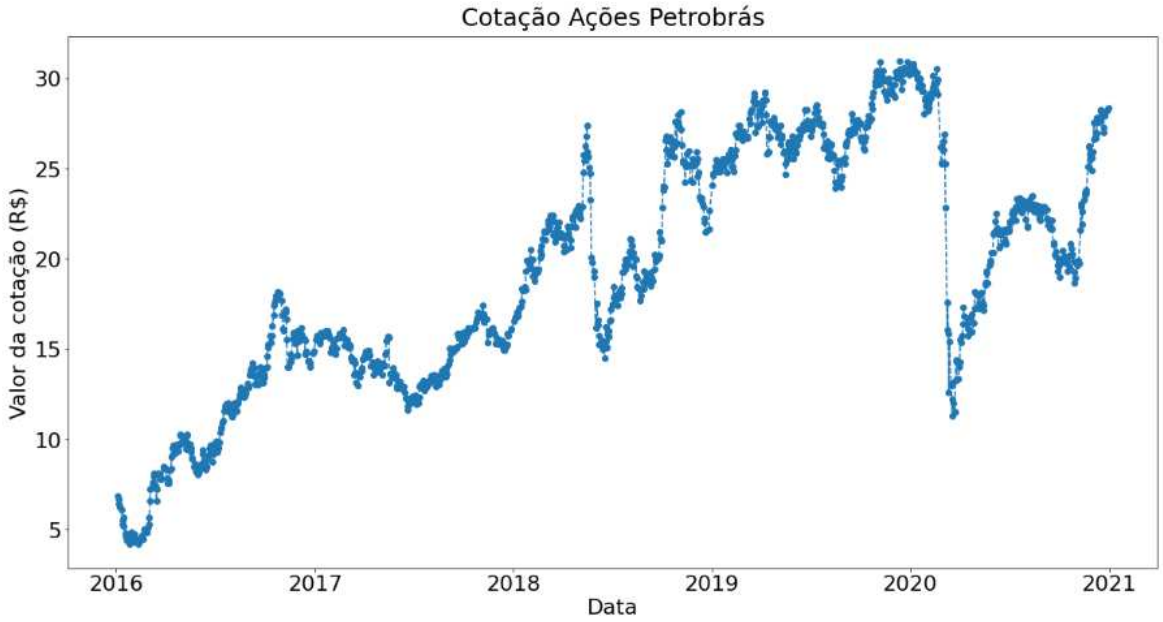
	datetime	ticker	open	close	high	low	volume
0	1998-03-16	PETR4	276.0	275.0	276.0	273.99	27078933.9
1	1998-03-17	PETR4	275.0	277.0	278.0	274.50	41049029.3
2	1998-03-18	PETR4	277.0	275.5	279.0	275.00	40506405.9
3	1998-03-19	PETR4	275.5	275.5	277.0	272.00	29801256.4
4	1998-03-20	PETR4	277.0	279.0	280.0	276.00	26713421.3

Fonte: Autoria Própria (2022).

Nota-se em uma primeira avaliação na Figura 6, que o valor da cotação das ações se comporta de maneira pouco previsível e com grandes oscilações.

Dados Selic: a taxa Selic é a taxa básica de juros da economia brasileira, definida pelo Banco Central em reunião que ocorre a cada 45 dias, e serve como referência para empréstimos

Figura 6 – Gráfico de cotação das ações da Petrobras



Fonte: Autoria Própria (2022).

e financiamentos. Ao contrário do gráfico de ações, a taxa Selic tem um comportamento mais estável, e como se pode notar na Figura 7, não apresenta aumentos e quedas repentinas de uma vez.

Essa taxa tende a influenciar o preço das ações, pois a maioria das empresas tem dívidas, e essas dívidas são em parte atreladas à Selic. Isso quer dizer que, quanto maior a taxa de juros, maior o custo da dívida, menor o lucro da empresa e consequentemente, uma tendência de menor valor em sua cotação.

No caso das ações da Petrobras, observou-se uma correlação negativa entre taxa Selic e preço da ação de -0,73 durante o período análisado (dados desde 2016).

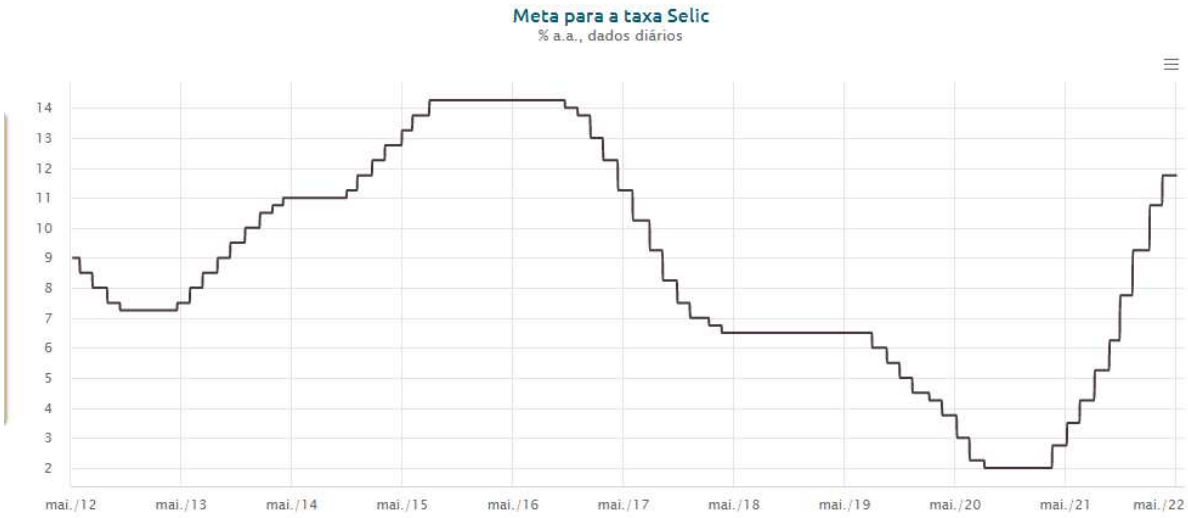
Dados *Brent*: sabendo que o principal negócio da Petrobras é a extração e venda do petróleo, é natural que o seu lucro, e consequentemente o valor de sua ação, tenha uma correlação com o valor do barril de petróleo – neste caso o Brent, a cotação mais comum quando se fala do preço de petróleo no mundo.

O valor do barril se comporta de maneira mais imprevisível (de acordo com a Figura 8), assim como o valor de uma ação, pois depende fortemente de oferta e procura. Nota-se no que no período inicial da COVID-19 o preço caiu significativamente, refletindo uma preocupação com a demanda baixa devido às operações de *lockdown*. Essa cotação apresentou uma correlação de 0,53 com a ação da Petrobras.

Correlação dos dados: para que a ideia da análise multivariável seja coerente, é importante que os dados tenham uma correlação considerável, de maneira que uma rede neural recorrente possa extrair aprendizado com base em todas as características. Nota-se na Figura 9 uma correlação acima de 0,5 para todas as variáveis, incluindo também uma correlação negativa (Selic), que significa que os dados tendem a se comportar de maneira oposta, ou seja,

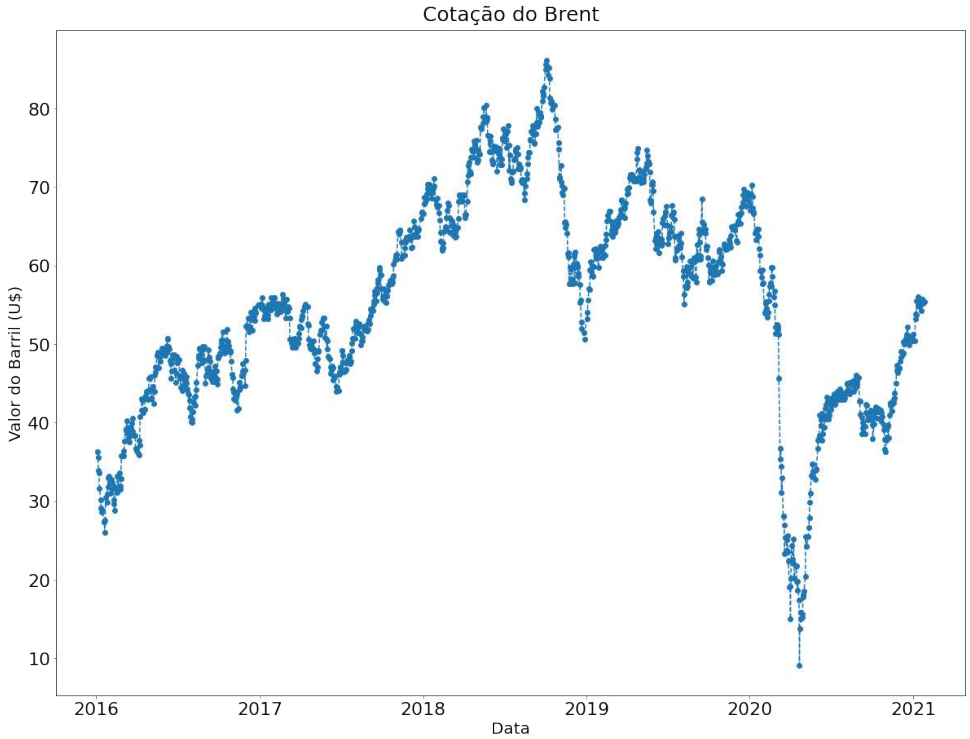
Fonte: Autoria Própria (2022).

Figura 7 – Gráfico taxa Selic



Fonte: https://www.bcb.gov.br/estatisticas/grafico/graficoestatistica/metaselic

Figura 8 – Cotação do Brent



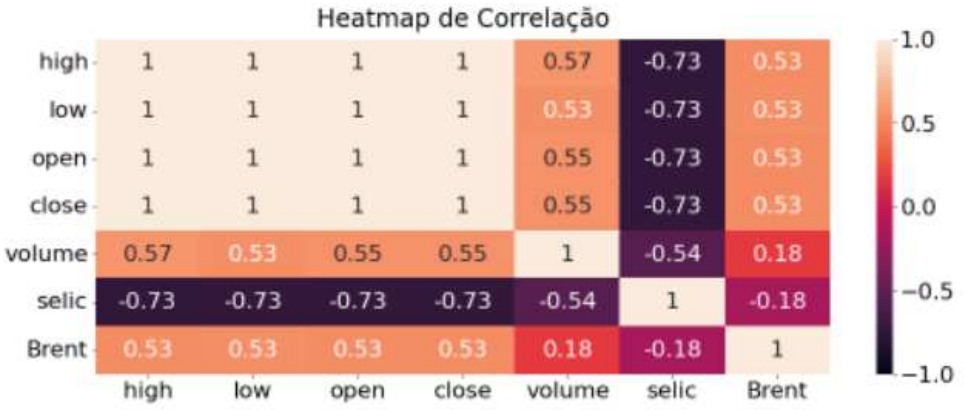
Fonte: Autoria Própria (2022).

quando a Selic cai, o preço da ação tende a subir.

3.3 Preparação dos dados

Inicialmente o trabalho foi dividido em duas etapas – primeiro a preparação de dados para o LSTM univariável e posteriormente para o multivariável. Os dois dados receberam um pré-processamento semelhante, que incluíram o dimensionamento dos dados e a divisão entre

Figura 9 – Correlação dos dados



Fonte: Autoria Própria (2022).

dados de treino e de teste.

Separação dos dados: através dos *datasets* iniciais, foram criados dois *dataframes*, o primeiro univariável, com os dados apenas da data e valor da cotação da ação, e um segundo, multivariável, com as variáveis mostradas nas Figuras 10 e 11.

Figura 10 – *Dataframe* univariável

	datetime	close
4319	2016-01-04	6.87
4320	2016-01-05	6.68
4321	2016-01-06	6.40
4322	2016-01-07	6.26
4323	2016-01-08	6.27

Fonte: Autoria Própria (2022).

Figura 11 – *Dataframe* multivariável

	datetime	open	close	high	low	volume	selic	Brent
4319	2016-01-04	6.57	6.87	7.03	6.55	314449892.0	0.000525	36.28
4320	2016-01-05	6.92	6.68	7.00	6.62	198582402.0	0.000525	35.56
4321	2016-01-06	6.53	6.40	6.54	6.40	434823901.0	0.000525	33.89
4322	2016-01-07	6.19	6.26	6.43	6.08	359180016.0	0.000525	33.57
4323	2016-01-08	6.38	6.27	6.45	6.13	327367768.0	0.000525	31.67

Fonte: Autoria Própria (2022).

Normalização dos dados: para que os métodos de predição tenham uma boa acurácia, é necessário que nenhuma característica tenha um peso muito maior do que outra. No trabalho proposto, por exemplo, a quantidade de ações negociadas de um dia para o outro, tem variações na ordem de milhões com relação ao volume negociado, enquanto a cotação da ação tem uma variação na média em menos de R$1,00.

Desta maneira, todos os dados foram pré-processados para que suas escalas fossem reduzidas, de maneira que estivessem entre um valor de 0 a 1, tornando os valores normalizados. Para aplicar essa técnica foi utilizada a função MinMaxScaler, do scikit-learn.

Divisão de dados treino e teste: além disso, os dados foram divididos entre treino e teste, em uma proporção de 75% dos dados para treino e os outros 25% para teste. O LSTM requer também que os dados sejam separados em janelas de predição, desta forma o algoritmo lê os dados da janela x e tenta prever o valor de $x + 1$. O intervalo escolhido foi de 60 dias.

3.4 Modelagem dos dados

Com os dados divididos e tratados foi criada a rede neural recorrentes de sete camadas: Três LSTM com 50 neurônios recorrentes, três camadas de *Dropout* e uma camada responsável por compilar os dados em uma só saída. O modelo pode ser observado na Figura 12, onde visualiza-se três chamadas ao método *Dropout* e uma de compilação (*Dense*).

Figura 12 – Modelagem rede LSTM

```
model=Sequential()
model.add(LSTM(50,return_sequences=True,input_shape=(60,1)))
model.add(Dropout(rate = 0.2))
model.add(LSTM(50,return_sequences=True))
model.add(Dropout(rate = 0.2))
model.add(LSTM(50))
model.add(Dropout(rate = 0.2))
model.add(Dense(1))
model.compile(loss='mean_squared_error',optimizer='adam')
```

```
model.summary()
```

```
Model: "sequential_7"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_21 (LSTM)              (None, 60, 50)            10400

 dropout_18 (Dropout)        (None, 60, 50)            0

 lstm_22 (LSTM)              (None, 60, 50)            20200

 dropout_19 (Dropout)        (None, 60, 50)            0

 lstm_23 (LSTM)              (None, 50)                20200

 dropout_20 (Dropout)        (None, 50)                0

 dense_7 (Dense)             (None, 1)                 51

=================================================================
Total params: 50,851
Trainable params: 50,851
Non-trainable params: 0
_________________________________________________________________
```

Fonte: Autoria Própria (2022).

O modelo foi executado por cem épocas e que utilizou como otimizador o Adam, que

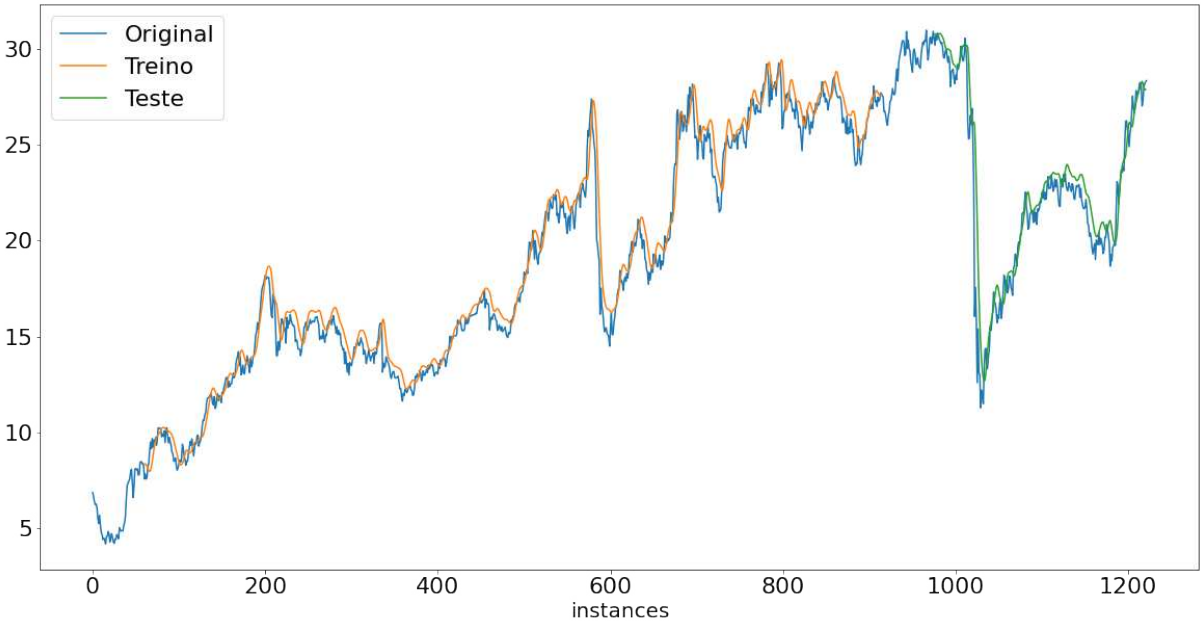
é um método estocástico de gradiente descendente que se baseia na estimativa adaptativa de momentos de primeira e segunda ordem.

Os dois grupos de dados (uni e multivariáveis) foram submetidos à mesma rede neural, com o objetivo de manter a semelhança do processo para melhor aferir os resultados.

3.5 Avaliação de resultados

O modelo se comportou conforme o esperado nos dois casos, seguindo de perto os valores reais do Ibovespa, com uma melhor performance do *dataset* univariável (demonstrado na Figura 13).

Figura 13 – Resultado do LSTM aplicado no *dataset* univariável



Fonte: Autoria Própria (2022).

Embora na Figura 13 os resultados pareçam seguir com precisão os valores reais da cotação, o modelo apresentou uma MAE = 0,88, o que na prática não seria um modelo confiável para utilização em mercado financeiro, pois os valores estariam significativamente longe dos valores reais da ação.

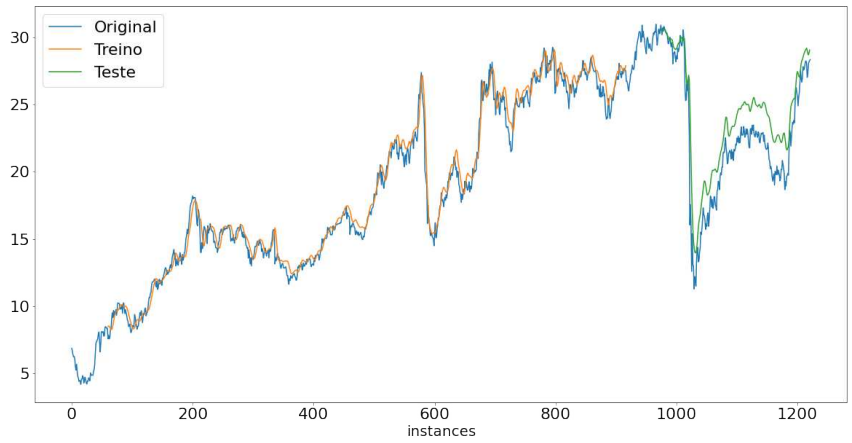
Na Figura 14 observa-se que os resultados do modelo multivariável também seguiram os padrões do resultado reais da ação da Petrobras, porém com uma diferença maior para o real, no caso um MAE = 1,89.

As diferenças entre os resultados dos dois modelos podem ser visualizadas na Tabela 1:

3.6 Execução

Na Tabela 1 nota-se que os dois métodos adotados para aferir a acurácia do modelo resultaram em um valor menor para o modelo UniVar, ou seja, o erro da predição foi menor

Figura 14 – Resultado do LSTM aplicado no dataset multivariável



Fonte: Autoria Própria (2022).

Tabela 1 – Resultado dos testes.

	MAE	MSE 2
UniVar	0,88	2,19
MultVar	1,89	5,57

Fonte: Autoria Própria (2022).

neste modelo.

Para garantir que nenhum hiperparâmetro (definições gerais do modelo, como tamanho da amostra e quantidade de *dropout* aplicados) específico favorecesse um dos LSTM, alguns foram alterados e os resultados constam na Figura 14:

A primeira parte da Figura 15, onde são comparados os resultados de MAE podem ser observadas visualmente no gráfico mostrado na Figura 16. Os resultados demonstram que não importa o hiperparâmetro alterado, o MAE do LSTM aplicado sobre os univariados são menores, ou seja, com um resultado mais próximo do real. Nota-se isso ao observar na Figura 15 que em todas as linhas o resulta de Uni são menores que os resultados de Mult.

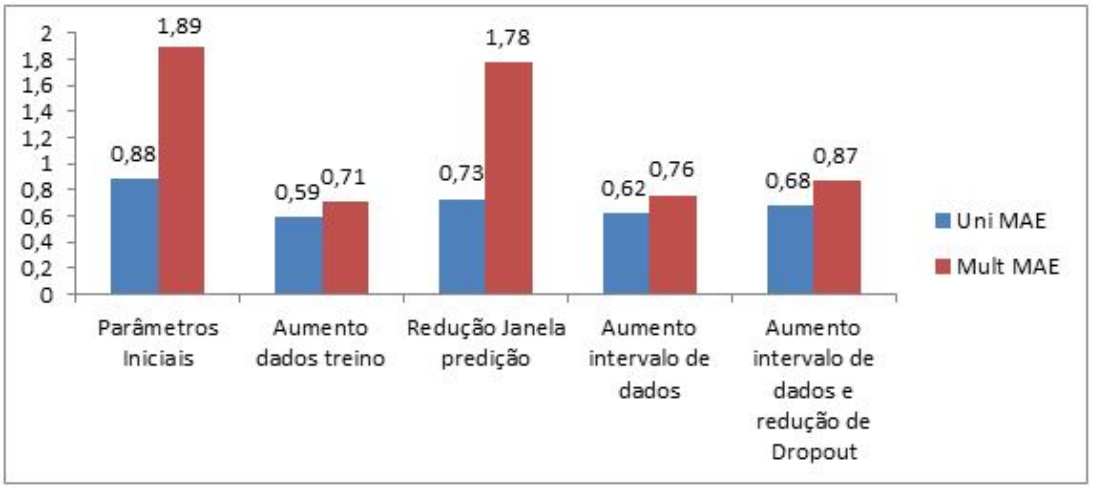
Os resultados de MSE também seguiram o mesmo padrão, conforme Figura 17.

Figura 15 – Resultados alterando hiperparâmetros

	Parâmetros	Uni MAE	Mult MAE	Uni MSE	Mult MSE
Parâmetros iniciais	Rede Neural com 3 *Dropout* 75% Treino Janela de previsão 60 dias Dados desde 2016	0,88	1,89	2,19	5,57
Aumento dos dados de treino	Rede Neural com 3 *Dropout* **85% Treino** Janela de previsão 60 dias Dados desde 2016	0,59	0,71	0,62	0,88
Redução no tamanho da sequência	Rede Neural com 3 *Dropout* 85% Treino **Janela de previsão 30 dias** Dados desde 2016	0,73	1,78	1,59	4,89
Aumento do intervalo de dados	Rede Neural com 3 *Dropout* 75% Treino Janela de previsão 60 dias **Dados desde 2014**	0,62	0,76	0,89	1,1
Aumento do Intervalo de dados e redução *Dropout*	**Rede Neural com 1 *Dropout*** 75% Treino Janela de previsão 60 dias **Dados desde 2014**	0,68	0,87	0,87	1,36

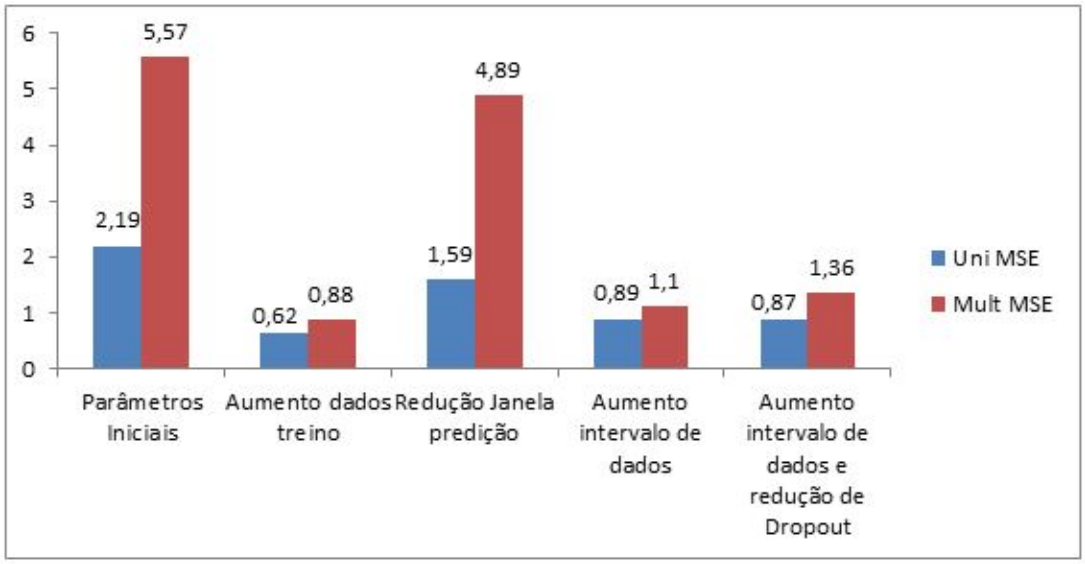
Fonte: Autoria Própria (2022).

Figura 16 – Comparação gráfica dos MAEs de alterações nos hiperparâmetros



Fonte: Autoria Própria (2022).

Figura 17 – Comparação gráfica dos MSEs de alterações nos hiperparâmetros



Fonte: Autoria Própria (2022).

4 CONCLUSÃO

Seguindo a metodologia CRISP-DM o estudo passou pelos seis passos de resolução de problemas proposto, e obteve um resultado conclusivo para a comparação dos dois modelos.

Como objetivo geral, conclui-se que embora o modelo LSTM, modelado conforme descrito na preparação dos dados, siga a tendência do mercado financeiro, seu erro médio absoluto na predição de uma ação cujo menor valor foi de 0,59, não permite que o investidor tenha uma assertividade relevante para buscar retornos financeiros, não sendo então factível para uma aplicação real.

O resultado demonstrou também que o LSTM univariável teve um desempenho melhor, ou seja, um erro menor frente aos dados reais do que o LSTM multivariável. Embora todas as variáveis utilizadas no modelo Multi tivessem uma boa correlação com a cotação final da ação da Petrobras, o modelo foi superado em todos os testes com diferentes alterações nos hiperparâmetros.

4.1 Limitações

Este estudo teve como limitação a análise das ações de apenas uma empresa, a Petrobras, cujo seu maior acionista é o Governo Federal, e está, portanto, suscetível a oscilações de preço devido a interferências políticas que são imprevisíveis.

Outra limitação do estudo foi a utilização das redes neurais recorrentes LSTM, que não traduziram a boa correlação das variáveis do *dataset* multivariável em redução de erro na predição de valores reais, podendo esta conclusão estar atrelada somente a essa rede neural específica.

4.2 Trabalhos Futuros

Para avaliar mais profundamente a importância de atributos que tenham boa correlação de uma ação na predição de uma cotação futura seria recomendado testar outras metodologias (como o modelo *Prophet*).

Além disso, seria recomendado testes com outras ações e mercados distintos, com o objetivo de minimizar as oscilações de preço causadas por ações políticas na Petrobras e no mercado brasileiro.

Outra abordagem que pode ser utilizada é o aumento do tamanho da amostra. Nota-se que houve uma melhor performance do modelo quando a quantidade de dados foi aumentada (na mudança de hiperparâmetros em que os dados foram considerados à partir de 2014).

Referências

BEAUFAYS, H. S. e Andrew Senior e F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. 2014. Disponível em: <https://arxiv.org/abs/1402.1128>. Citado na página 12.

CHEN, Y. Z. e. F. D. K. A lstm-based method for stock returns prediction: A case study of china stock market. **IEEE International Conference on Big Data**, p. 2823–2824, 2015. Citado na página 14.

CLINTON, P. C. e J. Crisp-dm 1.0: Step-by-step data mining guide. 2000. Disponível em: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>. Citado na página 15.

DSA, E. **The Deep Learning Book**. 2022. Disponível em: <https://www.deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory>. Acesso em: 01 de junho de 2022. Citado na página 13.

FAWCETT, F. P. e T. **Data Science para Negócios**. Rio de Janeiro: Alta Books, 2016. Citado na página 14.

G.I., S. C. e W. **Mean Squared Error**. 2011. Disponível em: <https://link.springer.com/referenceworkentry/10.1007/978-0-387-30164-8_528>. Acesso em: 01 de junho de 2022. Citado na página 14.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado 2 vezes nas páginas 11 e 12.

KOTB, A. S. e M. Unsupervised pre-training of a deep lstm-based stacked autoencoder for multivariate time series forecasting problems. Sci Rep 9. 2019. Citado na página 14 .

MORRIS, R. C. S. e E. R. Understanding lstm - a tutorial into long short-term memory recurrent neural networks. arXiv, p. 2, 2019. Citado na página 12.

NETO, A. A. **Valuation, métricas de valor e avaliação de empresas**. São Paulo: Atlas, 2020. Citado na página 11.

SELIYA, J. M. A. e Rushit Dave e N. Applications of recurrent neural network for biometric authentication anomaly detection. **MDPI**, v. 20, n. 1, p. 1–2, 2021. Citado na página 12.

TOROINVESTIMENTOS, E. **O que é Ibovespa**. 2022. Disponível em: <https://blog.toroinvestimentos.com.br/o-que-e-ibovespa>. Acesso em: 20 de maio de 2022. Citado na página 11.

VERSIGNASSI, A. **Crash, Uma breve história da economia**. Rio de Janeiro: Harper Collins, 2019. Citado na página 11.

WERBOS, P. J. Backpropagation through time: what it does and how to do it. **IEEE**, v. 78, n. 1, p. 1550 – 1560, 1990. Citado na página 12.