



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA,
CONTABILIDADE E SECRETARIADO EXECUTIVO – FEAACS
BACHARELADO EM FINANÇAS

PABLO ANTHONY COSTA CASTRO

ANÁLISE DE SENTIMENTO DOS TWEETS E SUA INFLUÊNCIA SOBRE O
PREÇO DO BITCOIN

FORTALEZA

2023

PABLO ANTHONY COSTA CASTRO

ANÁLISE DE SENTIMENTO DOS TWEETS E SUA INFLUÊNCIA SOBRE O PREÇO DO
BITCOIN

Monografia apresentada ao Curso de Finanças da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharelado em Finanças.

Orientador: Prof. Dr. Francisco Gildemir Ferreira da Silva.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

C353a Castro, Pablo.
ANÁLISE DE SENTIMENTO DOS TWEETS E SUA INFLUÊNCIA SOBRE O PREÇO DO BITCOIN
/ Pablo Castro. – 2023.
36 f. : il.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Faculdade de Economia,
Administração, Atuária e Contabilidade, Curso de Finanças, Fortaleza, 2023.
Orientação: Prof. Dr. Francisco Gildemir Ferreira da Silva.

1. Bitcoin. 2. análise de sentimento. 3. Twitter. 4. mercado financeiro. 5. criptomoedas. I. Título.
CDD 332

PABLO ANTHONY COSTA CASTRO

**ANÁLISE DE SENTIMENTO DOS TWEETS E SUA INFLUÊNCIA SOBRE O
PREÇO DO BITCOIN**

Trabalho de Conclusão do curso de
Finanças, da Universidade Federal do
Ceará, como requisito para obtenção do
título de Bacharel em Finanças.

Aprovada em: __/__/____.

BANCA EXAMINADORA

Prof. Dr. Francisco Gildemir Ferreira da Silva (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Prof. Dr. Vitor Borges Monteiro
Universidade Federal do Ceará (UFC)

Prof. Prof. Dr. Victor Aguiar Evangelista de Farias
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

À minha família, pelo apoio e respeito em todas as decisões que tomei na minha vida.

Ao Prof. Dr. Francisco Gildemir Ferreira da Silva, pela excelente orientação, paciência e apoio durante toda esta jornada acadêmica.

Aos meus amigos que fiz durante toda essa jornada universitária e na vida, tudo foi mais fácil com vocês ao meu lado, obrigado por tudo.

Resumo

As redes sociais têm o poder influenciar os sentimentos e as decisões dos indivíduos, e no mercado financeiro a situação não é diferente. Nos últimos anos, a criptomoeda Bitcoin atravessou por grandes variações no seu preço, decorrente de diversas razões. Com base nessas observações, foi possível conduzir um estudo de como os sentimentos dos investidores, através de suas publicações na rede social Twitter, influenciam o preço de fechamento da criptomoeda Bitcoin. Primeiramente, uma base de dados, contendo várias variáveis e com um número expressivo de tweets, foi encontrada no site Kraggle, que abrange o período entre 10/02/2021 e 06/01/2023. A partir dessa base de dados, foi empregada a biblioteca NLTK para a etapa de pré-processamento dos dados, a qual inclui análise semântica e estatística. A biblioteca VADER foi utilizada para mensurar a polaridade dos sentimentos expressos nas publicações. Adicionalmente, foi utilizada uma base de dados contendo as informações sobre o Bitcoin durante o mesmo período. Com a avaliação dos sentimentos dos tweets e os dados relativos ao preço de fechamento do Bitcoin, foi possível criar um modelo de regressão ARDL (Autoregressive Distributed Lag) para avaliar se as variáveis selecionadas influenciam no preço do Bitcoin. Os resultados revelaram que as variáveis selecionadas possuem influência sobre o preço do Bitcoin. Consequentemente, a hipótese de que os tweets influenciam o preço do Bitcoin foi não rejeitada.

Palavras-chaves: Bitcoin; análise de sentimento; Twitter; mercado financeiro; criptomoedas

Abstract

Social media has the power to influence the feelings and decisions of individuals, and in the financial market, the situation is no different. In recent years, the cryptocurrency Bitcoin has experienced significant price fluctuations due to various reasons. Based on these observations, a study was conducted on how investors' sentiments, as expressed in their Twitter posts, influence the closing price of the Bitcoin cryptocurrency. Firstly, a database containing multiple variables and a substantial number of tweets was found on the Kraggle website, covering the period from 10/02/2021 to 06/01/2023. From this database, the NLTK library was used for data preprocessing, which includes both semantic and statistical analysis. The VADER library was employed to measure the polarity of sentiments expressed in the posts. Additionally, a database containing information about Bitcoin during the same period was utilized. By evaluating the sentiments of tweets and the data related to the closing price of Bitcoin, an Autoregressive Distributed Lag (ARDL) regression model was created to assess whether the selected variables influence the price of Bitcoin. The results revealed that the selected variables do have an influence on the price of Bitcoin. Consequently, the hypothesis that tweets influence the price of Bitcoin was not rejected.

Key-words: Bitcoin; sentiment analysis; Twitter; Bitcoin; financial market; cryptocurrencies.

LISTA DE FIGURAS

Figura 1 – Fluxograma do funcionamento da análise de sentimento	17
Figura 2 – Passo-a-passo de uma etapa do pré-processamento dos tweets.....	18
Figura 3 – Número de tweets distribuídos entre os sentimentos positivo, neutro e negativo.....	21

LISTA DE GRÁFICOS

Gráfico 1 – Média do grau de sentimento em 1 dia com preço de fechamento do Bitcoin	22
Gráfico 2 – Média móvel de 5 dias do grau de polaridade dos tweets com o preço de fechamento do Bitcoin	22
Gráfico 3 – Variável dependente (Close) x independente (compound)	25
Gráfico 4 – Variável dependente (Close) x independente (Volume)	26
Gráfico 5 – Correlação entre as variáveis	27
Gráfico 6 – Gráfico de homoscedasticidade dos resíduos.....	30
Gráfico 7 – Quadro com a distribuição dos resíduos na série temporal, histograma dos resíduos, a autocorrelação e Q-Q plot.....	31
Gráfico 8 – Teste de autocorrelação parcial	32

LISTA DE TABELAS

Tabela 1	– Base de dados após o pré-processamento e com a polaridade definida	19
Tabela 2	– Sentimentos definidos em cada tweet	20
Tabela 3	– Amostra da base de dados com as variáveis da análise	21
Tabela 4	– Estatística descritiva dos dados	25
Tabela 5	– Resultados estatísticos do modelo de regressão linear múltipla	28
Tabela 6	– Resultados dos coeficientes	29
Tabela 7	– Modelo ARDL	32

LISTA DE ABREVIATURAS E SIGLAS

HME	Hipótese do Mercado Eficiente
RWT	Random Walk Theory
NLTK	Natural Language Toolkit
VADER	Valence Aware Dictionary and Sentiment Reasoner
ARDL	Autoregressive Distributed Lag

Sumário

1. INTRODUÇÃO	11
2. REVISÃO DE LITERATURA	12
2.1. HIPÓTESE DOS MERCADOS EFICIENTES	12
2.2. A INFLUÊNCIA DAS REDES SOCIAIS NO SENTIMENTO DOS INVESTIDORES	13
2.3. O MERCADO BITCOIN	15
3. METODOLOGIA	16
3.1. PRÉ-PROCESSAMENTO DA BASE DE DADOS	17
3.2. ANÁLISE DE SENTIMENTOS	19
3.3. MODELO ARDL PARA ANÁLISE	23
4. RESULTADOS E DISCUSSÃO	24
4.1. ANÁLISE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA	27
4.2. ANÁLISE DOS RESÍDUOS	29
5. CONSIDERAÇÕES FINAIS, LIMITAÇÕES E RECOMENDAÇÕES	33
6. REFERÊNCIAS.....	34

1. INTRODUÇÃO

No mercado financeiro há várias teorias e hipóteses em que o investidor pode se guiar para melhor orientar os seus investimentos. Uma hipótese que deu base para um amplas discussões sobre o preço das ações foi a Hipótese dos Mercados Eficientes (HME), de Eugene Fama (1970). A hipótese sugere que os preços do mercado sempre refletem todas as informações existentes, e que não existe ação cara ou barata, pois o mercado é eficiente em precificar o seu valor justo e que os agentes são racionais, no longo prazo. A partir desta hipótese, pode ser observado como novas informações influenciam o sentimento do investidor, até então considerados racionais. Com o passar do tempo, outros estudos questionaram a hipótese de racionalidade dos investidores.

O mercado financeiro tem se tornando cada vez mais tecnológico e informatizado, onde qualquer informação ou dado pode influenciar os preços, pois os agentes podem tomar ações de curto prazo ou aguardar mais dados para tomar decisões. No século XX tivemos ideias como o *Random Walk Theory*, de Karl Pearson (1905), e as Hipóteses do Mercado Eficiente (HME), de Eugene Fama (1970), em que essas teorias são contraditórias, segundo o famoso gestor Peter Lynch e também do investidor bilionário Warren Buffett. A teoria do passeio aleatório possui a premissa de que não importa o número de informações, tudo é aleatório. Enquanto nas HME, especialmente a terceira hipótese, sustenta que todas as informações, sejam públicas ou privadas, são acessíveis para todos, então torna-se difícil obter lucros a longo prazo. Como observado, os conceitos destes autores guiaram o mercado de capitais por décadas.

A partir dessa observação e de casos como por exemplo da varejista norte-americana GameStop se valorizaram mais de 5000% de forma artificial, em poucos dias, dado um grupo de investidores individuais em um fórum no site Reddit¹, e como consequência fundos perderam bilhões. O caso exemplifica que notícias e opiniões influenciam no nosso cotidiano, geralmente trazendo algum viés para o respectivo leitor sobre algo, influenciando-o e a sua racionalidade, principalmente com investidores iniciantes. As moedas digitais, especialmente o Bitcoin, tem muita volatilidade nos seus preços, por vezes afetadas por indícios de regulação dos países,

¹https://pt.wikipedia.org/wiki/Caso_GameStop

oferta de fundos de criptomoedas por grandes corporações do sistema financeiro e também tweets de influenciadores, além de outros.

O objetivo deste trabalho é analisar se os tweets possuem influência no preço da criptomoeda Bitcoin, analisando se os sentimentos dos investidores refletem nos preços de fechamento da moeda, de acordo a segunda Hipótese do Mercado Eficiente de Eugene Fama.

2. REVISÃO DE LITERATURA

2.1. HIPÓTESE DOS MERCADOS EFICIENTES

Eugene Fama, famoso economista norte americano, por desenvolver a Hipótese do Mercado Eficiente (HME) em 1970, contribuiu grandemente para o mercado financeiro, na época, mostrando três hipóteses de como o mercado funcionaria dado o nível de acesso as informações. As três hipóteses foram definidas como:

- Eficiência fraca: O mercado é eficiente em refletir todas as informações públicas disponíveis, logo, os retornos nos mercados são independentes, isso quer dizer que, retornos passados não ajudam a prever retornos futuros;
- Eficiência semiforte: Abrange a primeira hipótese, e sugere que as informações são absorvidas pelo mercado de forma instantânea. Com isto, os investidores não conseguem ganhos acima do mercado com as informações disponíveis;
- Eficiência forte: Engloba as outras duas hipóteses e afirma que todo tipo de informação reflete imediatamente nos preços dos ativos, seja ela pública ou privada. Desta forma, nenhum investidor conseguiria ganhos acima do mercado mesmo com novas informações.

Fama (1970) também aponta que a presença de correlação serial no preço das ações, no curto prazo, foi considerada restrita para se transformar em lucros, afirmando que profissionais não poderiam tirar lucros com essas pequenas tendências. O problema dessas hipóteses, era que assumiam os investidores como racionais. Com o avanço dos estudos, da tecnologia e da arbitragem, anomalias no mercado começaram a se intensificar e a racionalidade do investidor começou a ser questionada. Fama (1991) afirmou que a previsibilidade do horizonte do longo

prazo é os resultados das bolhas irracionais nos preços ou grandes oscilações nos retornos esperados.

Grandes economistas como Robert Shiller (1981), questionou a HME e mostrou que a volatilidade explica qualquer ponto de vista racional para o futuro, pois os investidores esperam um retorno naquilo que investem. Segundo Schumaker *et al.* (2012 *apud* LeBaron, Arthur *et al.*, 1999) observaram um aparente atraso no comportamento dos investidores ao absorverem novas informações, o que ajudou a descartar as teorias de correção instantânea com a nova entrada de informações, e que deu a ideia de mercados podem ser previstos num curto espaço de tempo com as novas informações.

Com o passar do tempo surgiram outras ideias como Lima (2003) afirma, de que os teóricos, na sua formulação mais moderna, sustentam que a tendência de preços é identificada como um passeio aleatório (Teoria do Passeio Aleatório), significando que se a informação já está capitalizada no preço de uma ação, conhecer a informação não o tornará um bom investidor.

O trabalho de Sousa (2019) estudou o sentimento dos investidores durante o período de eleições de 2018, os resultados encontrados corroboram com a teoria dos mercados eficientes de Fama (1970, 1990), em sua forma semiforte, de que os retornos passados mais as informações publicamente disponíveis, aqui representadas pelos tweets e os sentimentos que deles foram extraídos, possuem efeitos sobre os retornos contemporâneos das ações.

Nas últimas pesquisas acadêmicas, autores como Sousa (2020), utilizaram a segunda hipótese, das HME, em que afirma que os preços refletem todas as informações disponíveis no momento e que os preços mudam de acordo com que surgem novas informações. Também foi observado uma grande influência das redes sociais para confrontar essas hipóteses.

2.2. A INFLUÊNCIA DAS REDES SOCIAIS NO SENTIMENTO DOS INVESTIDORES

Os autores Summers e Schleiffer (1990, p. 19-20) trazem duas hipóteses interessantes sobre os sentimentos dos investidores:

“Nossa abordagem se baseia em duas hipóteses. Em primeiro lugar, certos investidores não são totalmente racionais e sua demanda por ativos de risco é afetada por suas crenças ou sentimentos que não são inteiramente justificados por informações a respeito dos fundamentos. Em segundo lugar, a arbitragem definida como as decisões tomadas por investidores inteiramente racionais, não sujeitos a tais sentimentos, envolve risco e é, portanto, limitada. As duas hipóteses, em conjunto, implicam que mudanças nos sentimentos do investidor não são integralmente contrabalançadas pelos arbitadores e assim afetam os retornos dos títulos.”

De acordo com Brav e Heaton (2002, apud Mussa et al., 2008), para as finanças comportamentais, os mercados podem ser ineficientes, porque investidores irracionais podem levar os preços além dos níveis racionais por algum tempo, e por esses preços irracionais que se pratica a arbitragem, ou a segunda hipótese das HME, onde investidores, mais preparados geralmente, procuram anomalias nesses comportamentos.

Segundo o site Traders Club (2021), e Brown e Cliff (2014), o sentimento do investidor nada mais é que um viés de especulador, com excesso de otimismo ou pessimismo. Niederhoffer (1971) analisou o como mercado reage a notícias de diferentes categorias e observou que o mercado tem tendencia de ter reações exageradas a notícias negativas.

Com este desejo de expressar o que estava sentindo, de acordo com as notícias e movimentações do mercado financeiro, em tempo real, o Twitter tornou-se uma rede social popular entre influenciadores, gestores, analistas e investidores individuais. Bollen, Mao e Zeng (2011) acreditam ser razoável supor que o humor e o sentimento do público podem conduzir os valores do mercado de ações tanto quanto as notícias.

De acordo com Medeiros e Borges (2019) alguns investidores expressam seus sentimentos sobre mercado financeiro nas redes sociais, principalmente o Twitter, e podem tomar decisões baseadas no que é escrito em seus perfis. O Twitter é atualmente a rede social mais popular da internet, pois permite que os usuários digitem o que estão fazendo ou pensando em tempo real (Silva, 2016). Brown e Cliff (2014) também observaram que nos últimos anos, os pesquisadores estão voltando a sua atenção para captar o sentimento dos investidores de forma instantânea e on-line.

Segundo Elbagir e Yang (2019), o Twitter promove uma rica fonte de dados que são utilizados nos campos de mineração de opinião e análise de sentimentos, este último sendo o

foco do trabalho. Recentemente, a análise de sentimentos através das publicações dos usuários desta rede social tem atraído a atenção de pesquisadores e analistas do mercado financeiro.

Pesquisas como a de Nunes (2019) afirma que os tweets de Elon Musk não têm impacto significativo nas rentabilidades do Bitcoin. Também Nunes (2019, apud Bollen et al., 2011), descobre que as mudanças no estado de humor do público podem ser verificadas à partir de publicações no Twitter, por meio de técnicas de processamento de texto. Ainda assim, apenas alguns estados de humor conseguem prever o índice Dow Jones (Nunes (2019, apud Bollen et al., 2011)).

Silva (2022) menciona que com o desenvolvimento das redes sociais e a popularidade do Bitcoin, uma análise de sentimento é extremamente relevante para uma análise mais completa do seu retorno.

2.3 O MERCADO BITCOIN

O Bitcoin, originalmente criado para facilitar trocas comerciais, porém a sua objetividade foi mudando e atualmente é um dos principais veículos de investimentos, sendo amplamente conhecido no mercado de criptomoedas, fatiando quase 80% do total de capitalização das criptomoedas comercializadas, em junho de 2016, segundo Phillip et al. (2018). De acordo com Silva (2022, apud Sensoy, 2019; Sovbetov, 2018), a volatilidade tem um efeito negativo sobre a eficiência dos preços do Bitcoin, mostrando reações mais severas no curto prazo.

Silva (2022, apud Kristoufek, 2013), em sua visão sobre a formação de preços do Bitcoin, não pode ser explicado por padrões macroeconômicos, como a paridade da taxa de juros descoberta, pois o Bitcoin não é circulado por nenhum Banco Central específico, portanto, está desconectado da economia mundial. Silva (2019) também cita que durante o período de 2009-2015, a oferta e demanda de Bitcoin têm um importante impacto em seu preço e sua importância tende a aumentar com o tempo, e, não rejeita a hipótese de que o comportamento especulativo dos investidores afeta o preço do Bitcoin a curto e longo prazo.

Segundo Kyriazis, Papadamou & Corbet (2020), os preços no mercado de criptomoedas podem variar substancialmente, dado que os investidores tendem a ter reações exageradas a notícias ruins, desta forma os resultados dos modelos podem tornar-se questionáveis.

3. METODOLOGIA

Neste capítulo, serão abordadas as seguintes etapas: definição das variáveis das bases de dados, descrição da amostra, tratamento dos dados, apresentação dos dados, modelo econométrico e análise de como o sentimento do investidor sobre o Bitcoin, através dos tweets influencia o preço de fechamento da cripto moeda.

Os próximos parágrafos detalham o processo metodológico adotado neste estudo. É importante ressaltar que, de acordo com a literatura apresentada por Nunes (2019): “*os tweets publicados na rede social Twitter, sejam eles de pessoas conhecidas ou não, exercem impacto tanto na decisão financeira, como nos preços e volumes de negociação dos mercados financeiros*”. Essa compreensão da influência dos tweets fornece uma base sólida para a metodologia empregada neste trabalho.

A base de dados foi obtida por meio do site Kaggle², que é uma plataforma de competição de ciência de dados. Nesta plataforma, usuários compartilham projetos contendo bases de dados prontas, onde a partir delas é lançados desafios, a qual está sujeito a notas, prêmios ou somente diversão. A nossa base de dados é composta por aproximadamente 4,6 milhões de tweets escritos sobre a criptomoeda Bitcoin, com a grande maioria escrita na língua inglesa. A base foi escolhida por possuir uma boa diversidade de variáveis e que mesmo pegando reduzindo o tamanho da base, a amostra ainda possui um número significativo de variáveis. A base original possui 13 colunas de variáveis, porém em decorrência da dimensão da base de dados e da grande demanda de processamento do computador, foram escolhidas 3 variáveis: localização de onde os tweets foram originados (*user_location*), data de publicação (*date*) e o conteúdo dos *tweets (text)*.

Além desta base de dados, é relevante mencionar que o *Yahoo Finance* é um website voltado para o mercado financeiro, onde este website desempenhou um papel fundamental para

² <https://www.kaggle.com/datasets/kaushikuresh147/bitcoin-tweets>

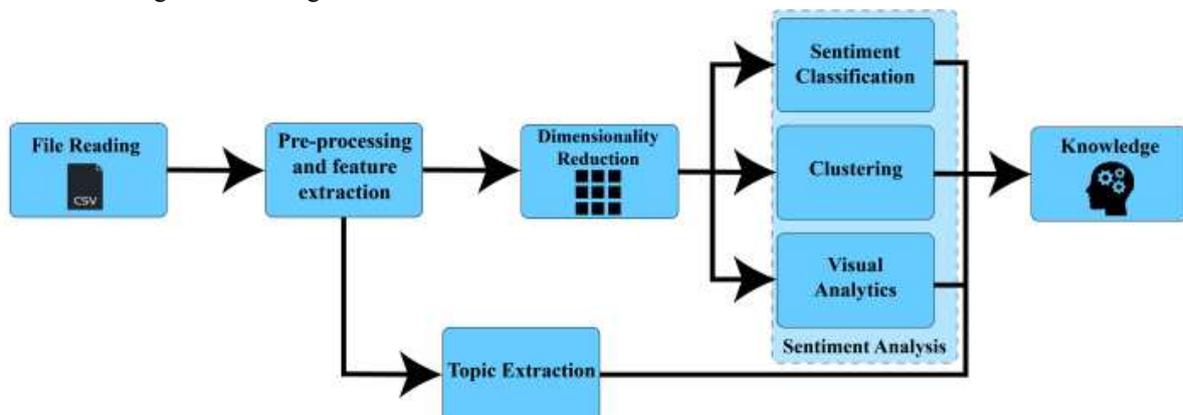
a pesquisa. Este portal oferece notícias sobre o mercado, acompanhamento de ações e índices, e principalmente dados financeiros, por estes motivos, a base de dados sobre o Bitcoin foi obtida através deste. As variáveis extraídas desta base de dados incluem o preço de abertura, o preço de fechamento, maior e menor preço negociado durante o dia e volume de transações. No âmbito da nossa pesquisa, o foco é o preço de fechamento da criptomoeda.

A serie temporal das duas bases de dados envolveu o período entre 10/02/2021 e 06/01/2023. O estudo investiga se as variáveis possuem influência no preço do Bitcoin. Portanto, será estimado se as variáveis escolhidas possuem influência na formação do preço do Bitcoin.

3.1. PRÉ-PROCESSAMENTO DA BASE DE DADOS

Para analisar estes sentimentos de uma forma mais automatizada e que seja possível a leitura de um grande volume de dados, provindos das redes sociais, a biblioteca *Natural Language Toolkit* (NLTK) é utilizada em larga escala pelos referenciados neste trabalho. Para Ebagir e Yang (2019), a NLTK é adequado para quem deseja trabalhar com dados textuais em um processo de linguagem natural e análise textual. Esta biblioteca fornece um jeito fácil de usar a interface e seus recursos lexicais. Incluindo um grupo de bibliotecas de pré-processamento para classificação, *tokenização*, análise e raciocínio semântico.

Figura 1 - Fluxograma do funcionamento da análise de sentimento

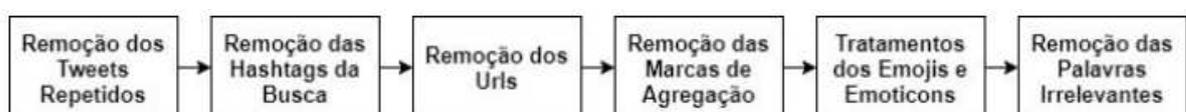


Fonte: Elbagir; Yang, 2019.

Na Figura 1 (Elbagir; Yang, 2019) é possível observar o fluxograma de como é feita a pesquisa. Após o download é feito o pré-processamento dos dados, através das bibliotecas *re* (tratamento textual), dentro da biblioteca NLTK, as técnicas de pré-processamento de texto escolhidas foram:

- Remoção de tweets duplicados: No *twitter* existe uma ferramenta chamada *Retweet*, em que é possível repostar um tweet de outro usuário, então é importante remover esses tweets repetidos. Foram removidos aproximadamente 47 mil tweets repetidos.
- Remoção de *Stopwords*: *Stopwords* (palavras vazias) são palavras que podem ser ocultadas numa frase e não afetam o sentido dela, no campo computacional, mas que ocupam muito espaço em memória e aumento do tempo de processamento. Exemplo destas palavras são: “as”, “de” e “para”. As bibliotecas mais eficientes para esta remoção estão em inglês
- Remoção de URL e outros: As URLs, assim como hashtags, emojis, pontuação, acentos e links, não tem papel importante na identificação da polaridade dos tweets, portanto, se tornam desnecessárias para o processamento.
- Remoção de espaços extras: Os usuários que utilizam o Twitter possuem uma linguagem mais informal, sendo comum encontrar tweets com múltiplos espaçamentos e isto também é tratado.
- Stemização: É o processo de reduzir a sua palavra ao radical, cada palavra é analisada individualmente e este processo foca no seu significado. A *bag-of-words* representa as palavras por um vetor, sinalizando quais delas estão naquele texto ou não. Com a Stemização é possível ficar com um vetor bem menor por conter menos palavras desnecessárias e repetidas, ajudando no desempenho operacional. A redução ao seu radical pode levar a uma outra gramaticalmente incorreta, porém ainda com valor para a análise.

Figura 2 – Passo-a-passo de uma etapa do pré-processamento dos tweets



Fonte: Kansaon; Brandão; Pinto, 2018.

3.2. ANÁLISE DE SENTIMENTOS

Após o pré-processamento, a base de dados está pronta para ser usada na análise de sentimentos. Para a análise de sentimentos, existe dois classificadores, pertencentes a biblioteca NLTK, TextBlob e VADER. Para a pesquisa é utilizado o método VADER, para determinar a polaridade dos sentimentos (Tabela 1) e classificação destes sentimentos (Tabela 2). Segundo Gilbert e Hutto (2014), afirmam que é uma ferramenta de análise de sentimento muito mais rápido do que algoritmos de aprendizado normalmente são. Na biblioteca VADER, cada sentimento ter uma proporção e classificação final para os sentimentos positivo, negativo ou neutro, diferente do TextBlob em que os resultados são a polaridade e a subjetividade. Com a escolha do classificador, o processo de polaridade do sentimento é criado, assim adicionando uma nova coluna na base de dados chamada de compound. Gilbert e Hutto (2014) define o compound como sendo uma métrica que calcula a intensidade da palavra, transformando-a em uma pontuação composta normalizada e ponderada.

Tabela 1 - Base de dados após o pré-processamento e com a polaridade definida

Date	Text	compound
2021-02-10	blue ridg bank share halt nyse bitcoin atm announc http co xaazmajkiv myblueridgebank http co sgbxmkp si	0.2960
2021-02-10	today thursday take friend leowandersleb btc wallet secur exp http co go adgrml	0.4939
2021-02-10	guy even read articl btc would like share http co qxczgmuy b http co wn ppkvi	0.5719
...
2023-01-06	Ethereum price updat eth usd bitcoin btc follo...	0.0000
2023-01-06	detail chang h low high bitcoin btc	-0.2732
2023-01-06	Earn crypto play fun game onlin get reward wat...	0.8555

Fonte: Elaboração própria (2023)

Com o grau de polaridade (*compound*) realizado, pôde-se classificar os sentimentos entre positivos, negativos e neutros. Os valores que Hutto e Gilbert (2014) propõem são:

- Sentimento positivo: *compound score* ≥ 0.05 ;
- Sentimento neutro: (*compound score* > -0.05) e (*compound score* < 0.05);
- Sentimento negativo: *compound score* ≤ -0.05 .

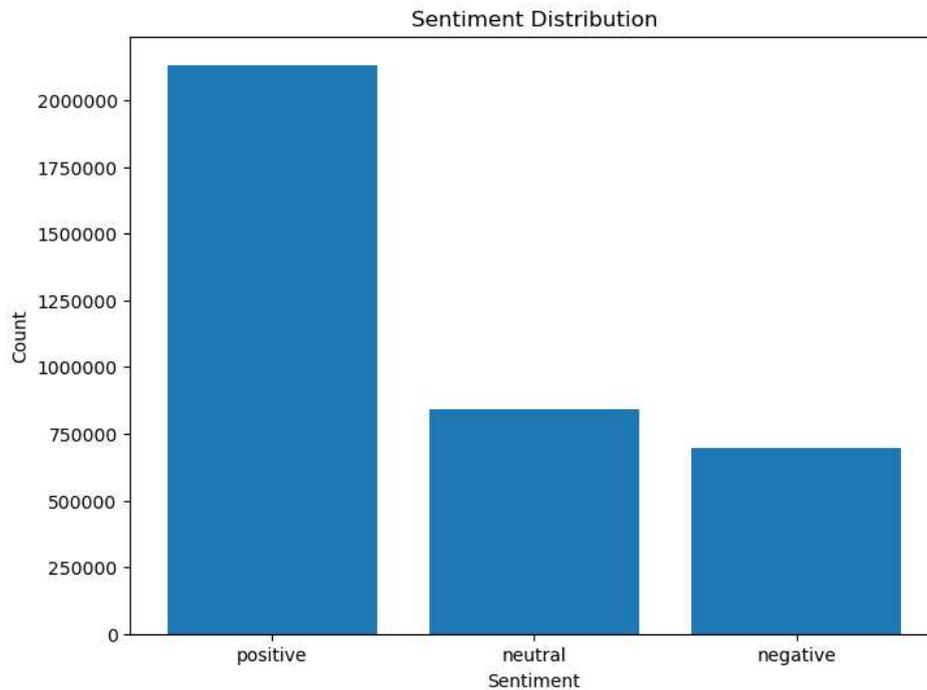
Tabela 2 – Sentimentos definidos em cada tweet

Date	Text	compound	sentiment
2021-02-10	blue ridg bank share halt nyse bitcoin atm announc http co xaazmajkiv myblueridgebank http co sgbxmkp si	0.2960	positive
2021-02-10	today thursday take friend leowandersleb btc wallet secur exp http co go adgrml	0.4939	positive
2021-02-10	guy even read articl btc would like share http co qxczgmuy b http co wn ppkvi	0.5719	positive
...
2023-01-06	Ethereum price updat eth usd bitcoin btc follo...	0.0000	neutral
2023-01-06	detail chang h low high bitcoin btc	-0.2732	negative
2023-01-06	Earn crypto play fun game onlin get reward wat...	0.8555	posivite

Fonte: Elaboração própria (2023)

Ao considerar toda a janela de extração dos dados (10-02-2021 a 06-01-23), a distribuição do grau de polaridade o seguinte efeito: o sentimento positivo teve como resultado 2.132.180 *tweets*; o sentimento neutro obteve 842.708 *tweets* e o sentimento negativo obteve 697.523 *tweets* no período. No final do pré-processamento foi obtido 3.672.411 *tweets* para análise, como observado na Figura 3.

Figura 3 - Número de *tweets* distribuídos entre os sentimentos positivo, neutro e negativo.



Fonte: Elaboração própria (2023)

Após a classificação dos tweets, foi realizada a média da variável polaridade, por dia, e a criação de uma nova base de dados mais enxuta (Tabela 3). Também é possível observar no Gráfico 1, a média dos sentimentos na série temporal com o preço de fechamento do Bitcoin.

Tabela 3 – Amostra da base de dados com as variáveis da análise

Date	compound	Close	Volume
2021-02-10	0.236377	44918.183594	87301089896
2021-02-13	0.231739	47105.515625	70250456155
2021-02-14	0.270487	48717.289062	71248675228
...
2022-12-26	0.160055	16919.804688	11886957804
2022-12-27	0.167015	16717.173828	15748580239
2023-01-06	0.198358	16951.968750	14413662913

Fonte: Elaboração própria (2023)

Gráfico 1 - Média do grau de sentimento em 1 dia com preço de fechamento do Bitcoin



Fonte: Elaboração própria (2023).

Em uma análise visual, foi utilizado o método da média móvel, que suaviza os movimentos do preço, retirando os ruídos presentes, representados pelas oscilações mais fortes, neste caso é os sentimentos dos tweets. No Gráfico 2, é possível observar que a média móvel de 5 dias, se comporta bem parecido com a movimentação do Bitcoin, fica bem nítido este comportamento entre os meses de julho de 2022 até meados de outubro de 2022.

Gráfico 2 - Média móvel de 5 dias do grau de polaridade dos tweets com o preço de fechamento do Bitcoin.



Fonte: Elaboração própria (2023).

Com base no que foi apresentado até o momento, foi possível levantar duas hipóteses para esse trabalho:

H0: O sentimento expresso, via Twitter, possui influência no preço de fechamento do Bitcoin;

H1: Não se pode afirmar que o sentimento expresso, via Twitter, possui influência no preço de fechamento do Bitcoin.

3.3. MODELO ARDL PARA ANÁLISE

O modelo ARDL, é a sigla para AutoRegressive Distributed Lag, o qual é um tipo de modelo de análise de séries temporais. Este modelo geralmente é utilizado para analisar a relação de longo prazo entre as variáveis econômicas, considerando tanto os efeitos de curto prazo quanto os de longo prazo. O diferencial deste modelo consiste em incluir termos de defasagem, o que particularmente é útil para analisar modelos que possuem forte cointegração entre as variáveis, sendo a cointegração a presença de uma relação de equilíbrio de longo prazo entre as variáveis não estacionárias. Sendo o modelo ARDL utilizado para estimar essa relação e avaliar como as variáveis independentes afetam a variável dependente no curto e longo prazo.

Para testar a hipótese em questão, foi utilizado o modelo ARDL, o qual é uma extensão do modelo AR, e que apesar de pouco conhecido na literatura brasileira em comparação aos modelos de Engle e Granger (1987), Phillips e Hansen (1990) e Johanssen (1990), destaca-se pela sua flexibilidade. O modelo ARDL não exige que os métodos de cointegração sejam de ordem zero, conforme salientado por Gama *et al.* (2019).

De acordo com Gama *et al.* (2019), uma das vantagens do modelo ARDL é sua capacidade de estimar de forma simultânea parâmetros de curto e longo prazos, enriquecendo assim o conjunto de informações sobre os dados. O modelo foi aplicado para avaliar se os tweets ainda influenciavam os tweets ao longo dos dias.

Para a construção do modelo, aplicou-se o logaritmo a todas as variáveis a fim de padronizá-las. As variáveis grau de polarização (*compound*) e volume foram utilizadas como variáveis independentes, enquanto o preço de fechamento do Bitcoin, como variável dependente, como representado na Equação 1 subsequente.

$$\Delta \ln(\text{Preço}_{BTC_t}) = \alpha + \beta_1 \ln(\text{Preço}_{t-1}) + \beta_2 \Delta \ln(\text{compound}_t) + \beta_3 \Delta \ln(\text{Volume}_t) + \varepsilon_t$$

Onde:

- Δ representa a diferença entre os valores das variáveis em t e $t-1$, o que ajuda a tornar as séries temporais estacionárias;
- $Preço_{BTC_t}$ é o preço do Bitcoin no período t ;
- $compound_t$ é a variável "compound" no período t ;
- $Volume_t$ é o volume de negociações do Bitcoin no período t ;
- α é o intercepto, que captura os efeitos constantes ou não observados;
- β_1 é o coeficiente que representa o efeito de curto prazo do preço do Bitcoin em relação ao seu valor anterior;
- β_2 é o coeficiente que representa o efeito de curto prazo do "compound" no preço do Bitcoin;
- β_3 é o coeficiente que representa o efeito de curto prazo do volume no preço do Bitcoin;
- ε_t é o erro aleatório no período t

4. RESULTADOS E DISCUSSÃO

Primeiramente, procede-se uma análise descritiva dos dados, após a limpeza da base de dados, com o intuito de observar os dados em sua forma mais bruta. Através destas observações, busca-se obter uma compreensão do que os dados informam, além de identificar quaisquer valores discrepantes, que possam afetar o valor das demais medidas.

Conforme apresentado na Tabela 4, a média da variável “*compound*” (sentimento dos investidores) demonstrou um valor positivo em torno de 0.22. O qual indica que ao longo de toda a série temporal, o sentimento do investidor manteve-se positivo na média. Além disso, é possível constatar que a média do preço de fechamento do Bitcoin foi de US\$ 37,594.05, com seu valor mínimo em US\$ 15,787.28 e seu valor máximo atingindo US\$ 65,992.84.

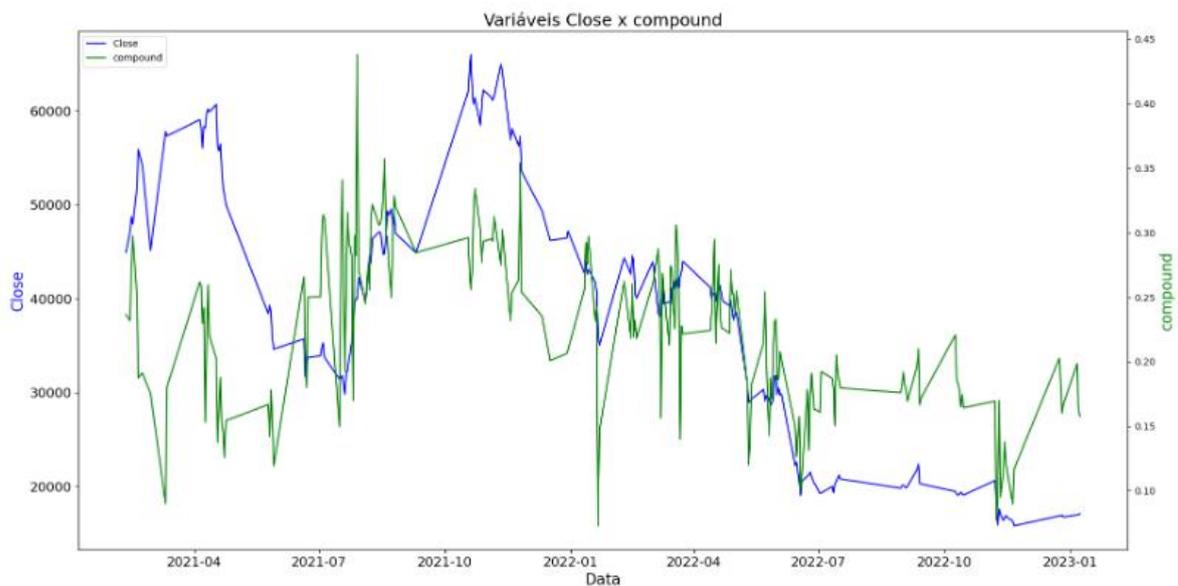
Tabela 4 – Estatística descritiva dos dados

	compound	Close	Volume
Count	216.000000	216.000000	216.0000e+02
Mean	0.220956	37594.045473	3.667874e+10
Std	0.063339	13836.068173	1.763889e+10
Min	0.072619	15787.284180	7.714767e+09
25%	0.173749	22522.242188	2.494371e+10
50%	0.221061	39372.863281	3.298254e+10
75%	0.270524	46512.927734	4.288384e+10
Max	0.438071	65992.835938	1.189925e+11

Fonte: Elaboração Própria (2023)

Nos Gráficos 3 e 4, foram representados graficamente as relações entre a variável dependente e as variáveis independentes, com o propósito de avaliar possíveis correlações no período analisado. No Gráfico 3, as variáveis “Close” e “compound” foram plotadas para verificar se as variáveis possuem correlação entre si. Conforme observado no Gráfico 3, ao longo da maior parte da série temporal, o preço de fechamento manteve-se em linha os sentimentos dos investidores.

Gráfico 3 – Varável dependente (Close) x independente (compound)

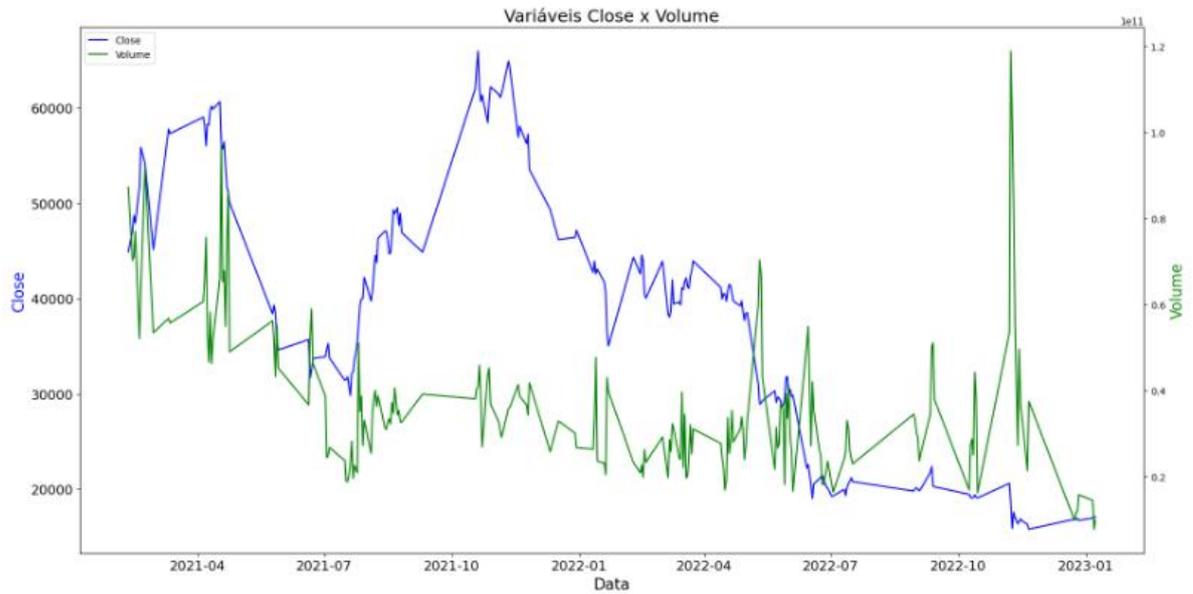


Fonte: Elaboração Própria (2023)

No Gráfico 4, é possível observar que existe correlação entre o volume de transações e o preço de fechamento, embora seja importante notar que houve um aumento substancial no volume a partir de abril de 2022, atingindo seu pico em meados de novembro. Essa elevação em novembro pode ser atribuída a possibilidade de quebra da corretora FTX². Após este período, o preço manteve-se estável e houve uma ligeira diminuição do volume de transações.

Além disso, durante o segundo semestre de 2021, enquanto o preço do Bitcoin atingia as máximas históricas, o volume não apresentou variações significativas.

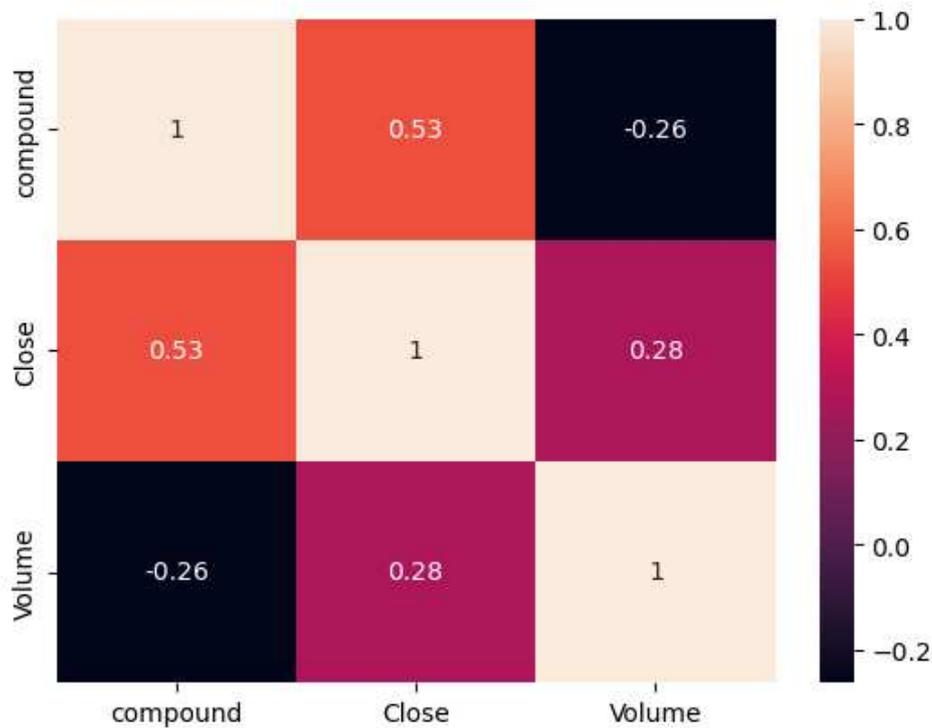
Gráfico 4 - Varável dependente (Close) x independente (Volume)



Fonte: Elaboração Própria (2023)

No Gráfico 5, realizou-se uma análise estatística de correlação entre as variáveis do modelo ARDL. Nesse contexto, a correlação entre as variáveis “preço de fechamento” e “*compound*” demonstrou ser positiva em 0.53. Além disso, observou-se uma correlação positiva de 0.28 entre as variáveis “preço” e “volume”. No entanto, a correlação entre as outras duas variáveis foram negativas, com um valor de -0.26.

Gráfico 5 - Correlação entre as variáveis



Fonte: Elaboração Própria (2023)

4.1. ANÁLISE DO MODELO DE REGRESSÃO LINEAR MÚLTIPLA

Para a parte estatística da pesquisa, foi utilizada a biblioteca *statsmodels*, no Python, que consiste em analisar e estimar modelos, esta biblioteca é aplicável em uma grande variedade de modelos. O *statsmodels.api* é o principal módulo da biblioteca, este módulo é capaz de aplicar modelo de regressão linear múltipla, o qual usaremos para a pesquisa, além de também ser útil para análise de séries temporais, ANOVA, ARIMA e outros modelos.

Com este módulo, é possível utilizar o método dos mínimos quadrados (OLS), o qual é aplicado nesta pesquisa pela sua definição das variáveis dependentes e independentes, obtendo assim o resultado do modelo de regressão.

De acordo com o resultado do modelo de análise de regressão linear múltipla, **a hipótese nula é rejeitada**, não existindo evidência estatística de que o sentimento expresso via *Twitter* possui correlação com o preço de fechamento do Bitcoin. O resumo dos resultados está nas Tabela 4 e 5.

A primeira observação dos resultados é olhar o R^2 (R-squared) do modelo, que é uma proporção que ajuda a entender o quão próximo estão os dados da linha de regressão ajustada, ou seja, é a porcentagem da variação da variável resposta, que é explicada por um modelo linear, cuja variação é entre 0 e 1, quanto mais próximo de 1, melhor o modelo se ajusta aos dados.

No resultado do modelo, o R^2 foi de 0.474, indicando que o modelo é fraco para explicar a variabilidade dos dados respostas ao redor de sua média. Somente o R^2 não determina se as previsões são tendenciosas ou não, logo é indicado avaliar os Gráficos dos resíduos, que será apresentado no tópico 4.3.

Tabela 5 – Resultados estatísticos do modelo de regressão linear múltipla

OLS Regression Results			
Dep. Variable:	Close	R-squared:	0.474
Model:	OLS	Adj. R-squared:	0.469
Method:	Least Squares	F-statistic:	95.79
No. Observations:	216	Prob (F-statistic):	2.12e-30
Df Residuals:	213	Log-Likelihood:	-2296.3
Df Model:	2	AIC:	4599
Covariance Type:	nonrobust	BIC:	4609

Fonte: Elaboração própria (2023)

O *p-valor* é o nível de significância individual dos parâmetros, sendo a hipótese nula considerada verdadeira, a qual devemos rejeitar ou não no seu teste. No resultado do teste, todos os coeficientes são zero em relação ao preço de fechamento, valores bem menores do que os 0.05 para não ser rejeitada a hipótese nula, neste caso, as variáveis independentes não afetam o preço de fechamento do Bitcoin.

Com estes testes, é possível concluir que as variáveis são estatisticamente significantes, esta significância pode ser explicada pela alta correlação entre as variáveis.

Tabela 6 – Resultados dos coeficientes

	coef	Std err	t	P > t 	[0.0025	0.975]
Const	-6848.5955	3286.667	-2.084	0.038	-1.33e+04	-370.035
Compound	1.426e+05	1.13e+04	12.662	0.000	1.2e+05	1.65e+05
Volume	3.529e-07	4.04e-08	8.729	0.000	2.73e-07	4.33e-07
Omnibus:	1.972					
Prob(Omnibus):	0.373					
Skew:	0.190					
Kurtosis:	3.189					
			Durbin-Watson:	0.497		
			Jarque-Bera (JB):	1.626		
			Prob(JB):	0.443		
			Cond. No.	6.89e+11		

Fonte: Elaboração própria (2023)

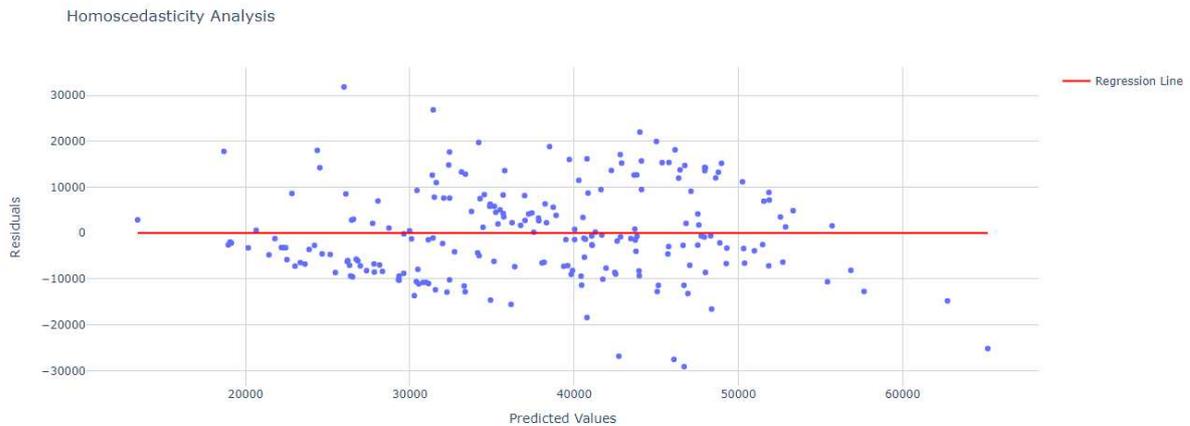
4.2. ANÁLISE DOS RESÍDUOS

Pelo Gráfico 6, é possível observar de forma bem clara, que os resíduos não seguem uma distribuição normal. Através do mesmo Gráfico, também é possível observar que os resíduos não seguem a hipótese de homoscedasticidade.

Analisamos a homoscedasticidades dos resíduos (ou erros) ou seja identificar se os resíduos possuem uma variância constante em relação as variáveis independentes, a forma visual de identificar a homoscedasticidade dos resíduos é por meio do Gráfico de dispersão, nele os resíduos devem estar distribuídos uniformemente ao redor da linha zero.

No Gráfico 6, podemos observar que as variâncias estão bem dispersas, indicando não haver homoscedasticidade nos resíduos, ou seja, a variância dos resíduos está mudando de acordo com a mudança dos valores das variáveis independentes.

Gráfico 6 – Gráfico de homoscedasticidade dos resíduos



Fonte: Elaboração própria (2023)

Para analisar, de forma estatística, se os resíduos seguem uma distribuição normal, é aplicado o teste de Shapiro.

Hipóteses para o teste de Shapiro:

H0: Há normalidade na distribuição dos dados $\rightarrow p\text{-value} > 0.05$

H1: Não há normalidade na distribuição dos dados $\rightarrow p\text{-value} \leq 0.05$

Onde:

H0: Hipótese nula;

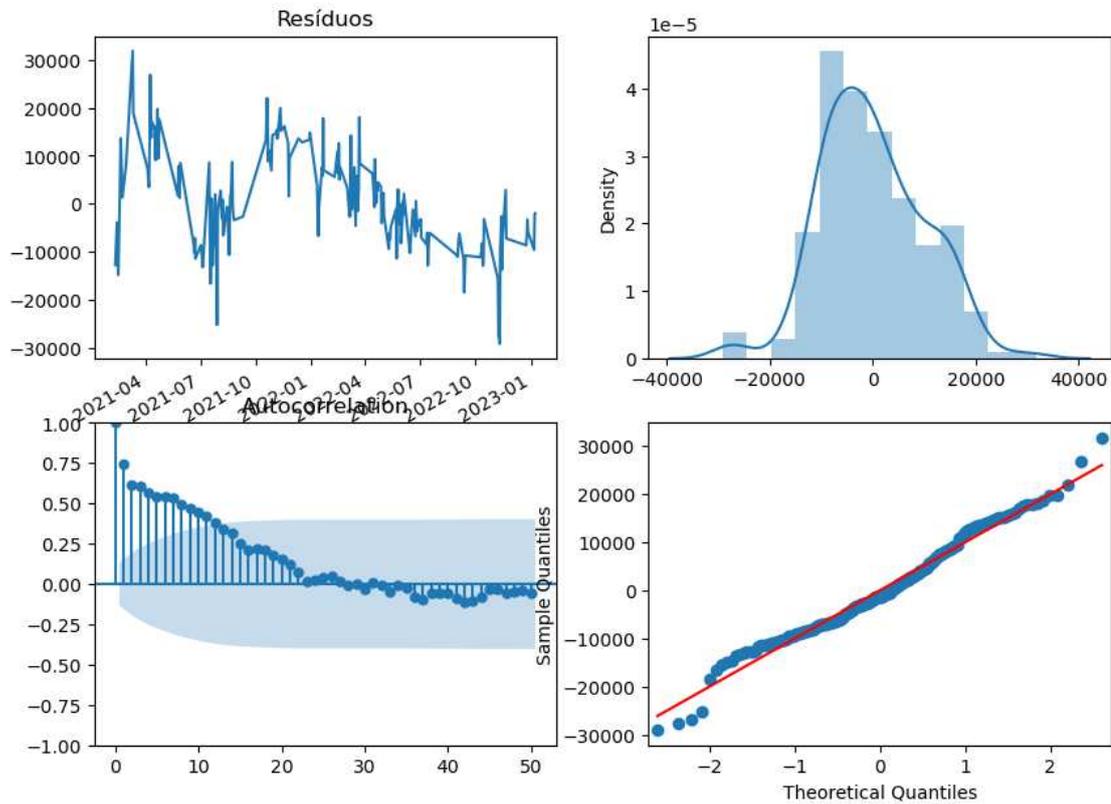
H1: Hipótese alternativa.

O *p-value* do teste de Shapiro foi em de aproximadamente 0.0025, rejeitando-se a hipótese nula de que os resíduos são normalmente distribuídos. No Gráfico 7, é possível observar o histograma, e inferir que ele não se aproxima de distribuição normal.

No Gráfico 7, é possível observar um decaimento na primeira *lag* no Gráfico de Autocorrelação, sugerindo a possibilidade de ser um modelo AR. É importante notar que com base no histograma e no *Q-Q plot*, que analisam a distribuição dos resíduos, os resíduos não seguem uma distribuição normal.

Dadas essas informações, a análise do Gráfico de Autocorrelação Parcial mostra-se ideal para corroborar com o estudo e avaliar de forma mais profunda a partir desta informação.

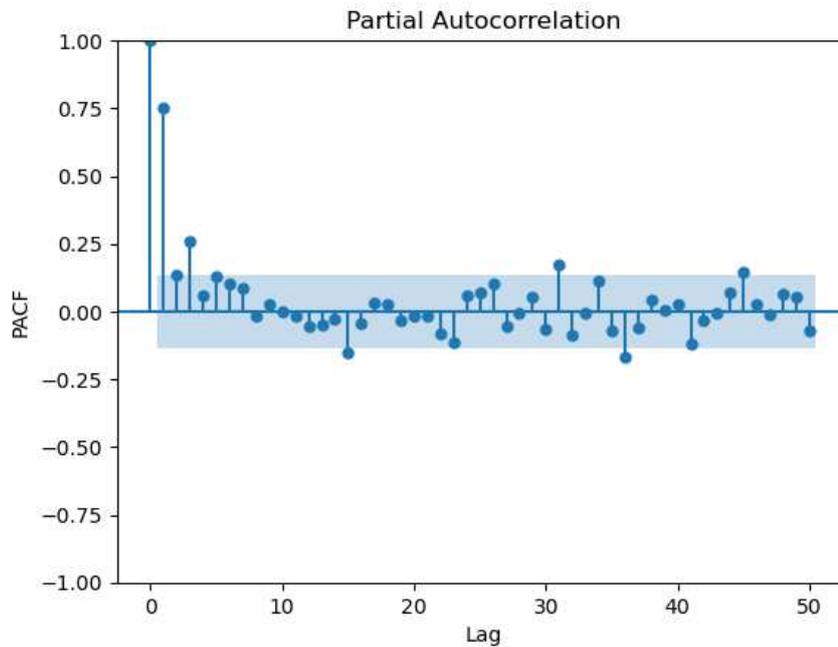
Gráfico 7 – Quadro com a distribuição dos resíduos na série temporal, histograma dos resíduos, autocorrelação e Q-Q plot



Fonte: Elaboração própria (2023)

No Gráfico 8, é aplicada a função de autocorrelação parcial (PACF), através dele é possível inferir que a correlação positiva nas 2 primeiras *lags*, as quais indicam que as variáveis são correlacionadas.

Gráfico 8 – Teste de autocorrelação parcial



Fonte: Elaboração própria (2023)

Com o teste de autocorrelação parcial, foi possível observar que as variáveis não são correlacionadas, sendo possível utilizar um modelo mais flexível, como o modelo ARDL. Na Tabela 7, o modelo gerado com a variável *compound*, que representa a polaridade dos *tweets*, resulta que os *tweets* ainda possuem influência nos preços do Bitcoin, mesmo após alguns dias.

Tabela 7 – Modelo ARDL (continua)

Modelo ARDL						
Dep. Variable:	lnClose	S.D. of innovations:	0.219			
Model:	ARDL (0, 5, 2)	AIC:	-21.588			
Method:	Conditional MLE	BIC:	15.540			
No. Observations:	216	HQIC:	-6.588			
Log Likelihood:	21.794					
	coef	Std err	z	P > z 	[0.025	0.975]
Const	-0.0631	1.081	-0.058	0.954	-2.195	2.069
ln.compound.L0	0.3609	0.076	4.750	0.000	0.211	0.511
ln.compound.L1	0.2325	0.084	2.769	0.006	0.067	0.398
ln.compound.L2	0.2667	0.083	3.221	0.001	0.103	0.430
ln.compound.L3	0.1342	0.080	1.677	0.095	-0.024	0.292
ln.compound.L4	0.1091	0.080	1.365	0.174	-0.049	0.267
ln.compound.L5	0.1295	0.072	1.808	0.072	-0.012	0.271
lnVolume.L0	0.1687	0.056	3.006	0.003	0.058	0.279

	coef	Std err	z	P > z 	[0.025	0.975]
lnVolume.L1	0.1247	0.064	1.949	0.053	-0.001	0.251
lnVolume.L2	0.2200	0.057	3.878	0.000	0.108	0.332

Fonte: Elaboração própria (2023)

Com base nos resultados obtidos, o presente estudo parece estar em contradição com o que é proposto por Engelen e Kulcsár (2023), na qual os autores questionam se o sentimento dos investidores influencia nos preços das criptomoedas e retornos associados a ela. Como exposto através do modelo ARDL (0,5,2), é possível inferir que através dos resultados os tweets e os seus sentimentos expressos neles, representados pela variável “*compound*”, podem ter influência sobre os preços da criptomoeda Bitcoin.

5. CONSIDERAÇÕES FINAIS, LIMITAÇÕES E RECOMENDAÇÕES

Pelo executado, pode-se concluir haver relação entre os *tweets* analisados e o preço do bitcoin, porém com divergência quando se analisa duas abordagens: o modelo ARDL indica a existência e o modelo de regressão linear não. Entretanto, dada a condição do modelo ARDL ser o mais adequado para o estudo, conclui-se haver relação entre sentimentos dos *tweets* e preços do bitcoin. Isso pode ser corroborado ao se estudar a relação conjunta dos *tweets* ou fatores exógenos a publicação do *tweets*, como publicações em jornais ou em outras mídias.

Infelizmente, durante a realização do trabalho, o *Tweeter* reduziu o número de *tweets* que poderiam ser baixados e como não há muitos dados disponíveis sobre o mercado acionário brasileiro, a situação foi contornada pelo uso de uma base pronta em inglês. Após a limpeza da base de dados, a nova base usando poucas variáveis, o programa ainda demorava em torno de 2h para poder fazer a leitura. Mas um computador de nova geração com SMD poderia reduzir o tempo computacional.

Futuros trabalhos poderão vir da combinação das variáveis de bancos de dados já estudadas, com modelos de regressão, como o modelo ARDL, em que não é tão conhecido, mas que possui uma funcionalidade excelente.

Como sugestão, é possível usar o modelo ARDL em estudos brasileiros já realizados e comparar com outros, qual o melhor a ser usado. Também usar bases de dados de outros países, pode enriquecer a literatura brasileira e facilitar quem está começando a estudar o tema, principalmente ligados ao tema de linguagem natural.

6. REFERÊNCIAS

BOLLEN, Johan; MAO, Huina; ZENG, Xiao-Jun. **Twitter mood predicts the stock market**. *Journal of Computational Science*, v. 2, n. 1, p. 1-8, 2011.

FAMA, Eugene. **Efficient Capital Markets: A Review of Theory and Empirical Work**. *The Journal of Finance*, v. 25, n. 2, p. 383-417, mai/1970.

ELBAGIR, Shihab; YANG, Jing. **Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment**. *Proceedings of the International Conference on Information Systems and Software Technologies*, 2019.

ENGELEN, Frank Van; KULCSAR, Levante. **Twitter Sentiment Analysis On The Cryptocurrency Market**, 2023.

FATALYEV, Kamaladdin; CHIVULUKA, Aneesh; PRASAD, Mukesh; LIU, Wei. **Stock Market Analysis with Text Data: A Review**. Disponível em: <https://doi.org/10.48550/arXiv.2106.12985>. Acesso em: 13/06/2022.

GAMA, Fabio; BASTOS, Suzana Quinet de Andrade; CARVALHO, Luíza; PAIVA, Luciano. **Assimetria na Comunicação do Banco Central e Investimento: Uma aplicação do modelo ARDL**. Disponível em: <http://corecon-sc.org.br/anais-cbe2019/arquivos/anais-cbe2019-3809.pdf>. Acesso em: 01/06/2023.

HUYNH, Huy; DANG, L. Minh; DOUG, Duc M. **A new model for stock price movements prediction using deep neural network**. *SoICT*, 2017. Disponível em: DOI: 10.1145/3155133.3155202. Acesso em: 16/06/2022.

KANSAON, Daniel P.; BRANDÃO, Michele A.; PINTO, Saulo A. de Paula. **Análise de Sentimentos em Tweets em Português Brasileiro**. In: BRAZILIAN WORKSHOP ON

SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM), 7., 2018, Natal. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2018. ISSN 2595-6094. DOI: <https://doi.org/10.5753/brasnam.2018.3578>.

KOSALANI, S.V.; ASSAF, R. **Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks**. Journal of Data Analysis and Information Processing, 8, 309-319, 2020. Disponível em: <https://doi.org/10.4236/jdaip.2020.84018>. Acesso em: 16/06/2022.

KRAAIJEVELD, Olivier; DE SMEDT, Johannes. **The predictive power of public Twitter sentiment for forecasting cryptocurrency prices**. Journal of International Financial Markets, Institutions and Money, 2020. Disponível em: <https://doi.org/10.1016/j.intfin.2020.101188>. Acesso em: 30/06/2023.

MASSARO, Andre. **Entenda a Hipótese dos Mercados Eficientes**. Andre Massaro, 2020. Disponível em: www.andremassaro.com.br/hipotese-dos-mercados-eficientes/. Acesso em: 01/02/2022.

MORRISSEY, Martha; WASSER, Leah; DIAZ, Jeremy; PALOMINO, Jenny. **Lesson 3. Analyze Word Frequency Counts Using Twitter Data and Tweepy in Python**. Earth Data Science, 2020. Disponível em: <https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/calculate-tweet-word-frequencies-in-python/>. Acesso em: 18/06/2023.

MORRISSEY, Martha; WASSER, Leah; DIAZ, Jeremy; PALOMINO, Jenny. **Lesson 5. Analyze The Sentiment of Tweets From Twitter Data and Tweepy in Python**. www.andremassaro.com.br/hipotese-dos-mercados-eficientes. Disponível em: <https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/analyze-tweet-sentiment-in-python/>. Acesso em: 18/06/2023.

MUSSA, Adriano; YANG, Edward; TROVAO, Ricardo; FAMA, Rubens. **Hipótese de Mercados Eficientes e Finanças Comportamentais – As Discussões Persistem**. Revista FACEF, São Paulo, v.11, n.1, pag. 5-17, mar/2008.

PEARSON, KARL. **Random Walk, 1905**. Disponível em: en.wikipedia.org/wiki/Random_walk. Acesso em: 28/01/2022.

REIS, TIAGO. **Hipótese do mercado eficiente: entenda mais sobre essa teoria**, 2018.
Disponível em: <https://www.sunoo.com.br/artigos/mercado-eficiente/>. Acesso em: 01/02/2022.

SCHUMAKER, Robert P.; ZHANG, Yulei; HUANG, Chun-Neng; CHEN. HSINCHUN.
Evaluating Sentiment in Financial News Articles, 2012.

SILVA, N. F. F. d. **Análise de sentimentos em textos curtos provenientes de redes sociais**.
Tese (Doutorado) — Universidade de São Paulo, 2016.

SILVA, Kenia Silveiro da. **Como variáveis econômicas afetaram o preço do Bitcoin antes e depois da pandemia**. Dissertação de mestrado – FGV, 2022.

SOUZA, Dyliane Mouri Silva de; et. Al. **O Efeito do Sentimento do Investidor Expresso via Twitter sobre o Comportamento do Mercado Acionário Brasileiro Durante o Período Eleitoral**. XIX USP International Conference in Accounting, 24 a 26 de Julho de 2019.

SOUZA, Dyliane Mouri Silva de. **Captando o sentimento do investidor na bolsa via Twitter**. Disponível em: <https://tc.com.br/blog/renda-variavel/twitter-bolsa-de-valores>. Acesso em: 07/02/2022

SOUZA, Dyliane Mouri Silva de. **Efeito Do Sentimento do Investidor Manifesto Via Twitter Sobre os Retornos e o Volume Negociado No Mercado Acionário Brasileiro**. Dissertação de mestrado – UFPB, 2020.

TEIXEIRA, Carolina Barcelos. **Análise De Sentimento Dos Usuários Do Twitter em Relação à Atual Situação Política Do Brasil**, 2019.