



Predicting Bitcoin Cryptocurrency Price Behavior based on ARIMA and NNAR modelling

Patrícia Virgínia de Santana Lima^{1*} 

David Venâncio da Cruz² 

Albaro Ramon Paiva Sanz³ 

¹²³ Departamento de Estatística, State University of Paraíba, Campina Grande, Brazil

Emails: ¹ patricia.lima@aluno.uepb.edu.br; ² davidvenacio4@gmail.com; ³ albaro.sanz@servidor.uepb.edu.br

* Corresponding author

How to cite this paper: Lima, P. V. de Santana, da Cruz, D. V., Sanz, A. R. P. (2024). Predicting Bitcoin Cryptocurrency Price Behavior based on ARIMA and NNAR modelling. *Socioeconomic Analytics*, 2(1), 121-129. <https://doi.org/10.51359/2965-4661.2024.265073>

RESEARCH ARTICLE

Socioeconomic Analytics
<https://periodicos.ufpe.br/revistas/SECAN/>
ISSN Online: 2965-4661

Submitted on: 11.11.2024.

Accepted on: 15.12.2024.

Published on: 27.12.2024.

Copyright © 2024 by author(s).

This work is licensed under the Creative Commons Attribution International License CC BY-NC-ND 4.0
<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>



Abstract

In this work, we develop specific models to predict the behavior of the Bitcoin cryptocurrency using a public database (Yahoo! Finance) to track price trends. The methodologies used are ARIMA and NNAR, and the validation of the models is carried out based on the daily closing values of assets. Both models fail to differ significantly. However, the adjusted model NNAR (2.2) fits slightly better with the original data series, presenting an MPE (Mean Percentage Error) of -0.102. Some prospects on the seasonality of data are discussed.

Keywords

prediction, Time Series Analysis, Bitcoin, forecast. ARIMA, NNAR.

1. Introdução

A criptomoeda Bitcoin que surgiu no ano de 2008, criada por anônimo denominado de Satoshi Nakamoto (Ulrich, 2004), com a publicação do seguinte trabalho “Bitcoin: A Peer-to-Peer Electronic Cash System” com o principal objetivo descentralização do mercado gerando independência da intervenção governamental nas transações. Diante de um vasto crescimento de investidores, cada vez aumenta a busca por ferramentas para análise desse ativo. Com isso muitos investidores tomam decisões através de indicadores como: Fibonacci, padrão de velas e modelos de previsão a exemplo apresentado neste trabalho (NAKAMOTO, 2009).

Desde de sua criação, vários trabalhos vem sendo desenvolvidos dentro da ciência de dados na tentativa de prever o comportamento das criptomoedas. Em 2018, Velankar, Valecha e Maji propuseram métodos para a previsão do Bitcoin com base dados da CoinMarketCap, depois de uma filtragem dos dados aplicou-se Regressão Bayesiana e Modelos lineares generalizados. Com relação as redes neurais recorrentes MCNally (2016) faz a comparação entre RNN (Rede neural bayesiana otimizada) e LSTM (Long Short Term Memory) em que LSTM supera o RNN em relação as medidas de ajuste, entretanto sem diferença significativa.

Nesta área especifica de previsão de criptomoedas ainda temos, Kim et al (2016) que propôs modelos de previsão baseados nos comentários e repostas divulgadas em comunidades on-line relacionadas as criptomoedas, de forma específicas nos fóruns de Bitcoins, Ethereum e Ripple, valendo que os comentários dos participantes influenciavam nas flutuações dos preços e volume das negociações das criptomoedas. Isto realizado por meio de um algoritmo denominado de VADER, os comentários dos participantes são filtrados, realizado uma análise estatística pelo sentimento que transmitem (HUTTO, GILBERTO, 2014), excluindo comentários considerados extremos positivos ou negativos, o resultado deste trabalho mostrou-se positivo para a ideia trabalhada, que a opinião dos usuários pode ser utilizada para prevê flutuações deste mercado, entretanto os autores recomendam critérios mais qualitativos para a previsão.

Os trabalhos mais recentes utilizando redes neurais para a predição de Ether, tem-se Duarte e Lima (2019), que compara duas redes neurais Multilayer perceptrons (MLP) e Bidirectional Long Short Term Memory Units (bi-LSTM), e conseguiram como resultado o bi-LSTM com um resultado um pouco melhor que MLP.

Com os modelos de séries temporais especialmente ARIMA, Azari 2009, propôs modelos com vários parâmetros para uma série com período de 3 anos e series diárias, que concluiu que para series longas os modelos não conseguiram “captar” todas as oscilações, pois o mercado de criptomoedas ainda é relativamente novo, ou seja, “emergente”, o que torna sensível as especulações, notícias, crises políticas, econômicas e militares. O que torna um mercado de alta volatilidade. De acordo com NuBank entre 2009 e 2011, a bitcoin obteve uma valorização impressionante de 30 mil por cento. Logo quanto menor o lag, “janela de tempo” melhor será a previsão de preços.

De acordo com Silva (2016), o maior problema da bitcoin e volatilidade. A mesma como

forma de contar a flutuação de preços tem um impacto profundo nas estratégias de investimento, ou seja, as series geralmente apresentam observações extremas e reações assimétricas e voláteis comparadas ao passado.

A utilização deste mercado tendo em vista a possibilidade de investimento e por fim de retorno financeiro, entretanto deve-se levar em conta que estes tipos de operações são de grade risco, alerta Radityo, Munajat e Budi (2017), que devido a alta volatilidade se faz necessário ferramentas de previsão para auxiliar possíveis investidores na tomada de decisão no câmbio de criptomoedas.

O presente trabalho tem como finalidade aplicar os modelos ARIMA e NNAR sobre os dados de criptomoedas, verificando os resíduos gerados pelos modelos ajustados por fim obter previsões de preços do Bitcoin, analisando os erros identificar qual o modelo obteve melhor resultado.

2. Material e métodos

A base de dados analisada, foi disponibilizada pela Yahoo Finance contendo 8430 linhas e 6 colunas, com valores diários com os seguintes atributos: menor valor atingido, maior valor atingido, abertura (valor inicial), fechamento (último valor atingido), ajustado e volume de negociações da criptomoeda Bitcoin, compreendido entre o período de 01/01/2021 e 05/11/2024.

2.1. Séries Temporais

Uma Série Temporal pode ser definida como uma sequência de dados observados em intervalos de tempo regulares, ou seja, uma variável aleatória ordenada no tempo. Assim, temos Y_t em que $(Y_t)_a^z = \{Y_a, \dots, Y_z\}$, onde Y_t é uma variável aleatória observada uma única vez regular (BOX; JENKINS, 2015).

2.2. Modelo auto-regressivo integrado de média móvel

De acordo com Cruz (2010), o modelo ARIMA é uma generalização do modelo auto-regressivo de média móvel (ARMA). A representação ARIMA (p, d, q) refere-se, respectivamente, às ordens de auto regressão, de integração e de média móvel: p é o operador de auto-regressivo, d número de diferenças, q é o número de termos da média móvel. Logo após o modelo ajustado ARIMA, verifica-se que os resíduos do modelo são independentes, denominados de ruído branco.

Denomina-se de modelo Autorregressivo de ordem AR(p) o modelo dado pela equação 2.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t \quad (2)$$

em que os termos $Z_{t-1}, Z_{t-2} + \dots + Z_{t-p}$ são independentes de a_t , ϕ_1 é o parâmetro (peso) que descreve como Z_t se relaciona com o valor para Z_{t-1} (a série de dados) e a_t é o erro aleatório.

A metodologia implantada de Box e Jenkins e Reinsel (2015) para construção de modelos

ARIMA é composta por três passos: Identificação; Estimação e Verificação. Logo, o passo 1 é mais relevante pois nele se determina os valores de P, a e b. Esta determinação é feita mediante as autocorrelações e autocorrelações parciais estimadas, as quais se espera que represente adequadamente as verdadeiras quantidades teóricas que são desconhecidas.

2.3 Neural Network Autoregression (NNAR).

A rede Neural Auto regressivo (NNAR), foi introduzida em 2018 por Hyndman e Athanasopoulos é um tipo de modelo de previsão de séries temporais baseado em redes neurais artificiais (ANNs). Uma das principais vantagens dos modelos NNAR é a sua capacidade de capturar relações não lineares entre os dados, o que os torna mais flexíveis e poderosos do que modelos lineares como o ARIMA [Maleki et al. 2018]. Esse modelo foi criado para uma rede neural feedforward com uma camada oculta, denotada por NNAR(p,k), em que p indica o número de neurônios nas camadas ocultas e k indica a entrada defasada.

O modelo Neural network autoregression é definido com uma rede multiplicas camadas envolvendo uma combinação linear como função de ativação, os pontos de saída são entradas para a próxima. Ainda de acordo com Thoplan, (2014), a função linear é dada pela seguinte expressão:

$$z_j = b_j + \sum_{i=1} w_{i,j} x_i \quad (3)$$

As entradas no neurônio oculto j são combinações linearmente, em que os pesos w e os parâmetros b são definidos a partir dos dados. Salientando que nas camadas ocultas a função não linear é utilizada como uma sigmoid, a função de ativação, definida pela seguinte equação (4):

$$f(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

2.4. Acurácia

Para mensurar os desempenhos dos modelos foram utilizados: o MAE (Erro médio absoluto) é calculado a partir da média dos erros absolutos, utiliza-se o módulo de cara erro, evitando a subestimação, logo o valor é menos afetado pelos pontos extremos (outliers). Cada erro é interpretado como a diferença entre y e $\hat{y}$, assim temos:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

E o MAPE (O erro percentual absoluto médio) é a média dos erros absolutos durante um determinado período multiplicado por 100% para que os resultados obtidos sejam em forma de porcentagens (Junita, et al. 2024), definido como:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100$$

(6)

2.5. Resíduos dos modelos

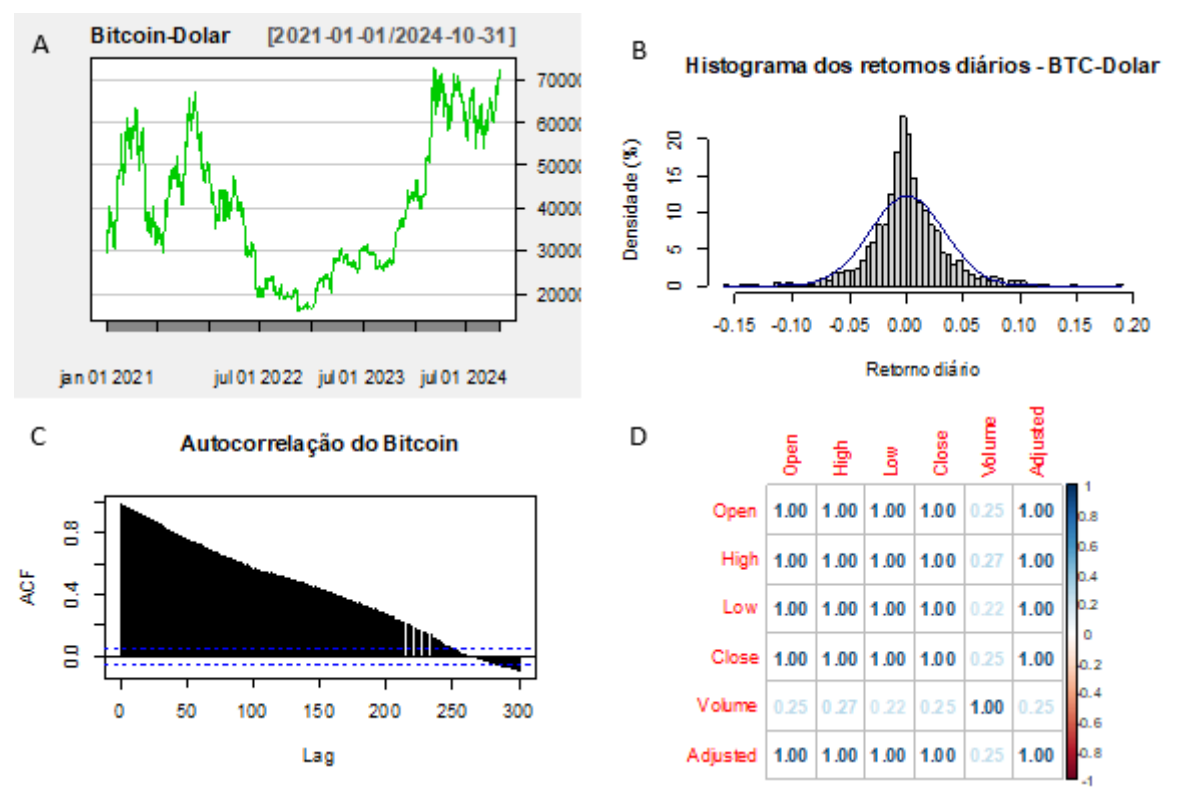
Para se validar o modelo ajustado os resíduos devem satisfazer as seguintes condições:

- 1- Aleatoriedade – Gráfico dos resíduos versus a ordem da coleta dos dados; e,
- 2- Função de autocorrelação. Segundo Morettin e Toloi (2018) para verificar se uma série é ruído branco, constrói-se o gráfico da função de autocorrelação com o seu respectivo intervalo de confiança. Caso as correlações residuais estejam dentro do intervalo, afirma-se que o processo é ruído branco.
- 3- Normalidade – Para a análise dos resíduos, deve-se observar a suposição de normalidade, sendo possível observá-la por meio do histograma e teste de normalidade.

3. Resultados e discussões

Esta seção apresenta os resultados obtidos, por meio de uma análise da série temporal, compreendida entre janeiro de 2020 a outubro de 2024 os gráficos apresentados abaixo, são possíveis constatar que:

Figura 1: Gráfico (A) Série temporal compreendida entre janeiro de 2020 a outubro de 2024; Gráfico (B) Retornos diários (oscilação) do Bitcoin; Gráfico (C) Autocorrelação dos preços; Gráfico (D) Matriz de correlação preços e volume.



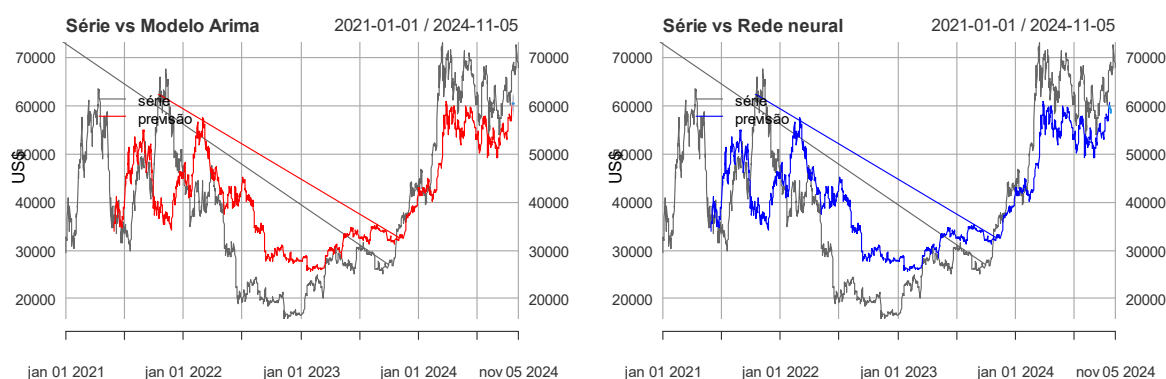
Observa-se no Gráfico (A) que a série referente aos valores de fechamento do Bitcoin entre julho de 2022 a março de 2023, o ativo sofreu uma depreciação no preço atingindo a menor marca no período de 15787.28 dólares, atualmente o ativo está em tendência de alta. Gráfico B, apresenta-se que as oscilações de preço (retornos) da criptomoeda Bitcoin na sua periodicidade diária, após serem quantificadas em bases logarítmicas, apresentam distribuição que se assemelha à distribuição normal. Para o gráfico de autocorrelação (C) Os valores próximos a 1 ou a -1 indicam uma forte influência do respectivo registro passado no registro atual, ao passo que um valor próximo a 0 indica uma fraca influência.

Para a série temporal analisada, isto é, no comportamento dos preços da criptomoeda Bitcoin, é possível verificar que os preços medidos diariamente, demonstram autocorrelação positiva aproximadamente o 270º dia antecedente. A partir disso, os preços indicam uma autocorrelação negativa. Com relação as correlações entre os preços (Gráfico D), apenas o volume de moedas negociadas em um determinado tempo não possui correlação significativa com a variação do preço do respectivo ativo.

3.1. Os modelos ajustados

Para fins de modelagem do comportamento do preço de fechamento diário do Bitcoin, foi utilizado dois modelos: Um modelo do tipo Arima (*AutoRegressive Integrated Moving Average model*) que a combinação entre métodos de diferenciação e os modelos de autorregressão e média móvel. O segundo modelo foi o *Neural Network Auto Regressive* (NNAR) é a Artificial Neural Network (ANN) que Hyndman e Athanasopoulos (2018), esse modelo tem à vantagem de ser capaz de prever com eficiência quando existe uma relação não lineares entre os meses e por isso não precisamos nos preocupar em fazer transformações, como a Box-Cox em tirar a diferença sazonal. Os modelos ajustados abaixo (Figura 2).

Figura 2: Série temporal dos preços de fechamento do ativo e os modelos ARIMA (1,1,0) e NNAR (2,2).



O modelo utilizado para a série é Arima (1,1,0), um modelo composto por um termo autorregressivo e apenas uma diferença necessária para tornar a serie estacionária, o segundo modelo NNAR (2,2) que considera 2 lag's na camada de entrada, 2 lag sazonal.

3.2. Métricas

A seguir é apresentado na Tabela 1, as métricas para a validação dos modelos ajustados: erro absoluto médio (MAE) e erro percentual absoluto médio (MAPE) para os modelos:

Tabela 1: Valores obtidos durante a validação dos modelos de previsão.

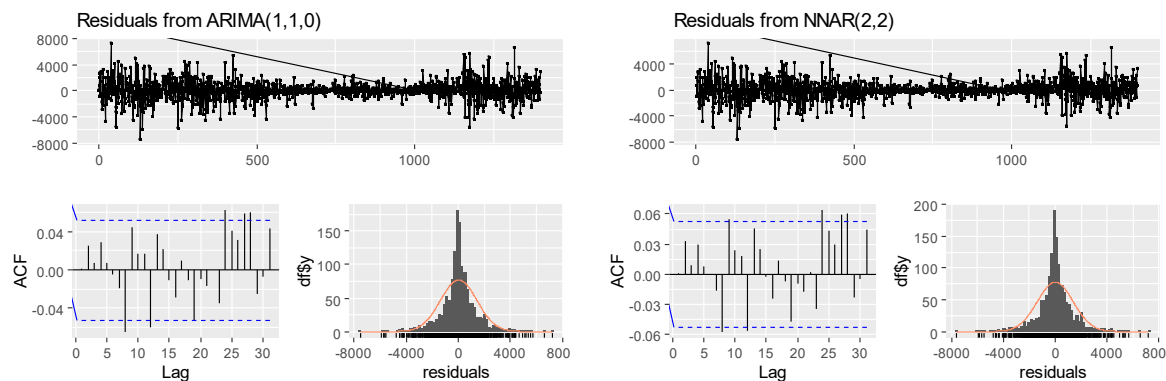
Modelos	Variância	MAE	MAPE
<i>Arima (1,1,0)</i>	1947146	916.6333	0.9921052
<i>NNAR(2,2)</i>	1927664	919.3747	0.9950723

Os modelos apresentam variâncias semelhantes, enquanto o MAE o modelo Arima (1,1,0) apresenta melhor desempenho, ou seja, obtém uma menor média entre os erros absolutos. Com relação ao MAPE, o segundo o modelo de previsão tem melhor desempenho quando o valor assumido é baixo, para isso há quatro níveis de classificação: altamente preciso (<10%); preciso (10% - 20%); regular (20% -50%) e impreciso (>50%), respectivamente (Junita, T. P., & Kartikasari, M. D. ,2024). Logo ambos os modelos se apresentam altamente precisos, com MAPE em torno 0,99%, possuindo pequena discrepância entre os valores observados e previstos.

3.3. Análise de resíduos

A próxima etapa foi verificar a qualidade do ajuste dos modelos, para isso vamos analisar os resíduos dos modelos, olhando para os erros gerados no processo de estimação, apresentado abaixo na Figura 3.

Figura 3: Analise de resíduos para os modelos ajustados a série dos ativos.



Observa-se que os resíduos obtidos em ambos os modelos se caracterizam como um ruído branco, as sequências de erros (resíduo) oscila em torno de zero, ou seja, à média dos resíduos é zero sem viés, a variância dos resíduos produzidos pelos modelos é contante, a dispersão dos resíduos não se altera ao longo do tempo, o que significa que os resíduos são independentes e identicamente distribuídos, assumindo uma covariância igual a zero.

A funções de autocorrelação para ambos os modelos mostra que as autocorrelações para os erros de previsão praticamente não excedem os limites de significância para defasagens

1-30, indicando que não há muita evidência de autocorrelações diferentes de zero nas defasagens 1-30. Portanto, podemos concluir que os resíduos não são auto correlacionados.

Além disso, os erros de previsão têm variância constante ao longo do tempo e são normalmente distribuídos com média zero, visto nos histogramas de Resíduos para ambos os modelos. Logo os modelos ARIMA (1,1,0) e NNAR (2,2) são válidos.

4. Conclusão

Com base nos resultados obtidos mostram que os modelos ARIMA e NNAR, mesmo apresentando índices satisfatórios para validação. A rede Neural NNAR (2,2), mesmo apresentando estruturas mais complexas, não conseguiu ter precisão maior que o modelo ARIMA (1,1,0), logo para estimar recomenda-se o uso deste modelo mais simples, dado que consegue representar a dinâmica do sistema com mesma precisão.

Observou-se, analisando a série e a literatura que é necessário obter modelos que capturem a sazonalidade, pois, o comportamento de preços com relação a esse ativo apresenta períodos ciclos de alta e baixa, influenciados por fatores externos e internos, conhecidos no mercado financeiro como “inverno e verão cripto”. Logo para trabalhos futuros recomenda-se modelos de métodos multiplicativo, dado que existe variações sazonais que mudam proporcionalmente ao nível da série.

Acknowledgements

Nossos agradecimentos vão primeiramente ao evento “Wanda 2024” e seus organizadores e a Universidade Estadual da Paraíba, pela disponibilidade concedida aos alunos.

Conflict of Interest Declaration

Os autores não têm conflitos de interesse a declarar. Todos os coautores concordam com o conteúdo do manuscrito e não há interesse financeiro em relatar.

References

1. Azari, A. (2019). Bitcoin price prediction: An ARIMA approach. *arXiv preprint arXiv:1904.05315*.
2. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
3. Cruz, D. V., Cunha Filho, M., & Falcao, A. P. S. T. (2014). Controle Estatístico de processos aplicado ao monitoramento do pH e turbidez das águas no abastecimento de Campina Grande-PB. *Rev. Bras. Biom*, 32(4), 459-477.
4. Duarte, F. L. C., & de Almeida Lima, N. C. (2019). Redes Neurais Aplicadas Na Predição do Preço do Ether: MLP vs bi-LSTM. *Anais do Encontro de Computação do Oeste Potiguar ECOP/UFERSA (ISSN 2526-7574)*, 1(3).

5. Finance, Y. (2020). Yahoo Finance. Retrieved from finance.yahoo.com: <https://finance.yahoo.com/recent-quotes> , (acesso em 02 de novembro de 2024).
6. Junita, T. P., & Kartikasari, M. D. (2024). Application of the neural network autoregressive (nnar) method for forecasting the value of oil and gas exports in Indonesia. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 18(1), 0341-0348.
7. Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one*, 11(8), e0161197.
8. Hyndman, R. J., & Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd edn. OTexts, Melbourne (2018).
9. Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).
10. Maleki, A., Nasser, S., Aminabad, M. S., & Hadi, M. (2018). Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics. *KSCE Journal of Civil Engineering*, 22, 3233-3245.
11. McNally, S. (2016). *Predicting the price of Bitcoin using Machine Learning* (Doctoral dissertation, Dublin, National College of Ireland).
12. Morettin, P. A., & Toloi, C. M. (2018). *Análise de séries temporais: modelos lineares univariados*. Editora Blucher.
13. Nakamoto, S., & Bitcoin, A. (2008). A peer-to-peer electronic cash system. *Bitcoin*.—URL: <https://bitcoin.org/bitcoin.pdf>, 4(2), 15.
14. NUBANK. Menu. Disponível em: <<https://nubank.com.br/>> Acesso em 15 nov. 2024.
15. Radityo, A., Munajat, Q., & Budi, I. (2017, October). Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods. In *2017 international conference on advanced computer science and information systems (ICACSYS)* (pp. 433-438). IEEE.
16. Shintate, T., & Pichl, L. (2019). Trend prediction classification for high frequency bitcoin time series with deep learning. *Journal of Risk and Financial Management*, 12(1), 17.
17. Thoplan, R. (2014). Simple v/s sophisticated methods of forecasting for mauritius monthly tourist arrival data. *International Journal of Statistics and Applications*, 4(5), 217-223.
18. Ulrich, F. (1892). Bitcoin-a moeda na era digital. *Journal*, volume, 2, 239.
19. Velankar, S., Valecha, S., & Maji, S. (2018). Bitcoin price prediction using machine learning. In *2018 20th International Conference on Advanced Communication Technology (ICACT)* (pp. 144-147). IEEE.