Topic:
Sentiment analysis on tweets for Bitcoin price trend prediction and validation, comparing tweets from the public and from Elon Musk's influence on Bitcoin price.

Database:
Tweets about Bitcoin: https://www.kaggle.com/alaix14/bitcoin-tweets-20160101-to-20190329
Actual Price for Bitcoin: https://www.kaggle.com/prasoonkottarathil/btcinusd
All Elon Musk's Tweets: https://www.kaggle.com/andradaolteanu/all-elon-musks-tweets
Dogecoin historical Data(08.03.2019 - 05.05.2021):
https://www.kaggle.com/zackerym/dogecoin-historical-data-03082019-05052021

Introduction:

The development of cryptocurrency, digital coins that are not issued by any legal entity like fiat currencies(dollar or yen),[4] in the past decade was rapid. Followed by a white paper published by Satoshi Nakamoto in 2008, the first cryptocurrency, bitcoin, was introduced into the world. Increasing gradually in value at the beginning, the price of bitcoin skyrocketed to $290 in mid-2015. [2] Eventually, The price of bitcoin climbed as high as $63,729.5, and it has fluctuated up and down unremittingly since then. However, bitcoin tumbled to $32, 639, 7% of its value in one morning. It is not the only case, many other famous cryptocurrencies burgeoned followed by the growth of bitcoin. A massive number of them went down severely in a ridiculously short period. Dogecoin was down 10% to $0.209537, and Ether was down 10.0% to $2,149.15. All happened in the same morning Bitcoin devalued, and thus all of the cryptocurrencies are considered very volatile.

The reason behind this abnormality and "chaos" is that there existed some relationships between social media and the price of the cryptocurrency, bitcoin specifically. As a result, similar to most commonly known currencies, cryptocurrencies are also affected by socially constructed opinions but on a more severe scale. [9] On January 29, 2021, Elon Musk changed the bio of his Twitter account to #bitcoin. This simple action has made the price of bitcoin rise around

$6,000[15]. Another example about how much Bitcoin/cryptocurrency is affected by social media is that on June 3, 2021, Musk simply tweeted a MEME about breaking up with bitcoin, and the price of bitcoin declined 5 percent as a consequence. Musk is only one example to show how the market of Bitcoin is deeply related to social media like Twitter.

Granted, grasping a precise understanding of cryptocurrency is not so easy[4]. It will still be helpful for prospective investors if there's a way to predict the trend of Bitcoin price by examining the sentiment of news, tweets specifically. In this research, we make contributions mainly on 1) implementing machine learning models to gain a more considered understanding of the relationship between sentiment of bitcoin-related tweets and bitcoin price and 2) comparing the impact of Elon Musk's tweets and all tweets related to Bitcoin on the price trend by the public . The rest of this paper is organized as follows…(will be completed later).

Literature Review

Abraham et al.(2018) [5] and Chen et al.(2020) [17] both leveraged Google Trend Search Volume Index(SVI) and found a strong correlation between SVI and bitcoin price. Abraham et al.(2018) [5] also includes tweets volume and found that sentiment of tweets was not a determined indicator when bitcoin price was falling. However, Stenqvist and Lonno(2017) [10] have analyzed sentiment fluctuation of over 2 million bitcoin-related tweets and gained a model predicting bitcoin price with 79% accuracy. Raju and Tarif [13] have also shown that bitcoin price is predictable using sentiment analysis of bitcoin-related tweets. They implemented machine learning method Recurrent Neural Networks(RNNs) with Long Short Term Memory(LSTM) and standard method ARIMA, and both models achieved a high accuracy. Thus, we believe that sentiment analysis can be an indicator of bitcoin price.

Ante(2021) [15] has revealed that Elon Musk, who has 44.7 million followers on Twitter, has extremely powerful influence on the cryptocurrency market and has caused a few abnormal trading volumes. They have also discovered abnormal returns of up to 18.99% for Bitcoin and 17.31% for Dogecoin after Musk's tweets. This research shows that magnates like Musk exert more impacts on the market. However, Mai et al.(2018) [19] implemented a VECM model to study the relationship between social media and the monetary value of bitcoin, and revealed that the silent majority is the group who exerts significant effects, that is to say, users who are less active are the ones who dominate the bitcoin price.

Sources:

1. Cryptocurrency and the problems
   https://www-jstor-org.ezproxy.drew.edu/stable/44000162?seq=1#metadata_info_tab_contents
   Can be Used to describe why to examine the matter of bitcoin is important

2. Bitcoin and the Future of Digital Payments
   https://www-jstor-org.ezproxy.drew.edu/stable/24562161?Search=yes&resultItemClick=true&searchText=bitcoin&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3Dbitcoin%2B%26filter%3D&ab_segments=0%2Fbasic_search_gsv2%2Fcontrol&refreqid=fastly-default%3A0ea2c0051f39c808888312dae9ede8a3&seq=1#metadata_info_tab_contents
   Can be Used to describe why to examine the matter of bitcoin is important

3. Bitcoin tumbles below $30,000 for the first time since January. (Not used for reference)
   https://www.nytimes.com/2021/06/22/business/bitcoin-cryptocurrency-price-decline.html?searchResultPosition=1

4. What drives Bitcoin price?
   http://www.accessecon.com/Pubs/EB/2016/Volume36/EB-16-V36-I2-P82.pdf
   Examine factors driving bitcoin price using Empirical Mode Decomposition

5. Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis

https://scholar.smu.edu/cgi/viewcontent.cgi?article=1039&context=datasciencereview
In this paper, we present a method for predicting changes in Bitcoin and Ethereum prices utilizing Twitter data and Google Trends data.

6. Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis
https://www.researchgate.net/profile/Prasanga-Neupane/publication/329910345_Recurrent_Neural_Network_Based_Bitcoin_Price_Prediction_by_Twitter_Sentiment_Analysis/links/5cf524a3a6fdcc847500a397/Recurrent-Neural-Network-Based-Bitcoin-Price-Prediction-by-Twitter-Sentiment-Analysis.pdf
Using RNN to accurately predict bitcoin price

7. Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis
http://cs229.stanford.edu/proj2015/029_report.pdf
"Our approach to cleaning data and applying supervised learning algorithms such as logistic regression, Naive Bayes, and support vector machines leads to a final hour-to-hour and day-to-day prediction accuracy exceeding 90%"

8. A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19
https://www.mdpi.com/2504-2289/4/4/33
We explore the effect of different preprocessing functions, features, and time lengths of data on the correlation results. Out of 13 strategies, we discover that splitting sentences, removing Twitter-specific tags, or their combination generally improve the correlation of sentiment scores and volume polarity scores with Bitcoin prices. The prices only correlate well with sentiment scores over shorter time spans. Selecting the optimum preprocessing strategy would prompt machine learning prediction models to achieve better accuracy as compared to the actual prices.

9. Predicting Bitcoin price fluctuation with Twitter sentiment analysis
https://www.diva-portal.org/smash/get/diva2:1110776/FULLTEXT01.pdf

10. Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2607167
This paper uses time-series analysis to study the relationship between Bitcoin prices and fundamental economic variables, technological factors and measurements of collective mood derived from Twitter feeds.

11. Bitcoin Spread Prediction Using Social And Web Search Medi
    https://d1wqtxts1xzle7.cloudfront.net/40425066/Bitcoin_Spread_Prediction_Using_Social_A20151127-10804-12xrxm4-with-cover-page-v2.pdf?Expires=1626319696&Signature=eEZHzycm8Z0z1ysM1MyqNQnE05asLW7pWz1cU5xE01E5qHxHTgSq8pA3WtYOxFcY58hGEPPdHRlhzh1xh3Ct0z08pPgUv6p12onQEYgeCLGVFghDFJmvZ5m06Yhn8kRmdLP0HbH0xjFhPcks56pFmjL~FTh6DnNVlqfz2A4p9bfJ3Dy63bDpJE2bM3RDFhJb3PF5qRv8ueAVi~5mNGsxSQFAHCM0tN6U9pJXrII0uhKwClQi3be1MiIIDUcms2dnuOZevVRlS~5RFVkdPTN9w3iDMWSZB~ZeceyFLaGLI97RS-7aZVKKasmefZNEP~MKR33Hqq3R3iklTl5oyGHEuw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
    In this work, we investigated if the spread of the Bitcoin's price is related to the volumes of tweets or Web Search media results. We compared trends of price with Google Trends data, volume of tweets and particularly with those that express a positive sentiment. We found significant cross correlation values, especially between Bitcoin price and Google Trends data, arguing our initial idea based on studies about trends in stock and goods market.

12. An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information
    https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8125674
    This paper reveals the effect of Bayesian neural networks (BNNs) by analyzing the time series of the Bitcoin process. We also select the most relevant features from Blockchain information that is deeply involved in Bitcoin's supply and demand and use them to train models to improve the predictive performance of the latest Bitcoin pricing process.

13. Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis

    https://arxiv.org/pdf/2006.14473.pdf

The objective of this paper is to determine the predictable price direction of Bitcoin in USD by machine learning techniques and sentiment analysis.

14. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning
    https://www.mdpi.com/1099-4300/21/6/589
    In this paper, we propose the usage of common machine learning tools and available social media data for predicting the price movement of the Bitcoin, Ethereum, Ripple and Litecoin cryptocurrency market movements. We compare the utilization of neural networks (NN), support vector machines (SVM) and random forest (RF) while using elements from Twitter and market data as input features.

15. How Elon Musk's Twitter Activity Moves Cryptocurrency Markets

    https://www.researchgate.net/profile/Lennart-Ante/publication/349007665_How_Elon_Musk%27s_Twitter_activity_moves_cryptocurrency_markets/links/601ae03d92851c4ed548f55c/How-Elon-Musks-Twitter-activity-moves-cryptocurrency-markets.pdf
    Our study shows the significant impact that social media activity of influential and well-known individuals can have on cryptocurrencies.

16. Twitter-Based Uncertainty and Bitcoin Before and during the Pandemic

    https://www.mdpi.com/2227-7072/9/2/28/htm
    We investigated the differential impacts of a new Twitter-based Market Uncertainty index (TMU) and variables for Bitcoin before and during the COVID-19 pandemic.

17. Zheshi Chen, Chunhong Li, Wenjun Sun, Bitcoin price prediction using machine learning: An approach to sample dimension engineering, Journal of Computational and Applied Mathematics, Volume 365, 2020, 112395, ISSN 0377-0427, https://doi.org/10.1016/j.cam.2019.112395.

    This paper presents a representative process of price prediction. The authors leveraged Google Trend search volume index and Baidu media search volume with typical features including price and number of transactions. In order to enrich our research, we could take the search volume index into account as well.
    The Introduction section is inspiring. We could learn from it and start ours with how Bitcoin price is strongly dominated by investors' confidence and how we evaluate confidence. What's creative is that we take real reviews and comments on social platforms instead of indexes as features.
    Also, the Implementation and Results sections are worth a look. It's organized and articulated.

18. David Yermack, Chapter 2 - Is Bitcoin a Real Currency? An Economic Appraisal, Editor(s): David Lee Kuo Chuen, Handbook of Digital Currency, Academic Press, 2015, Pages 31-43, ISBN 9780128021170, https://doi.org/10.1016/B978-0-12-802117-0.00002-3.

    This article introduces Bitcoin from an economic aspect and gives us a basic idea of how Bitcoin differs from widely used currencies. It could be paraphrased and used in our introduction section.

19. Feng Mai, Zhe Shan, Qing Bai, Xin (Shane) Wang & Roger H.L. Chiang (2018) How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis, Journal of Management Information Systems, 35:1, 19-52, https://web.a.ebscohost.com/bsi/detail/detail?vid=0&sid=9ad7fc9a-2a81-40c8-a8e0-602418f8cb42%40sessionmgr4006&bdata=JnNpdGU9YnNpLWxpdmU%3d#AN=128814841&db=buh

    This paper has an elaborated methodology from which we can learn and strong analysis based on their VECM model. They studied the relationship between social media and the monetary value of bitcoin and revealed that the silent majority is the one who exerts significant effects, that is to say, users who are less active are the ones who dominate the bitcoin price.
    *This result is interesting and inspiring. We don't need to specify our topic at this point, but once we got something from the trained models, we could dive deeper and look for something interesting like the conclusion of this paper, e.g. which sentiment/word has the strongest correlation with price, posts at what time are more important.

20. Dominique Guégan, Thomas Renault, Does investor sentiment on social media provide robust information for Bitcoin returns predictability?, Finance Research Letters, Volume 38, 2021, 101494, ISSN 1544-6123, https://doi.org/10.1016/j.frl.2020.101494.

    This paper contributes to the discussions on the efficiency of the Bitcoin market and on the determinants of Bitcoin returns by exploring the relationship - at various time-frequency - between investor sentiment on social media and the evolution of the price of Bitcoin. The dataset they used was from StockTwits on which users classify their posts as bullish or bearish by themselves. The dataset is simpler than what we plan to use, but the authors only focused on whether investor sentiment provides information, so this dataset matches their goal. However, a self-classified dataset might not be that accurate, we will include processing the raw dataset and classification as part of our research.

21. Burnie, A., & Yilmaz, E. (2019). Social media and bitcoin metrics: which words matter. *Royal Society open science*, *6*(10), 191068. https://doi.org/10.1098/rsos.191068