



UFOP

Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

**Redes Neurais LSTM e Google Trends
aplicados para previsão de Séries
Temporais do Mercado Financeiro no
contexto de Criptomoedas**

Rosabel Vieira Braga

**TRABALHO DE
CONCLUSÃO DE CURSO**

ORIENTAÇÃO:

Fernando Bernardes de Oliveira

COORIENTAÇÃO:

Marcos Antônio Alves

Agosto, 2023

João Monlevade–MG

Rosabel Vieira Braga

**Redes Neurais LSTM e Google Trends aplicados
para previsão de Séries Temporais do Mercado
Financeiro no contexto de Criptomoedas**

Orientador: Fernando Bernardes de Oliveira

Coorientador: Marcos Antônio Alves

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Agosto de 2023

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

B813r Braga, Rosabel Vieira.

Redes neurais LSTM e Google Trends aplicados para previsão de séries temporais do mercado financeiro no contexto de criptomoedas. [manuscrito] / Rosabel Vieira Braga. - 2023.

64 f.

Orientador: Prof. Dr. Fernando Bernardes de Oliveira.

Coorientador: Dr. Marcos Antônio Alves.

Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Graduação em Sistemas de Informação .

1. Análise de séries temporais. 2. Bitcoin. 3. Inteligência computacional. 4. Mercado financeiro - Moedas. 5. Previsão. 6. Redes neurais (Computação). I. Oliveira, Fernando Bernardes de. II. Alves, Marcos Antônio. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 004.8:519.2

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



FOLHA DE APROVAÇÃO

Rosabel Vieira Braga

**Redes Neurais LSTM e Google Trends aplicados para previsão de
Séries Temporais do Mercado Financeiro no contexto de Criptomoedas**

Monografia apresentada ao Curso de Sistemas de Informação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação

Aprovada em 01 de setembro de 2023

Membros da banca

Prof. Dr. Fernando Bernardes de Oliveira - Orientador (Universidade Federal de Ouro Preto)
Dr. Marcos Antônio Alves - Coorientador (Fundação para Inovações Tecnológicas - FITec)
Prof. Dr. Elias Silva de Medeiros - Avaliador (Universidade Federal da Grande Dourados - UFGD)
Prof. Petrônio Cândido de Lima e Silva - Avaliador (Instituto Federal de Educação Ciência e Tecnologia do Norte de Minas Gerais - IFNMG)

Fernando Bernardes de Oliveira, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 29/09/2023



Documento assinado eletronicamente por **Fernando Bernardes de Oliveira, PROFESSOR DE MAGISTERIO SUPERIOR**, em 29/09/2023, às 15:48, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0598367** e o código CRC **A6401B3E**.

Este trabalho é dedicado ao meu Starman, meu pai, Gilberto.

Agradecimentos

Primeiramente, agradeço a Deus, fonte inesgotável de força e inspiração, por guiar meus passos ao longo desta jornada acadêmica e por me conceder a sabedoria e a perseverança necessárias para alcançar este marco em minha vida.

À minha família, especialmente aos meus pais e irmãos, que sempre estiveram ao meu lado, oferecendo apoio emocional, encorajamento constante e valiosos ensinamentos. Sem o amor e o apoio de vocês, este feito não teria sido possível.

Aos meus amigos, verdadeiros pilares de apoio, agradeço por compartilharem comigo a jornada, pelas conversas estimulantes e pelo incentivo nos momentos mais desafiadores. Suas risadas e amizade fizeram toda a diferença.

À comunidade acadêmica, que proporcionou um ambiente rico em conhecimento, debates e descobertas, expresso minha profunda gratidão. A troca de ideias e a colaboração foram fundamentais para o meu crescimento.

Aos meus orientadores, cuja expertise e orientação foram cruciais para a realização deste trabalho. Suas críticas construtivas, orientações cuidadosas e compromisso inabalável com minha formação acadêmica foram inestimáveis.

Por fim, agradeço a mim, a Rosabel do passado pela coragem de viver e enfrentar o novo.

“All models are wrong, but some are useful.”

— George E. P. Box (1919 – 2013),
in: Empirical Model-Building and Response Surfaces.

Resumo

O Bitcoin foi originado com o objetivo de criar uma moeda única, sem encargos e independente de bancos. Porém, o seu crescente valor de mercado fez com que se popularizasse também como alternativa de investimento. Para obter lucro, os investidores devem, em um cenário ideal, comprar o ativo quando o seu preço estiver baixo e revendê-lo quando os preços aumentarem. Mas a extrema volatilidade da moeda, que sofre influência de diversos fatores e experimenta grandes flutuações de preço em curtos períodos de tempo, torna sua predição um desafio para pesquisadores em todo o mundo. Partindo disso, o presente trabalho busca identificar se a agregação de índices com a frequência de busca por palavras-chaves, ligadas ao Bitcoin, no Google, junto ao modelo de *Recurrent neural networks* (RNN), o *Long Short-Term Memory* (LSTM), pode contribuir para a predição do valor futuro do ativo. Para isso, foram utilizados dados do período de 2014 a 2022, divididos em três distribuições distintas dos dados, definidas como cenários de estabilidade, tendência de alta e tendência de queda do preço da moeda. As janelas de tempo definidas foram de 90 e 10 dias, e em cada um dos experimentos foi adotada a abordagem de modelo univariado, e multivariado. Considerando o contexto experimental estabelecido, os resultados sugerem que a agregação do sentimento, em uma janela de tempo de 10 dias no cenário de queda, apresentou uma maior capacidade de generalização do preço futuro da criptomoeda, quando observado as métricas de avaliação utilizadas.

Palavras-chaves: Séries Temporais. Inteligência Computacional. Redes Neurais Recorrentes. Criptomoedas. Bitcoin. Google Trends.

Abstract

Bitcoin was originally conceived with the intention of creating a singular currency, free from charges and independent of banks. However, its burgeoning market value has propelled it to become a viable investment alternative. In an ideal scenario, investors aim to purchase the asset when its price is low and sell it when prices escalate, thus generating profit. Nevertheless, the substantial volatility of the currency, influenced by a myriad of factors and experiencing substantial price fluctuations within brief time intervals, bestows upon its prediction a challenge for researchers globally. From this premise, the current study endeavors to ascertain whether the amalgamation of indices alongside the frequency of keyword searches associated with Bitcoin on Google, combined with the Recurrent Neural Network (RNN) model - specifically, the Long Short-Term Memory (LSTM) architecture - can contribute to forecasting the future value of the asset. To this end, data spanning from 2014 to 2022 were employed, partitioned into three distinct data distributions that delineate scenarios of stability, upward trends, and downward trends in the currency's price. The designated time windows encompassed 90 and 10 days, and each experiment embraced both univariate and multivariate modeling approaches. Within the established experimental context, the results suggest that the aggregation of sentiment, within a 10-day time frame during a downturn scenario, exhibited a greater capacity for generalizing the cryptocurrency's future price, as observed through the employed evaluation metrics.

Key-words: Time Series. Computational Intelligence. Recurrent Neural Networks. Cryptocurrencies. Bitcoin. Google Trends.

Lista de ilustrações

Figura 1 – Neurônio Artificial	25
Figura 2 – Exemplo de uma Rede Neural Artificial de 2 camadas com 4 entradas e 2 saídas	26
Figura 3 – Rede neural totalmente recorrente	27
Figura 4 – Rede neural simples	27
Figura 5 – LSTM	28
Figura 6 – <i>Forget Gate</i>	29
Figura 7 – <i>Input Gate e Input Modulation Gate</i>	29
Figura 8 – Atualização da célula de memória	30
Figura 9 – Saída final	31
Figura 10 – Preço de fechamento do Bitcoin no período analisado	35
Figura 11 – Fluxograma da metodologia proposta	36
Figura 12 – Escala de valores em cada cenário observado	38
Figura 13 – Cenário 1 - Janela de 90 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos	42
Figura 14 – Cenário 1 - Janela de 10 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos	42
Figura 15 – Cenário 1 - Resíduo da diferença do valor real e do estimado	44
Figura 16 – Cenário 2 - Janela de 90 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos	45
Figura 17 – Cenário 2 - Janela de 10 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos	46
Figura 18 – Cenário 2 - Resíduo da diferença do valor real e do estimado	47
Figura 19 – Cenário 3 - Janela de 90 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos	48
Figura 20 – Cenário 3 - Janela de 10 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos	48
Figura 21 – Cenário 3 - Resíduo da diferença do valor real e do estimado	50
Figura 22 – Cenário 1 - Janela de 10 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis	60
Figura 23 – Cenário 1 - Janela de 90 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis	61
Figura 24 – Cenário 1 - Resíduo da diferença do valor real e das médias móveis	61
Figura 25 – Cenário 2 - Janela de 10 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis	62

Figura 26 – Cenário 2 - Janela de 90 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis	62
Figura 27 – Cenário 2 - Resíduo da diferença do valor real e das médias móveis . .	63
Figura 28 – Cenário 3 - Janela de 10 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis	63
Figura 29 – Cenário 3 - Janela de 90 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis	64
Figura 30 – Cenário 3 - Resíduo da diferença do valor real e das médias móveis . .	64

Lista de tabelas

Tabela 1 – Parâmetros da rede LSTM	37
Tabela 2 – Cenário 1 - Resultado das métricas	43
Tabela 3 – Cenário 2 - Resultado das métricas	46
Tabela 4 – Cenário 3 - Resultado das métricas	49
Tabela 5 – Média Móvel - Resultado das métricas	51

Lista de abreviaturas e siglas

ACF *Autocorrelation Function*

ABT *Analytical Base Table*

BTC *Bitcoin*

CRISP-DM *Cross Industry Standard Process for Data Mining*

EMA *Exponential Moving Average*

FCMs *Fuzzy Cognitive Maps*

MAE *Mean Absolute Error*

MAPE *Mean Absolute Percentage Error*

MDA *Mean Directional Accuracy*

MFE *Mean Forecast Error*

MSE *Mean Squared Error*

MM *Média Móvel*

LSTM *Long Short-Term Memory*

PACF *Partial Autocorrelation Function*

OMS *Organização Mundial da Saúde*

R² *R-Squared*

RMSE *Root Mean Squared Error*

RNN *Recurrent neural networks*

SEC *U.S. Securities and Exchange Commission*

SP500 *Standard & Poor's 500*

SMA *Simple Moving Average*

WMVFTS *Weighted Multivariate Fuzzy Time Series*

Lista de símbolos

Σ	AutoSoma
φ	Phi minúscula
$US\$$	Dólar americano
n	Número de observações na série temporal
y_t	Valor real da série temporal no período t
\hat{y}_t	Valor previsto pela modelagem no período t
X	Valor original
X_{norm}	Valor normalizado de X
X_{min}	Valor mínimo na amostra original
X_{max}	Valor máximo na amostra original

Sumário

1	INTRODUÇÃO	17
1.1	Motivação	17
1.2	Objetivos	18
1.2.1	Objetivo Geral	18
1.2.2	Objetivos Específicos	18
1.3	Metodologia	18
1.4	Organização do trabalho	19
2	REVISÃO BIBLIOGRÁFICA	20
2.1	Mercado Financeiro	20
2.1.1	Média Móvel	20
2.2	Criptomoeda	21
2.2.1	Bitcoin	21
2.3	Google Trends	22
2.4	Séries Temporais	23
2.4.1	Séries temporais não lineares	23
2.4.2	Análise de Séries Temporais	24
2.4.2.1	Autocorrelação	24
2.4.2.2	Causalidade	24
2.5	Redes Neurais Artificiais	25
2.5.1	Rede Neural Recorrente	26
2.5.2	LSTM	27
2.6	Trabalhos Relacionados	31
2.7	Considerações finais	33
3	MATERIAL E MÉTODO	34
3.1	Coleta de dados	34
3.1.1	Bitcoin	34
3.1.2	Google Trends	35
3.1.3	Definição da ABT	35
3.2	Implementação dos algoritmos	36
3.2.1	Desenvolvimento da LSTM	37
3.3	Definição do experimento	38
3.3.1	Métricas de avaliação	39
3.4	Considerações Finais	40

4	RESULTADOS E DISCUSSÃO	41
4.1	Cenário 1 - Estabilidade	41
4.1.1	Janela de 90 dias	41
4.1.2	Janela de 10 dias	42
4.1.3	Comparação dos resultados	43
4.2	Cenário 2 - Tendência de alta	44
4.2.1	Janela de 90 dias	44
4.2.2	Janela de 10 dias	45
4.2.3	Comparação dos resultados	46
4.3	Cenário 3 - Tendência de queda	47
4.3.1	Janela de 90 dias	47
4.3.2	Janela de 10 dias	48
4.3.3	Comparação dos resultados	49
4.4	Média Móvel	50
4.5	Discussão	51
4.6	Considerações finais	52
5	CONCLUSÃO	53
5.1	Limitações da metodologia proposta	54
5.2	Propostas de trabalhos futuros	54
	REFERÊNCIAS	55
	APÊNDICES	59
	APÊNDICE A – ANÁLISE GRÁFICA DA MÉDIA MÓVEL	60
A.1	Cenário 1 - Estabilidade	60
A.2	Cenário 2 - Tendência de alta	62
A.3	Cenário 3 - Tendência de queda	63

1 Introdução

O Bitcoin (BTC) foi criado com o objetivo de se tornar uma moeda única, sem encargos e independente de bancos (NAKAMOTO, 2008). A sua capacidade de facilitar movimentações financeiras, graças as características herdadas pela tecnologia *blockchain* - da qual é dependente - e do seu crescente valor de mercado, fez com que o ativo atraísse grande atenção de investidores, líderes mundiais, da mídia e do público em geral (PATEL et al., 2020). No entanto, o seu preço é altamente volátil e depende de vários fatores, como o custo de transação, dificuldade de mineração, tendências de mercado, popularidade, preço de moedas alternativas, mercado de ações, sentimentos e alguns fatores legais (SOVBETOV, 2018). Essa volatilidade tem sido uma preocupação para pesquisadores em todo o mundo, que buscam entender o seu comportamento e modelá-las para estimar seu valor futuro.

Partindo disso, este trabalho busca estudar o comportamento do Bitcoin - a criptomoeda com maior valor de mercado¹ -, e validar se a hipótese de que a indexação da frequência normalizada de busca por palavras-chaves no Google, pode contribuir para a previsão do seu valor futuro, aplicado a diferentes cenários (estabilidade, tendência de alta e tendência de queda).

1.1 Motivação

Investir no mercado financeiro é uma atividade complexa que requer um profundo conhecimento das melhores estratégias de investimento. Neste sentido, as técnicas de análise de dados têm se tornado cada vez mais populares entre os investidores, pois permitem a identificação de padrões e tendências do mercado, auxiliando na tomada de decisões mais assertivas. As criptomoedas, apesar da alta volatilidade e do risco associado, têm atraído a atenção de investidores em todo o mundo. No entanto, investir nestes ativos ainda é um campo em desenvolvimento, e a aplicação das técnicas apropriadas nesse contexto requer uma abordagem cuidadosa.

Com isso, este trabalho tem como objetivo aplicar técnicas avançadas de séries temporais e inteligência computacional para estimar o preço futuro do Bitcoin. Espera-se que os resultados deste trabalho possam contribuir para o desenvolvimento de estratégias que ajudem os investidores a alcançar melhores resultados financeiros e a minimizar seus riscos.

¹ <https://coinmarketcap.com/>

1.2 Objetivos

1.2.1 Objetivo Geral

Estimar o preço de fechamento do Bitcoin, pela frequência normalizada da busca por palavras-chaves no Google. O estudo utilizará uma rede neural recorrente LSTM, para fazer inferência do valor futuro do ativo.

1.2.2 Objetivos Específicos

Este trabalho tem os seguintes objetivos específicos:

- Identificar as principais características e comportamentos da série temporal do ativo.
- Treinar um método de *Deep Learning*, LSTM, para prever o preço de fechamento do BTC.
- Utilizar o modelo treinado para fazer a predição do preço de fechamento do BTC, em janelas de 90 e 10 dias, com e sem a agregação de variáveis exógenas.
- Investigar os melhores resultados para cenários de estabilidade, alta e baixa.
- Analisar a efetividade do uso do índice como fator de agregação, por meio de experimentos computacionais utilizando instâncias de teste e coleção de dados.

1.3 Metodologia

A metodologia empregada para a execução do projeto envolveu tanto uma pesquisa bibliográfica, quanto uma pesquisa experimental. As atividades realizadas neste estudo podem ser descritas da seguinte maneira:

1. Revisão da literatura: pesquisa e análise da literatura sobre tópicos relacionados a Séries Temporais, *Blockchain*, Criptomoedas, Inteligência Computacional, *Deep Learning*, entre outros.
2. Coleta de dados: pesquisa e construção de uma coleção de dados relevantes para a moeda e palavras-chave de interesse.
3. Análise descritiva da Série Temporal: análise dos dados históricos obtidos para identificação de suas características temporais.
4. Definição das técnicas de Inteligência Computacional: definição das técnicas adequadas para predição do comportamento do Bitcoin.

5. Implementação: implementação das técnicas de Inteligência Computacional escolhidas para o contexto estudado.
6. Planejamento e execução do experimento: planejamento e execução do experimento em relação às criptomoedas, incluindo todos os testes necessários para validar a abordagem proposta.
7. Análise e discussão: avaliação dos resultados obtidos por meio de técnicas estatísticas apropriadas, com as devidas considerações e ponderações.

1.4 Organização do trabalho

O restante deste trabalho é organizado como se segue. O Capítulo 2 aborda a revisão bibliográfica contendo os fundamentos teóricos mais importantes utilizados na solução do problema estudado, assim como alguns trabalhos relacionados ao tema. No Capítulo 3 é apresentado o material e método proposto para o desenvolvimento deste trabalho. No Capítulo 4 é dissertado a análise e discussão dos resultados gerados pelos experimentos realizados no estudo. Por fim, no Capítulo 5, é exposta as conclusões, bem como algumas limitações e sugestões para trabalhos futuros.

2 Revisão Bibliográfica

Neste capítulo são discutidos conceitos fundamentais para o melhor entendimento deste trabalho. Na Seção 2.1 é explicado o Mercado Financeiro, e na Seção 2.2 detalhamos as criptomoedas de maneira geral, elucidando sua origem, finalidade e comportamento. A Seção 2.3 apresenta o Google Trend, justificando o seu uso. A 2.4 explica de maneira sucinta o conceito de séries temporais, e aborda os tópicos principais. Na Seção 2.5 introduzimos o conceito de Redes Neurais Artificiais, descrevendo suas variações e abordando os conceitos do LSTM. A Seção 2.6 menciona trabalhos de previsão de séries temporais, e por fim na Seção 2.7 encerra o capítulo resumindo o que foi abordado.

2.1 Mercado Financeiro

O mercado financeiro é um ambiente onde ocorre a negociação de ativos financeiros, como ações, títulos, moedas, commodities e derivativos (HULL, 2016). É um componente crucial da economia global, proporcionando mecanismos para empresas e governos captarem recursos financeiros e para os investidores alocarem seus fundos visando retornos financeiros.

Os principais conceitos do mercado financeiro incluem: (1) oferta e demanda de ativos, que determinam os preços dos ativos financeiros; (2) risco e retorno, que estão interligados e influenciam as decisões de investimento; (3) liquidez, que se refere à facilidade de comprar ou vender um ativo sem afetar significativamente seu preço; (4) diversificação, que busca reduzir riscos investindo em diferentes ativos; (5) intermediários financeiros, como bancos e corretoras, que facilitam a negociação entre compradores e vendedores; e (6) informações e análises, que desempenham um papel crucial na tomada de decisões de investimento.

2.1.1 Média Móvel

Murphy (1999) explica que a Média Móvel (MM) é uma técnica amplamente utilizada no mercado financeiro para suavizar as flutuações de preços e identificar tendências em séries temporais de dados financeiros. Esta pode ser calculada de diversas maneiras, mas a *Simple Moving Average* (SMA) e a *Exponential Moving Average* (EMA) são as duas principais formas.

A SMA é calculada somando os preços de um período específico e dividindo pela quantidade de períodos. Ela atribui o mesmo peso a todos os preços dentro do intervalo de tempo. Já a EMA dá mais peso aos preços mais recentes, tornando-a mais sensível às mudanças recentes.

2.2 Criptomoeda

As criptomoedas são uma forma de moeda digital que utiliza criptografia para garantir a segurança das transações e controlar a criação de novas unidades. Diferentemente das moedas tradicionais, como o dólar americano ou o euro, as criptomoedas não são emitidas por um banco central ou autoridade governamental centralizada (NAKAMOTO, 2008; SWAN, 2015; RADZIWILL, 2018). Em vez disso, são baseadas em uma tecnologia de registro distribuído chamada *blockchain*, que permite que as transações sejam verificadas e registradas de forma segura e transparente, sem a necessidade de intermediários financeiros.

As criptomoedas são conhecidas por serem altamente voláteis em relação a outras moedas, com flutuações significativas de preço em curtos períodos de tempo. No entanto, também são consideradas uma forma de pagamento segura e privada, já que as transações são registradas de forma criptografada no *blockchain*, tornando-as muito difíceis de serem falsificadas ou rastreadas.

2.2.1 Bitcoin

O Bitcoin foi a primeira criptomoeda a ser implementada em 2009 por uma pessoa ou grupo de pessoas sob o pseudônimo de Satoshi Nakamoto (NAKAMOTO, 2008). A motivação por trás da sua criação era fornecer uma forma de transação eletrônica que fosse segura, privada e independente de governos ou instituições financeiras tradicionais (SWAN, 2015).

Desde então, o ativo se tornou uma das criptomoedas mais populares e valiosas do mundo, sendo amplamente utilizado como forma de investimento e meio de pagamento em diversos países (RADZIWILL, 2018). O seu valor é altamente volátil, com flutuações significativas de preço em curtos períodos de tempo. Em dezembro de 2017, o valor do Bitcoin atingiu o pico de cerca de US\$20.000 por unidade, mas caiu para menos de US\$4.000 em dezembro de 2018. Em 2021, o seu preço atingiu novos recordes históricos. Em abril, ele ultrapassou os US\$60.000 por unidade ¹. No entanto, a partir de 2022, o preço começou a cair e teve uma queda significativa em maio, atingindo cerca de US\$30.000 por unidade.

Apesar de ser amplamente utilizado como forma de investimento, o Bitcoin ainda não é amplamente aceito como meio de pagamento em muitos países, principalmente devido à falta de regulamentação e às preocupações com segurança e privacidade (RADZIWILL, 2018). No entanto, muitos dos seus defensores argumentam que ele pode ser uma forma mais segura e privada de fazer transações online do que outras formas de pagamento eletrônico, como cartões de crédito (SWAN, 2015).

¹ <https://www.coindesk.com/price/bitcoin>

2.3 Google Trends

Conforme apontado por [Caetano \(2021\)](#), a popularização da internet e a consequente disponibilidade de informações em tempo real levaram à necessidade de avaliação de tarefas na comunidade online. Com o advento das mídias digitais e a disseminação das redes sociais, as empresas passaram a perceber que poderiam utilizar essas plataformas para obter informações valiosas sobre seus clientes, como o comportamento de compra, o momento ideal para o lançamento de produtos e a localização geográfica dos consumidores.

Nesse contexto, o Google Trends se destaca como uma das ferramentas mais importantes para a busca de informações sobre os assuntos mais discutidos ([CAETANO, 2021](#)). Através desta, é possível identificar a maior tendência do movimento nas redes sociais, o assunto do momento e a localidade do mundo onde determinado tema tem mais impacto. Desde 2004, o Google disponibiliza o *trendline*, que auxilia os usuários programadores a obter informações sobre os assuntos do momento. Além disso, existe o *Pytrends*, uma ferramenta digital que permite que programas de computador baixem dados importantes sobre diversos assuntos sem a necessidade de acessar o site do Google. Dessa forma, é possível desenvolver estatísticas e algoritmos próprios para a área de expertise do analista. Em resumo, se trata de uma ferramenta gratuita que permite acompanhar a popularidade relativa de determinadas palavras-chave ou termos de pesquisa ao longo do tempo.

Diferentemente do Twitter ², cujo acesso é limitado no tempo, e do LexisNexis, que é um serviço baseado em assinatura, o Google Trends possui a vantagem de ser disponibilizado gratuitamente ([ASLANIDIS; BARIVIERA; LÓPEZ, 2022](#)). Trata-se de uma ferramenta simples de ser acessada e reflete amplamente a atenção de um perfil mais abrangente de investidores, o que possibilita sua utilização na previsão de valores de indicadores econômicos de curto prazo.

² Renomeada para X, em 24 de julho de 2023

2.4 Séries Temporais

Uma série temporal pode ser conceituada como um conjunto de observações que se encontram dispostas em conformidade com um parâmetro temporal, revelando uma interdependência sequencial. Um dos propósitos subjacentes à investigação de séries temporais reside na análise e na formulação de modelos que dão conta dessa interligação (SILVA; GUIMARÃES; TAVARES, 2008).

Tais séries podem ser categorizadas de acordo com critérios como a natureza discreta ou contínua, bem como a distinção entre determinismo e estocasticidade, além da possibilidade de serem multivariadas e multidimensionais. A representação dessas séries por meio da abordagem de componentes não observáveis envolve a fusão harmoniosa de quatro elementos: sazonalidade, padrões cíclicos, tendências e ruído branco (MENDENHALL; BEAVER; BEAVER, 1999).

Uma série temporal é dita sazonal se seu comportamento é influenciado por fatores sazonais e o mesmo tende a se repetir a cada “s” períodos de tempo. Uma série é cíclica se seus dados possuem quedas e picos que não são fixos a um determinado período de tempo e nem a algum aspecto do calendário. A tendência da série pode ser de crescimento, decrescimento ou constante a longo prazo. Essa tendência pode ter diversos padrões de alteração, podendo crescer ou decrescer de forma exponencial, linear ou amortecida. Componentes de erros que apresentam variações ascendentes ou descendentes após a ocorrência de um efeito de tendência, um efeito cíclico, ou um efeito sazonal, podem aparecer.

As séries temporais também podem ser classificadas como estacionárias ou não estacionárias, quando tratado da variabilidade das observações. Em uma série temporal estacionária, as estatísticas não são afetadas por variações no tempo. Por outro lado, na série considerada não estacionária, apenas a realização do processo estocástico é suficiente para se obter todas as estatísticas do mesmo (MORETTIN; TOLOI, 2006).

2.4.1 Séries temporais não lineares

Uma série temporal não linear é uma sequência de dados que não segue um padrão linear e pode apresentar comportamentos caóticos e imprevisíveis. De acordo com a teoria do caos, pequenas variações nas condições iniciais podem levar a grandes diferenças nos resultados finais, o que torna difícil prever o comportamento futuro da série (KLIOUTCHNIKOV; SIGOVA; BEIZEROV, 2017).

No contexto do Bitcoin, a série temporal não linear é uma característica importante, uma vez que o preço do Bitcoin é altamente volátil e pode ser influenciado por diversos fatores interconectados e complexos. A dinâmica caótica do Bitcoin pode ser atribuída a fatores como mudanças nas políticas governamentais, adoção global, flutuações do mercado

e mudanças nas percepções dos investidores.

Diante disso, é importante destacar que previsões de preços do Bitcoin baseadas em modelos lineares ou simplistas podem não ser precisas ou confiáveis. Em vez disso, é necessário levar em consideração a complexidade e a dinâmica não linear da série temporal do Bitcoin ao fazer previsões e análises de tendências (MOAT et al., 2013).

Portanto, é recomendável o uso de modelos mais sofisticados que considerem a complexidade da série temporal, como redes neurais artificiais e modelos baseados em teoria do caos. Essas técnicas podem ajudar a prever tendências futuras do preço do Bitcoin de maneira mais precisa e confiável (CHEN; LI; SUN, 2020).

2.4.2 Análise de Séries Temporais

A análise de séries temporais é uma ferramenta importante para prever o comportamento de uma série de dados ao longo do tempo. Nesse contexto, dois conceitos fundamentais são a autocorrelação e a causalidade.

2.4.2.1 Autocorrelação

A autocorrelação é uma medida estatística usada para avaliar a correlação entre os valores de uma série temporal e seus valores anteriores. É uma medida importante porque muitos modelos de previsão de séries temporais dependem de sua existência ou não (CHATFIELD; XING, 2019).

Existem duas principais formas de medir a autocorrelação: a *Autocorrelation Function* (ACF) e a *Partial Autocorrelation Function* (PACF). A ACF é uma medida da correlação entre uma série temporal e suas observações passadas, enquanto a PACF mede a correlação entre a série temporal e uma observação passada, controlando o efeito das observações intermediárias.

A memória em séries temporais é um conceito que se refere à dependência das observações presentes em relação às observações passadas, que podem ser identificadas por meio da ACF e PACF (ENDERS, 2015). Uma série temporal com memória é aquela em que os valores passados influenciam os valores presentes ou futuros, enquanto uma série sem memória é aquela em que os valores passados não afetam os valores presentes ou futuros, sendo esta uma característica importante para a modelagem e previsão dessas séries.

2.4.2.2 Causalidade

A causalidade refere-se à relação de causa e efeito entre duas variáveis; destacando que correlação não implica necessariamente causalidade. Para estabelecer causalidade, é necessário encontrar uma relação de causa e efeito bem definida entre as variáveis. A

análise de causalidade pode ser realizada usando técnicas de modelagem, como a análise de causalidade de Granger. Essa técnica estatística examina se as mudanças em uma variável podem ser usadas para prever mudanças em outra variável, sendo útil para identificar relações de causa e efeito entre variáveis em uma série temporal (GRANGER, 1969).

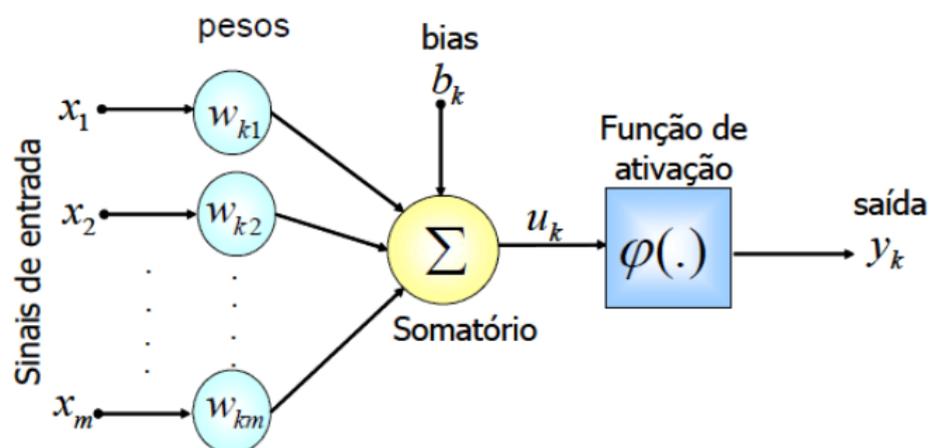
2.5 Redes Neurais Artificiais

As redes neurais artificiais são uma das bases fundamentais da inteligência artificial (GOODFELLOW; BENGIO; COURVILLE, 2016). Elas são compostas por algoritmos que buscam replicar o funcionamento do cérebro humano, permitindo que as máquinas possam aprender a partir de exemplos e experiências. O termo “neural” se refere à imitação das estruturas neurais biológicas, enquanto o termo “artificial” indica que essas redes são criadas por meio de programação computacional.

Essas redes são compostas por camadas de neurônios interligados, que processam informações a partir de entradas de dados. Cada camada é composta por um número variável de neurônios, e cada neurônio é conectado a todos os neurônios da camada seguinte. O número de camadas e neurônios varia de acordo com a complexidade da rede e do problema que se deseja resolver.

Entre as principais características das redes neurais, é citado a capacidade de aprender e generalizar, ou seja, de se adaptar a novos dados e reconhecer padrões em diferentes contextos. Além disso, as redes neurais são capazes de lidar com dados não-lineares e de alta dimensionalidade, o que as torna muito úteis em diversas áreas da ciência e tecnologia.

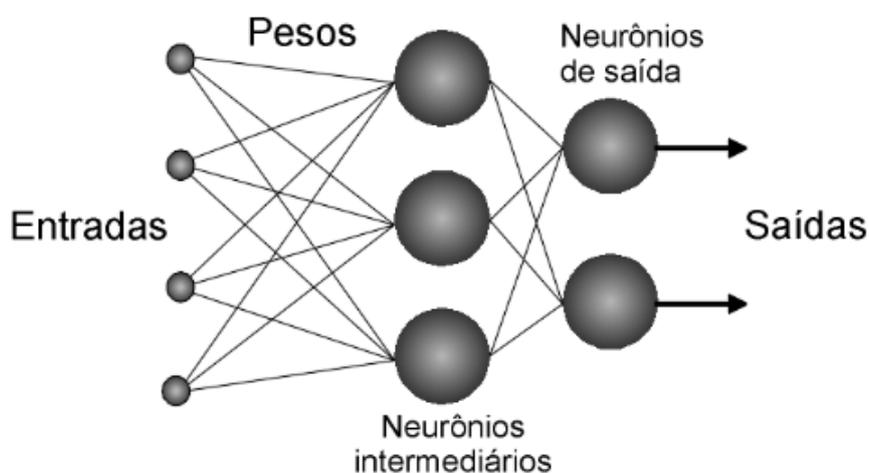
Figura 1 – Neurônio Artificial



Fonte: Soares e Silva (2011)

A Figura 1 ilustra um exemplo de um neurônio artificial com entradas iguais a X_1, X_2, \dots, X_m , bem como os pesos sinápticos W_i atribuídos a cada uma das entradas. A saída linear u_k é gerada a partir da soma das entradas X_i ponderadas pelos pesos W_i , que passam por uma função de ativação $\varphi(\cdot)$ para gerar a saída do neurônio (SOARES; SILVA, 2011). Na Figura 2, é possível visualizar um exemplo simples de uma rede neural artificial, onde as camadas de entrada, ocultas e de saída são ilustradas.

Figura 2 – Exemplo de uma Rede Neural Artificial de 2 camadas com 4 entradas e 2 saídas



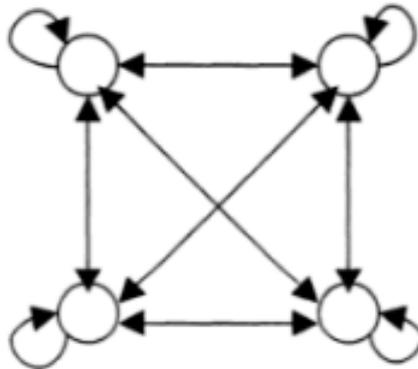
Fonte: Tafner (1998)

2.5.1 Rede Neural Recorrente

De acordo com Nielsen (2019), a RNN representa uma classe abrangente de redes neurais em que os mesmos parâmetros são aplicados repetidamente, mesmo quando as entradas mudam com o passar do tempo. Essas redes são amplamente utilizadas no campo acadêmico e no mercado para tarefas baseadas em sequência, linguagem, previsão e classificação de séries temporais.

Entre as suas principais características é destacado sua capacidade de analisar intervalos de tempo ordenados, preservar o estado de um intervalo de tempo para o outro e possuir parâmetros que ajudam a atualizar seu estado, incluindo um estado oculto. A Figura 3 ilustra uma rede neural totalmente recorrente, em que cada neurônio recebe como entrada a saída de cada um dos neurônios, inclusive de si mesmo (auto-realimentação).

Figura 3 – Rede neural totalmente recorrente



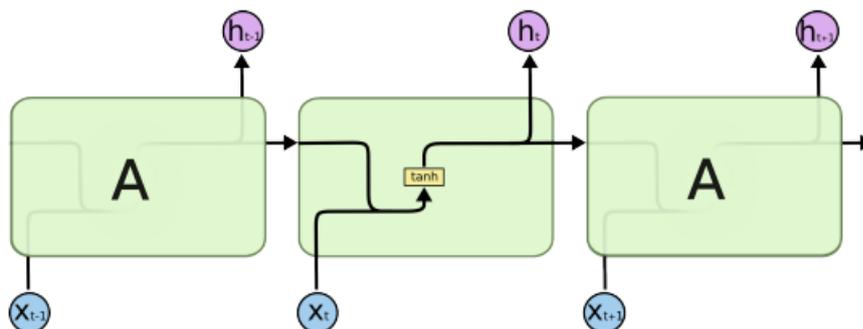
Fonte: Medsker e Jain (1999)

2.5.2 LSTM

Dentre as redes neurais recorrentes, a LSTM é destacada por possibilitar a manutenção de entradas anteriores por um longo período, característica essencial no processo de previsão de valores de saídas futuras. Ela possibilita aprender dependências de longo prazo e evitar que informações importantes sejam esquecidas com o passar do tempo (CHEN; ZHOU; DAI, 2015).

Esta é capaz de aprender e lembrar das informações mais relevantes por meio de uma célula que armazena a memória de longo prazo. Diferentemente das redes neurais tradicionais, que possuem uma cadeia de módulos repetidos de estruturas simples de camada única com transformação do tipo tangente hiperbólica (Figura 4), as LSTMs contêm quatro camadas interconectadas que trabalham em conjunto para realizar a tarefa de aprendizado (NELSON; PEREIRA; OLIVEIRA, 2017).

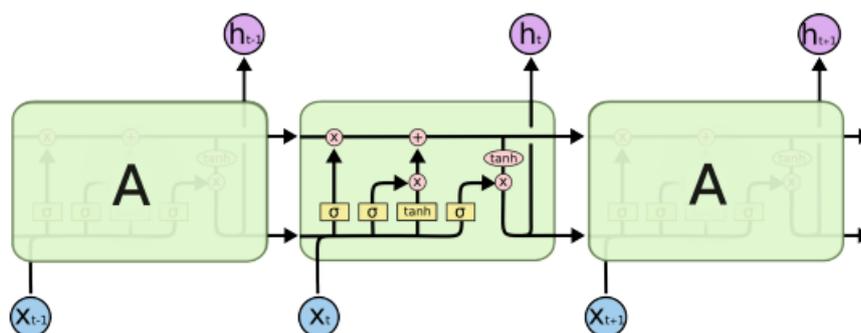
Figura 4 – Rede neural simples



Fonte: Olah (2015)

Uma LSTM é composta por blocos interligados, chamados de células (Figura 5). Cada célula contém três tipos de entrada: o Portão de Entrada (*Input Gate*), o Portão de Saída (*Output Gate*) e o Portão de Esquecimento (*Forget Gate*). Cada uma dessas entradas implementa funções específicas de escrita, leitura e restabelecimento da célula de memória. Essas entradas utilizam uma função sigmoideal de ativação e permitem que a LSTM aprenda dependências de longo prazo, evitando que informações importantes sejam esquecidas com o tempo (HOCHREITER; SCHMIDHUBER, 1997). Ao contrário das redes neurais tradicionais, que possuem uma cadeia de módulos repetidos com estruturas simples de camada única, as LSTMs possuem uma estrutura de cadeia com módulos que contêm quatro camadas de rede neural interagindo entre si (NELSON; PEREIRA; OLIVEIRA, 2017).

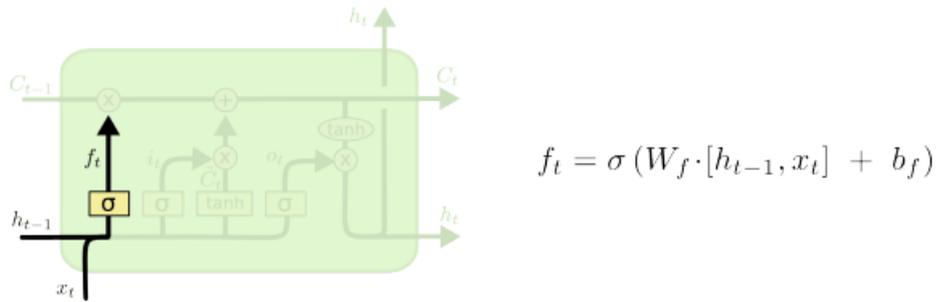
Figura 5 – LSTM



Fonte: Olah (2015)

O primeiro passo da LSTM é a determinação das informações que devem ser descartadas da célula de memória (HOCHREITER; SCHMIDHUBER, 1997). Essa tarefa é realizada pelo *Forget Gate*, que recebe como entrada o valor de saída do bloco anterior h_{t-1} e o valor de entrada do bloco atual x_t . Em seguida, gera como saída um valor entre 0 e 1 para cada entrada presente na célula de memória C_{t-1} do bloco anterior (Figura 6). Esses valores indicam a relevância de cada informação presente na célula de memória para a tarefa atual e são usados para atualizar a célula de memória C_t do bloco atual, mantendo somente as informações relevantes (LIPTON; BERKOWITZ; ELKAN, 2015).

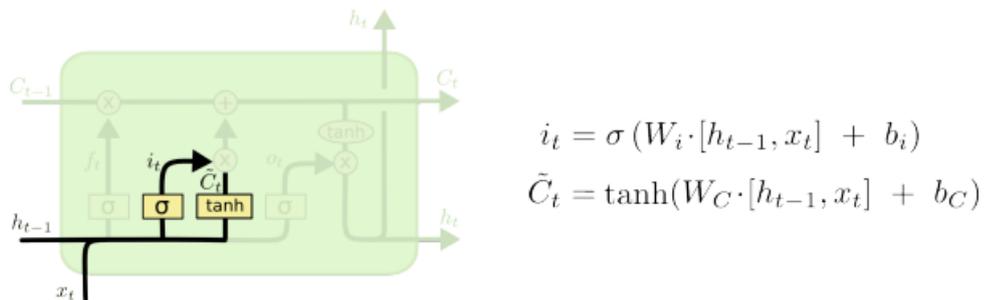
Figura 6 – Forget Gate



Fonte: Olah (2015)

O próximo passo da LSTM é decidir se a nova informação presente na entrada x_t deve ser armazenada na célula de memória. Esse processo é realizado em duas etapas: primeiro, a camada *Input Gate*, que utiliza a função sigmoideal de ativação, é responsável por determinar quais valores devem ser atualizados. Em seguida, a camada *Input Modulation Gate*, que utiliza a função tangente hiperbólica, gera um vetor de novos valores candidatos que podem ser adicionados à célula de memória (Figura 7).

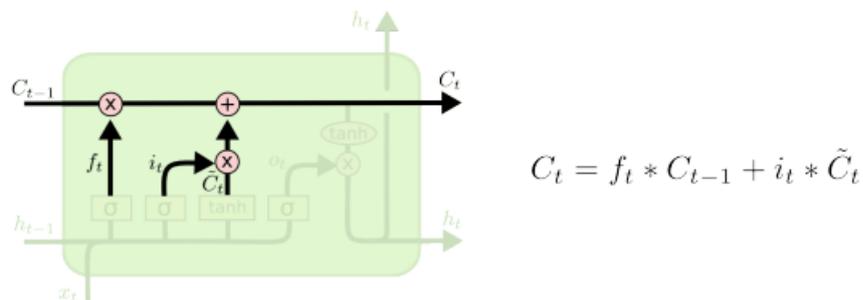
Figura 7 – Input Gate e Input Modulation Gate



Fonte: Olah (2015)

Após a determinação das informações relevantes pela etapa do *Forget Gate* e a decisão de quais valores serão adicionados à célula de memória pela etapa do *Input Gate*, é possível atualizar o estado da célula. Esse processo é realizado em duas etapas: primeiro, o estado anterior é modulado pela saída do *Forget Gate* para descartar as informações irrelevantes. Em seguida, os novos valores filtrados pela camada tangente são adicionados ao estado anterior, resultando no novo estado da célula de memória (Figura 8).

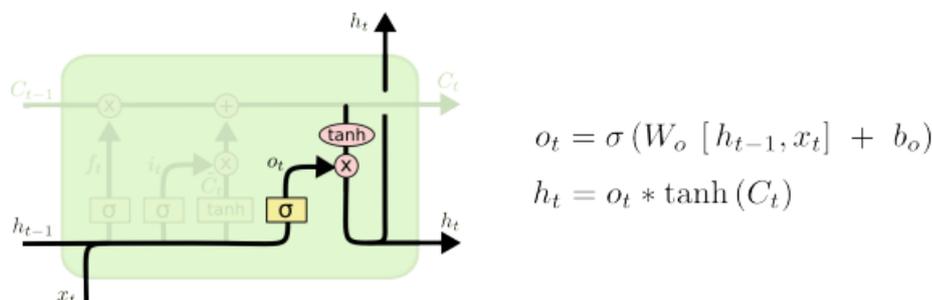
Figura 8 – Atualização da célula de memória



Fonte: Olah (2015)

A obtenção da saída final da rede LSTM envolve mais um passo de processamento na célula de memória. Primeiramente, a camada sigmoideal *Output Gate* é responsável por determinar qual parte da célula de memória deve ser utilizada na saída. Em seguida, a célula de memória é normalizada utilizando uma função tangente hiperbólica e multiplicada pela saída do *Output Gate*, de modo que a saída final seja apenas a parte relevante da célula de memória filtrada pela camada de saída (Figura 9). Esse processo de filtragem garante que a saída final esteja diretamente relacionada com as informações relevantes para a tarefa em questão, permitindo que a LSTM aprenda dependências de longo prazo e evite que informações importantes sejam perdidas ao longo do tempo.

Figura 9 – Saída final



$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

Fonte: Olah (2015)

2.6 Trabalhos Relacionados

Como descrito na Seção 2.2 o BTC é uma forma de dinheiro eletrônico que tem se popularizado como alternativa de investimento. Os investidores podem obter lucro ao comprar o ativo quando o seu preço estiver baixo e revendê-lo quando os preços aumentarem. No entanto, os preços flutuantes do BTC tornam sua previsão essencial para que os investidores tomem decisões mais informadas.

Nesse sentido, Séries Temporais tem sido amplamente utilizado como modelo de previsão, permitindo identificar padrões que ajudam a prever eventos futuros, como sugerido no estudo de [Ramadani e Devianto \(2020\)](#). Neste trabalho, os autores utilizaram o *Fuzzy Time Series Markov Chain*, o *Chen Logical*, e seus métodos segmentados devido a resultados de previsão desbalanceados. Como base, foram utilizados os dados de 2010 a 2020. Das abordagens propostas o método *Fuzzy Time Series Markov Chain* obteve o menor erro de precisão baseado no *Mean Absolute Percentage Error (MAPE)*, abaixo de 10%, quando comparado com o *Fuzzy Time Series Segmented Chen Logical Method* e o *Fuzzy Time Series Chen Logical Method*.

[McNally, Roche e Caton \(2018\)](#) optaram por descobrir com qual precisão a direção do preço do Bitcoin pode ser previsto. Para isso foram utilizados os modelos ARIMA, *Bayesian optimised recurrent neural network* e a LSTM. Destes, o LSTM obteve o melhor desempenho, alcançando o menor *Root Mean Squared Error (RMSE)* e a maior acurácia de classificação. Já com o objetivo de auxiliar na tomada de decisão de compra ou venda de ações, [Pires \(2018\)](#) optou por combinar técnicas a fim de tornar as decisões mais seguras. Para isso, foi utilizado padrões de *candle* para balizar as decisões da LSTM. No trabalho é proposto três estratégias de compra e venda de ações de três empresas. Na primeira abordagem foi utilizado apenas o resultado obtido através da previsão do modelo LSTM; na segunda, os padrões de *candle* só iriam influenciar, quando o resultado do modelo

indicasse uma compra; por fim, na terceira abordagem, os padrões influenciaram em todas as decisões, ou seja, incluindo os momentos em que a previsão do modelo indicava que a ação deve ser mantida em posse, ou mesmo que não deve ser comprada. Como conclusão, foi observado que as estratégias de compra que utilizaram os padrões de *candle*, em todos os casos, tenderam a ser mais seguras do que aquelas que não as utilizaram.

Kraaijeveld e Smedt (2020) buscaram identificar até que ponto o sentimento público do Twitter pode ser usado para prever os retornos de preços das principais criptomoedas, ativas no mercado. Para isso foi utilizada uma abordagem de análise de sentimento baseada em léxico específico para criptomoedas, dados financeiros e testes bilaterais de causalidade Granger. Descobriu-se que o sentimento do Twitter pode ser usado para prever os retornos de preços do Bitcoin, Bitcoin Cash e Litecoin. Os autores também constataram que os preditores mais fortes no nível diário são o sentimento obtido da rede social e o volume de mensagens; e o retorno dos preços é a variável preditora mais forte no nível intermediário. De forma semelhante, Dias (2021) adotou a abordagem de agregar índices de análise de sentimento, oriundos da mesma rede, com Conjuntos Nebulosos Hesitantes para a previsão de séries temporais no mercado financeiro. Foi empregado como algoritmo de *Machine Learning*, a *Weighted Multivariate Fuzzy Time Series (WMVFTS)*. Para os experimentos foram coletados os tweets postados pela Bloomberg e os preços de fechamento do *Standard & Poor's 500 (SP500)* Index e Nasdaq Composite Index. O autor concluiu que a principal característica entregue pelo método proposto foi a capacidade de melhorar o método Fuzzy Time Series (FTS) usando informações hesitantes, como as notícias postadas no Twitter.

Visando também a captura do sentimento, Huang, Rojas e Convery (2020) avaliaram as capacidades preditivas de dados de pesquisa na Internet, em específico o Google Trend, como métrica para a atenção dos investidores no mercado como um todo. Eles concluíram que os dados de pesquisa permitem construir um modelo para prever movimentos direcionais no índice SP500. Constatando que os dados podem, de fato, ser usados como sinais potenciais para o movimento do mercado de ações. E que o sinal direcional fornecido por um determinado índice de volume de pesquisa está condicionado à positividade ou negatividade do termo de pesquisa inicial. Aslanidis, Bariviera e López (2022) também adotaram o Google Trend em seu estudo, com o objetivo de descobrir a ligação existente entre estas e as criptomoedas. Foi encontrado um fluxo bidirecional de informações entre a atenção do Google Trends e retornos de criptomoedas em até seis dias; e identificada uma dependência significativa entre os retornos dos ativos analisados e o Google Trends, para as cinco criptomoedas analisadas e diferentes composições do índice das palavras-chave.

Por fim, na busca por ferramenta alternativa para modelar e analisar de sistemas complexos, Orang, Silva e Guimarães (2022) alcançaram resultados notáveis ao aplicar o *Fuzzy Cognitive Maps (FCMs)*. Estas apresentam semelhanças com redes neurais recorrentes e podem ser classificados como um método neuro-fuzzy. O que significa, segundo os

autores, que FCMs são uma mistura de lógica difusa, rede neural e sistema especialista, que atuam como uma ferramenta poderosa para simular e estudar o comportamento dinâmico de sistemas complexos. Onde as suas características mais interessantes incluem a interpretabilidade do conhecimento, características dinâmicas e capacidade de aprendizagem. O que os torna especialmente interessantes para a modelagem de sistemas complexos, como a flutuação do preço das criptomoedas.

2.7 Considerações finais

Neste capítulo descrevemos os principais conceitos para o entendimento do estudo, assim como os trabalhos relacionados a previsão de séries temporais. Sendo estes fundamentais para o entendimento do problema e das estratégias aplicadas. No Capítulo 3 é apresentado a técnica utilizada, bem como as discussões acerca da implementação.

3 Material e Método

Este capítulo apresenta uma breve contextualização sobre os dados e as técnicas utilizadas como base para a implementação dos algoritmos. O processo foi baseado na metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM), e passou pelas fases de: entendimento do problema; entendimento dos dados; pré-processamento dos dados; modelagem; e avaliação do resultado.

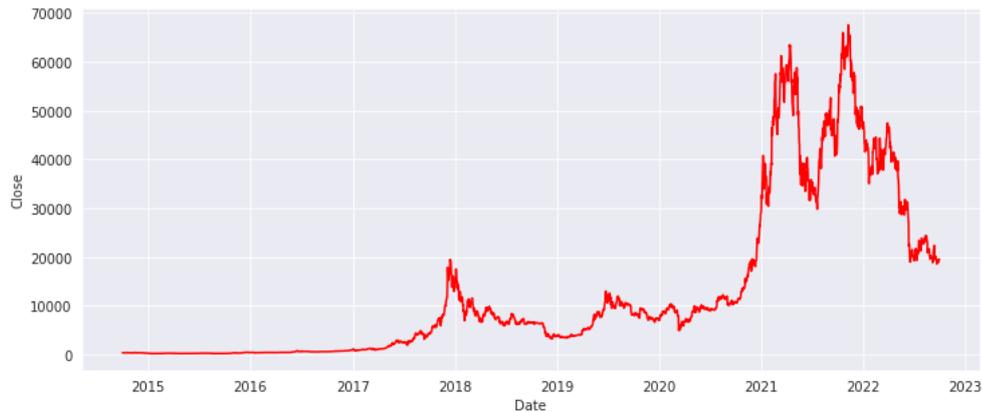
Na Seção 3.1 é descrito como os dados foram coletados, analisados e selecionados. A Seção 3.2 apresenta o desenvolvimento do método LSTM para previsão. Depois, é detalhado a definição do experimento na Seção 3.3. Por último, na Seção 3.4 é realizado a sintetização do que foi feito no capítulo.

3.1 Coleta de dados

3.1.1 Bitcoin

O conjunto de dados da cotação do preço do Bitcoin (BTC) foi obtido através da biblioteca *yFinance*, que utiliza dados obtidos através da API pública do Yahoo Finance. A amostra selecionada é do período de 01 de outubro de 2014 até 30 de setembro de 2022, totalizando 2.922 amostras, formando oito ciclos completos (oito anos). Esse conjunto é composto por registros diários contendo as seguintes informações: *Date* (Data do registro), *Open* (preço de abertura em US\$), *High* (preço de máxima em US\$), *Low* (preço de mínima em US\$), *Close* (preço de fechamento em US\$), e *Volume* (o valor total de US\$ do volume de negociação nas principais bolsas de Bitcoin). Abaixo é possível visualizar o preço de fechamento ao longo do tempo, na Figura 10.

Figura 10 – Preço de fechamento do Bitcoin no período analisado



Fonte: Elaborado pela autora

3.1.2 Google Trends

Para obter os dados da frequência normalizada da busca no Google, pelas palavras-chaves que estão mais relacionadas a economia do Bitcoin, identificadas via sites de notícias ¹, foi utilizado a biblioteca *PyTrends*. A amostra selecionada corresponde ao mesmo período de tempo, da amostra do BTC, também totalizando 2.922 amostras, por cada palavra obtida. As palavras selecionadas foram: Bitcoin, *Cryptocurrency*, *Inflation*, *Dollar*, *Metaverse*, *War*, *Global Crisis*, BTC, Blockchain e Crypto.

3.1.3 Definição da ABT

Com a finalidade de entendimento e seleção dos dados, para a modelagem, foi realizado análise descritiva, decomposição temporal, autocorrelação e correlação entre cada uma das *features*². Através desses processos pode-se observar que as séries são estocásticas, não estacionárias, possuem tendência de crescimento, sazonalidade e memória longa. Foi observado também que a série alvo, *Close*, é não linear, indicando que sofre influência de fatores externos, fortalecendo a necessidade da utilização de variáveis exógenas³ para a sua predição. Por fim, com o objetivo de auxiliar na seleção das *features*, foi realizado teste de causalidade de Granger, análise de correlação⁴, e identificação de multicolinearidade⁵. Por ser uma série temporal, técnicas mais avançadas de seleção de *features* foram descartadas.

¹ <https://www.cnnbrasil.com.br/economia/>, <https://www.infomoney.com.br/>

² Características ou atributos específicos

³ Séries auxiliares. Variáveis independentes que não são influenciadas por outras variáveis do modelo

⁴ Medida estatística que avalia o grau de associação ou relação entre duas variáveis

⁵ Fenômeno estatístico em análise de regressão e modelagem, onde duas ou mais variáveis independentes (preditoras) em um modelo estão altamente correlacionadas entre si

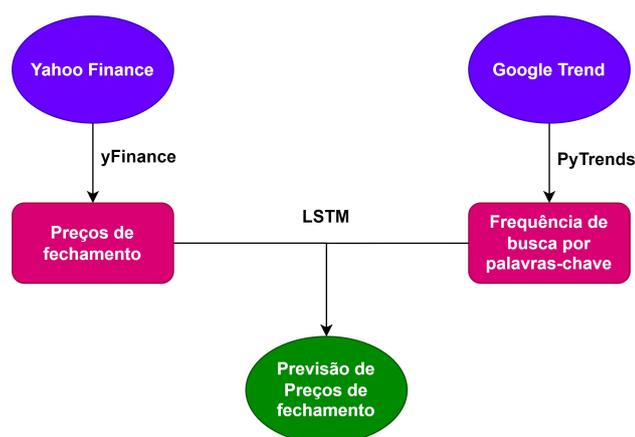
Optou-se por treinar os modelos com a série *Close*, a mesma a ser estimada, e com a frequência de busca normalizada das palavras “Crypto” e “Bitcoin“. Ambas foram aprovadas no teste de causalidade de Granger, e obtiveram resultado de correlação forte (0,9 e 0,7 respectivamente), quando comparadas com a série de valor de fechamento. Finalizando a Analytical Base Table (ABT)⁶ com as colunas: *Close*, Bitcoin e Crypto.

3.2 Implementação dos algoritmos

A implementação dos algoritmos foi realizada na linguagem Python 3.10, com o auxílio das bibliotecas `yFinance` e `PyTrends` para obtenção dos dados (Yahoo Finance, 2023; PyTrends, 2023); `Pandas` e `Numpy` para a manipulação (Pandas Development Team, 2023; NumPy Developers, 2023); `Matplotlib`, `Seaborn`, e `Plotly` para visualização (Matplotlib Development Team, 2023; WASKOM, 2023; Plotly Technologies Inc., 2023); `Scipy` e `Statsmodel` para testes estatísticos e decomposição temporal (SciPy Developers, 2023; Statsmodels Developers, 2023); `Sklearn` para normalização dos dados e implementação das métricas (Scikit-learn Developers, 2023); e `Keras` para a construção da rede neural LSTM (CHOLLET, 2023). O editor escolhido para a execução foi o Google Colaboratory, que fornece GPU para processamento, muito útil para algoritmos de Deep Learning (Google Colaboratory, 2023).

Na Figura 11 é apresentado o fluxograma da metodologia proposta.

Figura 11 – Fluxograma da metodologia proposta



Fonte: Elaborado pela autora

⁶ Tabela estruturada usada para construir modelos analíticos

3.2.1 Desenvolvimento da LSTM

A LSTM foi treinada com 4 camadas, sendo a primeira com 100 neurônios, as duas seguintes com 50 neurônios cada uma, e a última com 1 neurônio de saída. Foram definidos 100 épocas de treinamento. A função de ativação utilizada, estabelecida de forma empírica, foi a linear. O otimizador escolhido foi o RMSprop que divide a taxa de aprendizagem por uma média exponencialmente decrescente de gradientes quadrados tencionando otimizar o resultado. A função de perda adotada foi a *Mean Squared Error* (MSE); a métrica escolhida foi a *Mean Absolute Error* (MAE); o *batch size*⁷ foi 32; e o *dropout*⁸ foi definido, empiricamente, com 0,3 a fim de prevenir o *overfitting*⁹. A Tabela 1 apresenta as principais configurações do modelo.

Tabela 1 – Parâmetros da rede LSTM

Camada de Entrada
1 e 3 variáveis Janela: 90 e 10 dias
Camada Oculta
1 LSTM 100 neurônios 2 LSTMs 50 neurônios
Camada de saída
1 Dense - 1 neurônio Função de ativação: Linear
Compilador
Otimizador: RMSprop Função de perda: MSE Épocas: 100 Batch Size: 32

Fonte: Elaborado pela autora

Para otimizar o processo, também ocorreu a adição das funções: *EarlyStopping*, que força a parada do treinamento, quando não há mais melhoria do resultado; *ReduceLROn-Plateau*, que reduz a taxa de aprendizagem quando a métrica não apresenta melhora, a fim de que consiga provocar melhoria no valor resultante da métrica; e do *ModelCheckpoint*, que salva o melhor modelo, a partir do valor da métrica.

A escolha dos hiperparâmetros foram feitas de forma empírica. Devido ao custo computacional, foi descartado a fase de *tuning*¹⁰ do modelo, que pode vir a ser implementado em trabalhos futuros.

⁷ Número de exemplos de treinamento usados de uma só vez para ajustar os parâmetros de um modelo durante o treinamento

⁸ Técnica de regularização em redes neurais que ajuda a evitar *overfitting*, reduzindo a dependência excessiva de neurônios específicos durante o treinamento

⁹ Ocorre quando um modelo se ajusta muito bem aos dados de treinamento, mas não generaliza bem para novos dados, levando a resultados menos confiáveis

¹⁰ Processo automatizado para realização do ajuste dos hiperparâmetros do modelo, a fim de encontrar a melhor combinação

3.3 Definição do experimento

A amostra definida foi dividida em três conjuntos de dados, formando cenários diferentes a serem analisados, ficando da seguinte forma:

- 01/10/2014 a 30/09/2020, contendo os dados com maior estabilidade.
- 01/10/2014 a 08/11/2021, contendo os dados com tendência de alta.
- 01/10/2014 a 30/09/2022, contendo os dados com tendência de queda.

Na Figura 12 é possível observar a diferença de escala de valores no eixo vertical, em cada um dos conjuntos de dados separados para teste, utilizados no estudo.

Figura 12 – Escala de valores em cada cenário observado



Fonte: Elaborado pela autora

Apesar da série temporal apresentar autocorrelação de memória longa, optou-se por utilizar 10% dos dados para teste, em cada uma das investigações, correspondendo a pouco mais de seis meses disponíveis para cada cenário. Essa decisão foi tomada na tentativa de garantir que o modelo fosse capaz de capturar padrões e comportamentos passados da série, apesar da alta volatilidade do ativo.

Os dados foram normalizados com a função *MinMaxScaler* do *Sklearn* (Equação 3.1), em uma escala de 0 a 1, buscando obter uma melhor performance do algoritmo, fato comum em treinamentos de redes neurais. A janela de tempo definida para o treinamento foi de 90 e 10 dias. Para cada janela foram criados modelos univariados¹¹ (apenas a série *Close*), e multivariados¹² (com a adição das variáveis exógenas) totalizando doze modelos. Os resultados obtidos foram comparados com a SMA de 8 e 80 dias.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

3.3.1 Métricas de avaliação

As métricas utilizadas para avaliar a acurácia dos modelos criados foram a *R-Squared* (R^2), *Root Mean Squared Error* (**RMSE**), *Mean Absolute Error* (**MAE**), *Mean Absolute Percentage Error* (**MAPE**), *Mean Directional Accuracy* (**MDA**), e *Mean Forecast Error* (**MFE**).

O R^2 é uma métrica usada para medir o quão bem os valores previstos se ajustam aos valores observados (Equação 3.2). Ele fornece uma medida da proporção da variabilidade dos valores dependentes que é explicada pelas variáveis independentes do modelo. O seu valor varia entre 0 e 1, onde 0 indica que o modelo não explica nenhuma variabilidade dos dados e 1 indica que o modelo explica toda a variabilidade.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.2)$$

O RMSE é uma medida de desvio padrão dos erros de previsão, descrita na Equação 3.3. Quanto menor for o resultado, melhor é a precisão do modelo. A sua função é avaliar a magnitude do erro absoluto, sem levar em consideração a direção do erro (GEORGE; ANTONY; KRISHNAN, 2021). É uma medida difícil de interpretar por não ser normalizada, e sim estar na escala da medida analisada.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3.3)$$

¹¹ Modelo que envolve apenas uma única variável de interesse

¹² Modelo que abrange múltiplas variáveis de interesse simultaneamente

O MAE, detalhado na Equação 3.4, é uma medida da diferença média absoluta entre o valor real e o valor previsto. Quanto menor o seu valor, melhor é a precisão do modelo. A sua função é avaliar a magnitude do erro absoluto, sem levar em consideração a direção do erro. É uma medida de fácil interpretação, pois apresenta a média absoluta em dólares do erro, para mais ou para menos do preço real.

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (3.4)$$

O MAPE é uma medida da diferença média percentual entre o valor real e o valor previsto, detalhado na Equação 3.5. O seu valor é expresso em termos percentuais e é útil para avaliar a precisão do modelo em relação ao tamanho dos valores previstos. Quanto menor o MAPE, melhor é a precisão do modelo.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (3.5)$$

O MDA, detalhado na Equação 3.6, é uma medida da precisão de previsão de um modelo. Ela compara a direção prevista (para cima ou para baixo) com a direção real realizada. É uma métrica popular para prever o desempenho em economia e finanças (VISHWESH, 2023).

$$MDA = \frac{1}{N} \sum_{t=1}^N \mathbf{1}_{sgn(A_t - A_{t-1}) = sgn(\hat{A}_t - \hat{A}_{t-1})} \quad (3.6)$$

O MFE, detalhado na Equação 3.7, é a média dos erros de previsão. O seu valor pode ser positivo ou negativo, dependendo da direção do erro. Esta medida é útil para avaliar a precisão do modelo em termos da direção do erro.

$$MFE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t) \quad (3.7)$$

3.4 Considerações Finais

Este capítulo descreveu a obtenção dos dados e os métodos utilizados nesse estudo. Inicialmente, obtivemos os dados do BTC e dos indicadores de sentimentos. Em seguida, definimos a ABT utilizada no trabalho. Depois, aplicamos o método LSTM em cada um dos cenários definidos em janelas de tempo de curto e longo prazo. Por fim, apresentamos as métricas de avaliação utilizadas. Dessa forma, no Capítulo 4 apresentamos os resultados obtidos aplicando a sequência de técnicas aqui descritas.

4 Resultados e Discussão

Este capítulo apresenta os resultados e as análises dos doze modelos criados. Ao final, é realizada comparação entre estes modelos com a SMA com tendência de curto e longo prazo. Tendo o objetivo de verificar se a agregação da frequência normalizada das palavras-chaves selecionadas, em cada um dos cenários propostos, se mostra adequado a ser utilizado como ferramenta de auxílio em investimentos de criptomoedas.

Foram realizados experimentos em intervalos de tempo de 90 e 10 dias, com e sem a agregação das variáveis que contém os dados obtidos via o Google Trends. O ativo testado foi o Bitcoin, considerando os valores de fechamento.

As Seções 4.1, 4.2, 4.3 detalham os resultados dos modelos em cada um dos cenários estudados. Na Seção 4.4 é apresentado os valores obtidos pelas médias móveis. Depois, na Seção 4.5 é realizado a discussão. Por fim, na Seção 4.6 é efetivado o resumo do que foi trabalhado no capítulo.

4.1 Cenário 1 - Estabilidade

O primeiro cenário analisado corresponde ao período de 01/10/2014 a 30/09/2020, com 2.192 amostras. Sendo 1.972 amostras destinadas para treinamento, e 220 amostras para o teste. O desvio padrão¹ dos dados desta amostra corresponde a US\$ 4.152, quando analisado sua totalidade, e de US\$ 1.703, quando observado o conjunto de teste.

4.1.1 Janela de 90 dias

Na Figura 13 comparamos os resultados da predição, em uma janela de 90 dias, com e sem a agregação dos índices das palavras-chaves, oriundas do Google Trends.

A análise gráfica sugere que o modelo univariado apresentou uma capacidade de generalização maior, quando observado o padrão da curva ao longo da série temporal. No entanto, percebe-se que o modelo multivariado também acompanha o padrão de movimento da curva, e em alguns momentos obtém valores mais aproximados ao da série original.

¹ Medida estatística que quantifica a dispersão dos valores em relação à média de um conjunto de dados

Figura 13 – Cenário 1 - Janela de 90 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos



Fonte: Elaborado pela autora

4.1.2 Janela de 10 dias

A Figura 14 apresenta os resultados da predição em uma janela de 10 dias, com e sem a agregação dos índices que contém a frequência da busca pelas palavras-chaves ao longo do tempo. A figura sugere que os valores alcançados nos modelos, com e sem o indicador de sentimento, foi aproximado. A análise gráfica também indica que as séries resultantes do modelo apresentam um atraso na captura do movimento da curva.

Figura 14 – Cenário 1 - Janela de 10 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos



Fonte: Elaborado pela autora

4.1.3 Comparação dos resultados

A Tabela 2 traz os resultados dos modelos, quando aplicado as métricas de avaliação. Ela apresenta que os resultados alcançados na janela de tempo maior, obtiveram melhores resultados quando comparado o R^2 , RMSE e MAE. Os dados também sugerem que o treinamento em janelas de tempo mais curtas conseguem capturar melhor a direção do valor (alta ou queda), obtendo um valor melhor para o MDA. Os MFEs resultantes indicam que os modelos tenderam a prever valores abaixo do que foi realizado. O teste F^2 aponta que não há diferença estatística entre os experimentos realizados (p-valor³ de 0,96).

Tabela 2 – Cenário 1 - Resultado das métricas

Experimentos	R^2	RMSE	MAE	MAPE	MDA	MFE
Univariado - 90 dias	0,85	655,26	450,91	0,06	0,49	99,31
Multivariado - 90 dias	0,78	801,77	585,32	0,07	0,51	269,37
Univariado - 10 dias	0,81	738,11	492,42	0,06	0,51	72,32
Multivariado - 10 dias	0,78	803,10	599,98	0,07	0,54	296,30

Fonte: Elaborado pela autora

A Figura 15 ilustra a diferença do preço real com a do preço previsto pelo modelo, ou seja, o erro residual ao longo da série testada. Nesta figura identificamos que, em março de 2020, os modelos atingiram o maior erro residual ao longo da série testada. Mesmo período em que a Organização Mundial da Saúde (OMS) declarou a pandemia de coronavírus⁴. Também é apresentado que a partir de maio de 2020 os modelos univariados e multivariados obtiveram erros residuais semelhantes aos seus pares, independente da sua janela temporal.

² Teste estatístico que compara as médias de três ou mais grupos independentes para determinar se há diferenças significativas entre eles

³ Medida estatística que indica diferença entre os valores dos grupos observados. O nível de significância adotado neste trabalho foi de 0,05

⁴ <https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/11/oms-declara-pandemia-de-coronavirus.ghtml>

Figura 15 – Cenário 1 - Resíduo da diferença do valor real e do estimado



Fonte: Elaborado pela autora

4.2 Cenário 2 - Tendência de alta

O segundo cenário analisado corresponde ao período de 01/10/2014 a 08/11/2021, com 2.596 amostras. Sendo 2.336 amostras destinadas para treinamento, e 260 amostras para o teste. O desvio padrão dos dados desta amostra corresponde a US\$ 9.940, quando analisado sua totalidade, e de US\$ 9.722, quando observado o conjunto de teste.

4.2.1 Janela de 90 dias

A Figura 16 traz os resultados gerados pelos modelos em uma janela de 90 dias, no cenário de tendência de alta do preço do Bitcoin. A figura ilustra que os modelos criados tenderam a prever valores superiores ao da série original, e também sugere que eles apresentam atraso ao capturar o padrão de alta e baixa do preço do ativo.

Figura 16 – Cenário 2 - Janela de 90 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos



Fonte: Elaborado pela autora

4.2.2 Janela de 10 dias

Já na Figura 17 é ilustrado os resultados da predição em uma janela de 10 dias, com e sem a agregação dos índices que contém o indicador de sentimento. Percebe-se através da análise gráfica, que o resultado desse experimento, apresentou uma capacidade de ajuste melhor, do que a janela temporal mais longa. Pois sugere uma diferença menor com o preço real do BTC, embora apresente atraso na captura da tendência de movimento da série temporal.

Figura 17 – Cenário 2 - Janela de 10 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos



Fonte: Elaborado pela autora

4.2.3 Comparação dos resultados

Na Tabela 3 é descrito os valores das métricas de avaliação utilizadas. Ao contrário do observado no cenário de estabilidade, os resultados obtidos em uma janela de tempo menor, 10 dias, apresentou uma maior capacidade de generalização, quando comparado com a janela de 90 dias, observando todas as métricas propostas. Os resultados dos modelos univariados foram melhores, em todos os casos, quando comparado com a mesma janela de tempo, exceto o MFE obtido na janela de 90 dias. Os MFEs resultantes indicam que os modelos tenderam a prever valores acima do que foi realizado. O teste F sugere que o resultado das métricas não apresentam diferença estatística significativa (p-valor de 0,99).

Tabela 3 – Cenário 2 - Resultado das métricas

Experimentos	R²	RMSE	MAE	MAPE	MDA	MFE
Univariado - 90 dias	0,51	6.790,74	5.465,99	0,12	0,49	-4.713,85
Multivariado - 90 dias	0,41	7.459,55	5.425,55	0,11	0,46	-3.980,52
Univariado - 10 dias	0,74	4.910,55	3.772,22	0,09	0,49	-2.156,44
Multivariado - 10 dias	0,63	5.877,67	4.537,21	0,10	0,49	-3.356,13

Fonte: Elaborado pela autora

A Figura 18 traz o erro residual ao longo da série observada. Nesta figura fica visível que em maio de 2021 ocorreu o maior erro residual, chegando a prever um valor superior a US\$ 20.000, quando comparado ao preço real. Época em que a Tesla recuou na decisão de aceitar BTC na venda dos seus carros elétricos; e o banco central da China e a *U.S. Securities and Exchange Commission (SEC)* ameaçaram a não autorizar novas modalidades de aplicação em criptomoedas⁵.

Figura 18 – Cenário 2 - Resíduo da diferença do valor real e do estimado



Fonte: Elaborado pela autora

4.3 Cenário 3 - Tendência de queda

O terceiro cenário analisado corresponde ao período de 01/10/2014 a 30/09/2022, com 2.922 amostras. Sendo 2.629 amostras destinadas para treinamento, e 293 amostras para o teste. O desvio padrão dos dados desta amostra corresponde a US\$ 16.425, quando analisado sua totalidade, e de US\$ 10.176, quando observado o conjunto de teste. Os resultados obtidos com e sem a adição da variável exógena, pode ser observado abaixo.

4.3.1 Janela de 90 dias

Podemos observar na Figura 19 os resultados gerados pelos modelos em uma janela de 90 dias, no cenário de tendência de queda do preço do ativo analisado. A imagem sugere que o modelo multivariado apresentou resultados semelhantes ao univariado, indicando também atrasos na identificação da mudança da tendência.

⁵ <https://valorinveste.globo.com/mercados/cripto/noticia/2021/05/31/bitcoin-sangrou-em-maio-mas-ainda-sobe-31percent-em-2021-compare-opcoes-para-investir.ghtml>

Figura 19 – Cenário 3 - Janela de 90 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos



Fonte: Elaborado pela autora

4.3.2 Janela de 10 dias

Ao comparar os resultados obtidos em uma janela de tempo de 10 dias, na Figura 20, observamos que o padrão da série se assemelha ao alcançado pela janela de 90 dias. A figura sugere atraso na captura do movimento de alta e queda do ativo.

Figura 20 – Cenário 3 - Janela de 10 dias - Preço de fechamento do Bitcoin real e os estimados pelos modelos



Fonte: Elaborado pela autora

4.3.3 Comparação dos resultados

Na Tabela 4 observamos os resultados alcançados com os experimentos. Ao analisar a janela temporal de 10 dias, percebemos que os valores atingidos pelo modelo multivariado foram melhores quando comparado as métricas R^2 , RMSE, MAE, MAPE e MFE. Juntamente percebemos que os valores obtidos em uma janela temporal longa, pelo modelo univariado, foi superior quando comparado com os outros modelos utilizados, com exceção da métrica MDA e MFE. Que apresentaram a capacidade de acerto da direção menor, quando comparado com o modelo multivariado, na mesma janela de tempo. Os MFEs sugerem que os modelos tenderam a prever valores acima do que foi realizado. O teste F indica que os resultados alcançados pelas métricas não apresentaram diferença estatística significativa (p-valor de 0,99).

Tabela 4 – Cenário 3 - Resultado das métricas

Experimentos	R^2	RMSE	MAE	MAPE	MDA	MFE
Univariado - 90 dias	0,94	2.430,97	1.788,57	0,06	0,47	-758,69
Multivariado - 90 dias	0,92	2.943,23	2.123,65	0,07	0,49	-673,82
Univariado - 10 dias	0,89	3.381,05	2.641,08	0,09	0,49	-2.109,16
Multivariado - 10 dias	0,91	3.005,92	2.199,79	0,07	0,48	-925,78

Fonte: Elaborado pela autora

Na figura 21 podemos visualizar o erro residual ao longo da série observada. A imagem ilustra que em fevereiro de 2022 os modelos tenderam a prever valores abaixo do realizado. Já entre maio e junho, a mais do que foi realmente atingido. Neste período, as principais economias mundiais sofreram com o aumento da inflação⁶, o que acabou refletindo também no mercado dos ativos digitais.

⁶ <https://www.poder360.com.br/economia/bitcoin-cai-54-em-2022-entenda-mudancas-no-mundo-cripto/>

Figura 21 – Cenário 3 - Resíduo da diferença do valor real e do estimado



Fonte: Elaborado pela autora

4.4 Média Móvel

Os resultados obtidos pelos modelos implementados apresentaram valores discrepantes quando observado os valores obtidos ao longo da série temporal do erro residual, ilustradas nas Figuras 15, 18 e 21. E também quando observado os valores obtidos pelas métricas de avaliação utilizadas neste estudo, sobretudo em cenários de tendência de alta.

Com o objetivo de comparar os resultados com um modelo ingênuo ⁷, as mesmas métricas de avaliação foram aplicadas aos valores obtidos pela SMA de 8 e 80 dias ⁸. Números definidos de forma empírica, por capturar tendências de curto e longo prazo nos preços dos ativos.

Na Tabela 5 é detalhado os resultados obtidos nos diferentes cenários trabalhados. Como nos outros experimentos (Seções 4.1, 4.2, 4.3), o teste F sugere que não há diferença estatisticamente significativa entre os grupos (p-valor de 0,25).

⁷ Abordagem simples para resolver um problema ou prever algo, onde se assume uma regra básica sem considerar muitos detalhes ou fatores complicados

⁸ Representadas neste trabalho como MM8 e MM80, respectivamente

Tabela 5 – Média Móvel - Resultado das métricas

Experimentos	R²	RMSE	MAE	MAPE	MDA	MFE
Cenário 1 - MM8	0,92	483,39	324,35	0,04	0,60	16,14
Cenário 1 - MM80	0,33	1.393,53	1.110,29	0,13	0,51	349,25
Cenário 2 - MM8	0,92	2.743,13	2.137,23	0,05	0,63	120,32
Cenário 2 - MM80	-0,18	10.534,62	9.438,35	0,21	0,51	2.617,17
Cenário 3 - MM8	0,97	1.731,73	1.261,63	0,04	0,59	-354,42
Cenário 3 - MM80	0,53	6.994,21	5.572,38	0,18	0,54	-4.953,23

Fonte: Elaborado pela autora

Nota-se que os valores obtidos pela MM8 é superior quando comparado com todos os resultados alcançados pelos modelos utilizados neste estudo, destacando os valores da RMSE, MAE, MDA e MFE. Já a MM80 apresenta resultado inferior na maioria dos casos, desconsiderando quando comparado o valor atingido pela MDA destes modelos, com as obtidas pelo método LSTM. Sugerindo que, considerando o contexto experimental estabelecido, a tendência de curto prazo tende a ser um melhor indicador para a inferência da predição do ativo. No Apêndice A pode-se visualizar a análise gráfica dos resultados alcançados, quando observado o SMA das tendências observadas.

4.5 Discussão

Por meio dos experimentos podemos observar a capacidade de generalização dos modelos em cada cenário estudado. Apesar dos modelos univariados apresentarem um desempenho melhor, em alguns momentos os algoritmos multivariados atingiram uma capacidade de generalização superior. Outro ponto interessante foi que no cenário de tendência de alta do preço do ativo, janelas de tempo menores atingirem melhores resultados.

O estudo indica que a adição das variáveis exógenas podem proporcionar resultados melhores em métricas como o MDA e MFE. Sinalizando que a sua adição, combinada também com outras técnicas, podem vir a melhorar a precisão da predição dos ativos. Apesar dos valores atingidos pelo MDA não ser satisfatório nos cenários com tendência de alta e queda, pois ficaram abaixo de 0,5.

Ao comparar os resultados obtidos em cada cenário pelos modelos LSTM com a SMA, é sugerido que o método de RNN não supera os valores obtidos pela MM8, e apresenta o MDA inferior quando também comparada com a MM80.

4.6 Considerações finais

Neste capítulo foram apresentados os resultados com os experimentos realizados com a metodologia proposta. No próximo capítulo serão expostas as conclusões identificadas após o desenvolvimento deste estudo, bem como as limitações da metodologia e propostas de trabalhos futuros.

5 Conclusão

Este estudo trata-se de uma pesquisa empírica que busca estimar o comportamento do Bitcoin e inferir seu valor futuro, por meio da criação de uma rede neural recorrente, a LSTM, com a agregação de variáveis exógenas que contêm a frequência de busca por palavras-chave no Google. Para atingir esse objetivo, foram realizados experimentos para avaliar a eficácia da abordagem em diferentes cenários.

Por meio de revisão bibliográfica foram identificadas abordagens correlatas para a questão de predição de dados, essenciais para a proposta do algoritmo e do conjunto de dados selecionados. A partir disso, foram analisadas as técnicas mais adequadas de acordo com o desempenho, complexidade e tecnologia utilizada nos métodos.

Por intermédio das métricas utilizadas foi possível observar comportamentos diferentes nos experimentos realizados, sendo notado um melhor desempenho no conjunto de dados com menor desvio padrão. Os indicadores também apresentaram uma maior capacidade de generalização dos modelos ao agregar o indicador de sentimento, ao se observar as métricas MDA, MFE e R^2 , sobretudo no cenário de tendência de queda. Apesar disso, durante os experimentos notou-se que os modelos implementados não superaram os resultados obtidos pela média móvel simples de 8 dias, sugerindo haver possibilidade de melhorias.

No mais, os objetivos do estudo foram alcançados obtendo-se os resultados almejados na análise e predição da criptomoeda selecionada. Além de inúmeros aprendizados e propostas de melhorias para os algoritmos utilizados e abordagens diferentes que podem ser exploradas.

É relevante enfatizar que após análises verificou-se que os algoritmos desenvolvidos possuem escalabilidade para serem aplicados em conjuntos de dados mais atualizados e em outras moedas, sendo previsto uma performance comparável à obtida neste trabalho. Desse modo, esta pesquisa contribui para futuros estudos acerca da análise preditiva de moedas e seu valor no mercado financeiro.

5.1 Limitações da metodologia proposta

Considerando os resultados obtidos, é importante ressaltar que há algumas limitações identificadas no estudo. Embora tenha sido observado um desempenho satisfatório ao utilizar as variáveis exógenas no cenário de queda¹, o mesmo não ocorreu no cenário de valorização do ativo. Sinalizando a necessidade de explorar novas combinações e técnicas para capturar esses padrões de comportamento. Ressalta-se também que apesar de ter ocorrido bons resultados com a adição das variáveis exógenas, os modelos univariados, obtiveram melhor desempenho na maioria dos casos.

Além disso, é pertinente destacar que, devido às limitações do escopo do estudo, não foi possível comparar os resultados obtidos com os de outros pesquisadores, limitando-se apenas aos resultados gerados nesta pesquisa. Ademais, a implementação de técnicas mais robustas para a criação dos modelos também foram descartadas neste momento do projeto, deixando margem para ser explorada em trabalhos futuros.

5.2 Propostas de trabalhos futuros

Para trabalhos futuros, é interessante considerar a inclusão de novos fatores para entrada dos dados, como médias móveis de longo e curto prazo, volume de negociação, além de retornos sobre o investimento. A implementação de outros algoritmos comumente usados para previsão financeira, como o *Poisson Hawkes* e *WMVFTS*, pode gerar resultados interessantes, tanto em uma abordagem híbrida quanto em uma abordagem pura. Além disso, a combinação de outras técnicas, como o processamento de linguagem natural para identificar o sentimento do mercado, pode proporcionar uma capacidade de generalização maior para estimar os preços futuros do ativo, contribuindo para que o usuário final alcance melhores resultados financeiros e minimize os riscos da operação.

¹ Em uma janela de tempo mais curta

Referências

- ASLANIDIS, N.; BARIVIERA, A. F.; LÓPEZ, Ó. G. The link between cryptocurrencies and google trends attention. *Finance Research Letters*, Elsevier, v. 47, p. 102654, 2022. Citado 2 vezes nas páginas 22 e 32.
- CAETANO, M. A. L. *Python e mercado financeiro : programação para estudantes, investidores e analistas*. [S.l.]: Blucher, 2021. Citado na página 22.
- CHATFIELD, C.; XING, H. *The analysis of time series: an introduction with R*. [S.l.]: CRC press, 2019. Citado na página 24.
- CHEN, K.; ZHOU, Y.; DAI, F. A lstm-based method for stock returns prediction: A case study of china stock market. In: IEEE. *2015 IEEE international conference on big data (big data)*. [S.l.], 2015. p. 2823–2824. Citado na página 27.
- CHEN, Z.; LI, C.; SUN, W. Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, Elsevier, v. 365, p. 112395, 2020. Citado na página 24.
- CHOLLET, F. *Keras: The Python deep learning API*. 2023. Disponível em: <<https://keras.io>>. Citado na página 36.
- DIAS, B. C. D. Agregação de índices de análise de sentimentos com conjuntos nebulosos hesitantes para previsão de séries temporais no mercado financeiro. 2021. Citado na página 32.
- ENDERS, W. Applied econometric time series fourth edition. *New York (US): University of Alabama*, 2015. Citado na página 24.
- GEORGE, S. E.; ANTONY, S.; KRISHNAN, R. Evaluation of machine learning models for demand forecasting in supply chain management: A review. 2021. Citado na página 39.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016. Citado na página 25.
- Google Colaboratory. *Google Colaboratory: A free Jupyter notebook environment that includes GPU support*. 2023. Disponível em: <<https://colab.research.google.com>>. Citado na página 36.
- GRANGER, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, JSTOR, p. 424–438, 1969. Citado na página 25.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. 1997. Citado na página 28.
- HUANG, M. Y.; ROJAS, R. R.; CONVERY, P. D. Forecasting stock market movements using google trend searches. *Empirical Economics*, Springer, v. 59, p. 2821–2839, 2020. Citado na página 32.

- HULL, J. C. *Opções, futuros e outros derivativos*. [S.l.]: Bookman Editora, 2016. Citado na página 20.
- KLIOUTCHNIKOV, I.; SIGOVA, M.; BEIZEROV, N. Chaos theory in finance. *Procedia computer science*, Elsevier, v. 119, p. 368–375, 2017. Citado na página 23.
- KRAALJEVELD, O.; SMEDT, J. D. The predictive power of public twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, Elsevier, v. 65, p. 101188, 2020. Citado na página 32.
- LIPTON, Z. C.; BERKOWITZ, J.; ELKAN, C. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015. Citado na página 28.
- Matplotlib Development Team. *Matplotlib: A 2D plotting library*. 2023. Disponível em: <<https://matplotlib.org>>. Citado na página 36.
- MCNALLY, S.; ROCHE, J.; CATON, S. Predicting the price of bitcoin using machine learning. In: IEEE. *2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP)*. [S.l.], 2018. p. 339–343. Citado na página 31.
- MEDSKER, L.; JAIN, L. C. *Recurrent neural networks: design and applications*. [S.l.]: CRC press, 1999. Citado na página 27.
- MENDENHALL, W.; BEAVER, R.; BEAVER, B. *Introduction to probability and statistics. 10-th edition*. [S.l.]: California, USA: Duxbury Press, 1999. Citado na página 23.
- MOAT, H. S. et al. Quantifying wikipedia usage patterns before stock market moves. *Scientific reports*, Nature Publishing Group UK London, v. 3, n. 1, p. 1801, 2013. Citado na página 24.
- MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: *Análise de séries temporais*. [S.l.: s.n.], 2006. p. 538–538. Citado na página 23.
- MURPHY, J. J. *Technical Analysis of the Futures Markets: A Comprehensive Guide to Trading Methods and Applications, New York Institute of Finance*. [S.l.]: Prentice-Hall, 1999. Citado na página 20.
- NAKAMOTO, S. Bitcoin: A peer-to-peer electronic cash system. *Decentralized business review*, p. 21260, 2008. Citado 2 vezes nas páginas 17 e 21.
- NELSON, D. M.; PEREIRA, A. C.; OLIVEIRA, R. A. D. Stock market's price movement prediction with lstm neural networks. In: IEEE. *2017 International joint conference on neural networks (IJCNN)*. [S.l.], 2017. p. 1419–1426. Citado 2 vezes nas páginas 27 e 28.
- NIELSEN, A. *Practical time series analysis: Prediction with statistics and machine learning*. [S.l.]: O'Reilly Media, 2019. Citado na página 26.
- NumPy Developers. *NumPy: fundamental package for scientific computing with Python*. 2023. Disponível em: <<https://numpy.org>>. Citado na página 36.
- OLAH, C. Understanding lstm networks. 2015. Citado 5 vezes nas páginas 27, 28, 29, 30 e 31.

ORANG, O.; SILVA, P. C. de Lima e; GUIMARÃES, F. G. Time series forecasting using fuzzy cognitive maps: a survey. *Artificial Intelligence Review*, Springer, p. 1–62, 2022. Citado na página 32.

Pandas Development Team. *Pandas: Powerful data structures for data analysis, time series, and statistics*. 2023. Disponível em: <<https://pandas.pydata.org>>. Citado na página 36.

PATEL, M. M. et al. A deep learning-based cryptocurrency price prediction scheme for financial institutions. *Journal of information security and applications*, Elsevier, v. 55, p. 102583, 2020. Citado na página 17.

PIRES, V. T. d. C. Redes neurais lstm e padrões de candle aplicados no auxílio à tomada de decisão em aplicações financeiras. 2018. Citado na página 31.

Plotly Technologies Inc. *Plotly: Open-source graphing library*. 2023. Disponível em: <<https://plotly.com>>. Citado na página 36.

Pytrends. *Pytrends: Google Trends API for Python*. 2023. Disponível em: <<https://github.com/GeneralMills/pytrends>>. Citado na página 36.

RADZIWILL, N. Blockchain revolution: How the technology behind bitcoin is changing money, business, and the world. *The Quality Management Journal*, Taylor & Francis Ltd., v. 25, n. 1, p. 64–65, 2018. Citado na página 21.

RAMADANI, K.; DEVIANTO, D. The forecasting model of bitcoin price with fuzzy time series markov chain and chen logical method. In: AIP PUBLISHING LLC. *AIP Conference Proceedings*. [S.l.], 2020. v. 2296, n. 1, p. 020095. Citado na página 31.

Scikit-learn Developers. *scikit-learn: Machine learning in Python*. 2023. Disponível em: <<https://scikit-learn.org>>. Citado na página 36.

SciPy Developers. *SciPy: Open-source software for mathematics, science, and engineering*. 2023. Disponível em: <<https://www.scipy.org>>. Citado na página 36.

SILVA, A. L. A.; GUIMARÃES, T. B.; TAVARES, A. O. Séries temporais: teoria e aplicações. 2008. Citado na página 23.

SOARES, P. L. B.; SILVA, J. P. da. Aplicação de redes neurais artificiais em conjunto com o método vetorial da propagação de feixes na análise de um acoplador direcional baseado em fibra ótica. *Revista Brasileira de Computação Aplicada*, v. 3, n. 2, p. 58–72, 2011. Citado 2 vezes nas páginas 25 e 26.

SOVBETOV, Y. Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero. *Journal of Economics and Financial Analysis*, v. 2, n. 2, p. 1–27, 2018. Citado na página 17.

Statsmodels Developers. *Statsmodels: Econometric and statistical modeling with Python*. 2023. Disponível em: <<https://www.statsmodels.org>>. Citado na página 36.

SWAN, M. *Blockchain: Blueprint for a new economy*. [S.l.]: "O'Reilly Media, Inc.", 2015. Citado na página 21.

TAFNER, M. A. Redes neurais artificiais: aprendizado e plasticidade. *Cérebro & Mente*, v. 2, n. 5, 1998. Citado na página 26.

VISHWESH, K. Mean directional accuracy of time series forecast. 2023. Citado na página 40.

WASKOM, M. *Seaborn: statistical data visualization*. 2023. Disponível em: <<https://seaborn.pydata.org>>. Citado na página 36.

Yahoo Finance. *yFinance: Yahoo Finance market data downloader*. 2023. Disponível em: <<https://github.com/ranaroussi/yfinance>>. Citado na página 36.

Apêndices

APÊNDICE A – Análise gráfica da média móvel

Neste apêndice se encontra os gráficos gerados ao comparar a série de fechamento (*Close*) do Bitcoin, com as médias móveis simples de 8 e 80 dias. Também é possível equiparar com os resultados obtidos pelos modelos em cada cenário analisado.

A.1 Cenário 1 - Estabilidade

As Figuras 22 e 23 apresentam os valores da série real contrastada com os valores obtidos pelas MMs, e também pelos modelos utilizados no cenário mais estável. Já na Figura 24 observamos o resíduo das MMs quando diferenciado do valor de fechamento real.

Figura 22 – Cenário 1 - Janela de 10 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis



Fonte: Elaborado pela autora

Figura 23 – Cenário 1 - Janela de 90 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis



Fonte: Elaborado pela autora

Figura 24 – Cenário 1 - Resíduo da diferença do valor real e das médias móveis



Fonte: Elaborado pela autora

A.2 Cenário 2 - Tendência de alta

Como na seção anterior, a Figura 25 e a Figura 26 apresentam os valores da série real contrastada com os valores obtidos pelas MMs, e também pelos modelos utilizados no cenário com tendência de alta. Na Figura 27 exibimos o resíduo das MMs quando diferenciado do valor de fechamento real.

Figura 25 – Cenário 2 - Janela de 10 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis



Fonte: Elaborado pela autora

Figura 26 – Cenário 2 - Janela de 90 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis



Fonte: Elaborado pela autora

Figura 27 – Cenário 2 - Resíduo da diferença do valor real e das médias móveis



Fonte: Elaborado pela autora

A.3 Cenário 3 - Tendência de queda

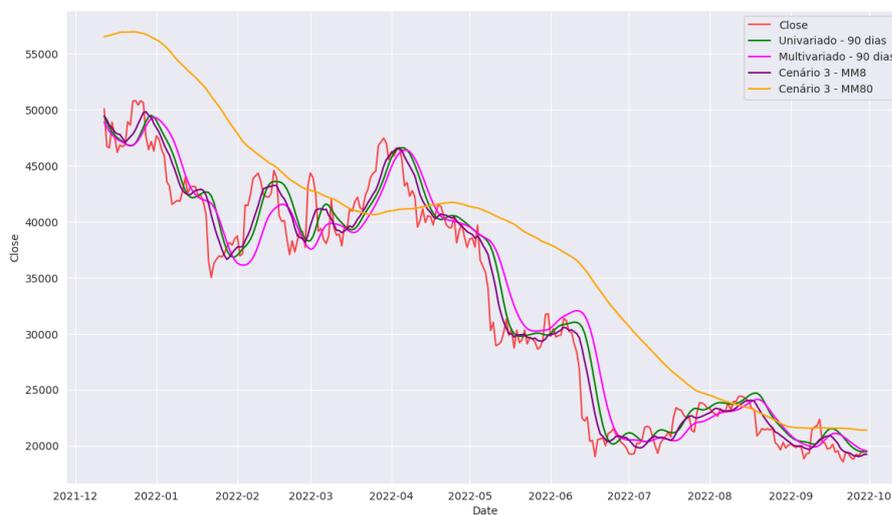
Por fim, como nas outras seções, a Figura 28 e a Figura 29 apresentam os valores da série real contrastada com os valores obtidos pelas MMs, e também pelos modelos utilizados no cenário com tendência de queda. Na Figura 30 ilustramos o resíduo das MMs quando diferenciado do valor de fechamento real.

Figura 28 – Cenário 3 - Janela de 10 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis



Fonte: Elaborado pela autora

Figura 29 – Cenário 3 - Janela de 90 dias - Preço de fechamento do Bitcoin real comparado com os estimados pelo modelo e médias móveis



Fonte: Elaborado pela autora

Figura 30 – Cenário 3 - Resíduo da diferença do valor real e das médias móveis



Fonte: Elaborado pela autora