Natural Language Processing for the Social Sciences (GR5067) Fall 2023

Uncovering the Relationship Between Elon Musk's Tweets & Bitcoin Stock Prices

Group 12

Yunxi Yang (yy3297), Zozo Chunyu (zc2691), Pei Liu (pl2848),
Yiming Ma (ym2912), Sam Barnett (snb2160), Nujailah Noor(nn2615)

Abstract

One of the most pressing questions about Bitcoin with respect to its validity as currency remains unresolved: Where does its value come from? Fluctuations in the price of Bitcoin and other cryptocurrencies have previously been found to be correlated with social media sentiment (Ifigeneia 2015), and individual statements from widely-followed accounts can possess an outsize influence on financial markets, including cryptocurrency markets (DeCambre 2021, Ante 2021). In this study, we aim to determine the influence of Tesla Motors CEO Elon Musk's Twitter posts on Bitcoin prices. We used sentiment analysis to predict the sentiment of Musk's tweets, and used the sentiment classification as one covariate in our Bitcoin price models. We also control for a number of other factors, including cryptocurrency volume and the values of several leading stock market indices, in an attempt to extend the correlational interpretation of previous studies. We find that the overall OLS-estimated effect of sentiment scores calculated from Musk's Twitter posts on Bitcoin prices is negligible and not statistically significant; other estimation procedures, including Lasso, produce similar results. We also built several machine learning models for Bitcoin price prediction, and compared the models using RMSEs and R2. We found the most optimal model for predicting Bitcoin's closing prices is Random Forest Regression with the Naive Bayes classification. We stress that these findings do not conclusively show that Musk's tweets have no influence on the cryptocurrency market.

Table of Contents

1 Introduction

Bitcoin is a decentralized digital currency, whose transactions are verified by network nodes through cryptography and recorded in a blockchain. This cryptocurrency was invented by Satoshi Nakamoto in 2008. Satoshi Nakamoto is known as the inventor of Bitcoin, however, it is a pseudonymous name of a person or persons who developed Bitcoin. The implementation of Bitcoin began in 2009, which marked the start of Bitcoin usage. In 2011, when Bitcoin got its first major users from the black market, the adjusted close price of Bitcoin was about 0.30 USD. Now, after 9 years, the price of Bitcoin has reached 29,993 USD on October 22, 2023, which means that the price of Bitcoin has grown about 64.59 times in value. Looking at the cryptocurrency market, we can find that the whole market has experienced significant growth in the past years, with Bitcoin emerging as one of the dominant cryptocurrencies. This significant growth of Bitcoin attracted several researchers to investigate the effects behind Bitcoin price, as it might reveal the economic condition and influence the investor's decision.

In 2015, Ifigeneia used 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 (Ifigeneia, 2015). They found that Twitter sentiment ratio is positively correlated with Bitcoin prices, and the number of Wikipedia search queries and the hash rate have a positive effect on the price of Bitcoin in the short-run analysis. Researchers believe that a higher degree of public recognition or interest in Bitcoin would lead to an increase in the market price of Bitcoin. They found that Twitter's sentiment ratio is positively correlated with Bitcoin prices. Among those tweets, there is one person whose words seem to have played a huge role in the impact of Bitcoin value. On January 29, 2021, Elon Musk, who was then the world's wealthiest person (Klebnikov, 2021), made a surprising move by changing his Twitter bio to #bitcoin. This led to a rapid increase in Bitcoin's value, soaring from about $32,000 to over $38,000 in just a few hours and boosting its market capitalization by $111 billion (Ante, 2021). The influence of Musk's tweets on financial markets has been evident in other instances as well. For example, his 2018 tweet about possibly taking Tesla private at $420 led to a fraud charge and a $40 million fine from the U.S. Securities and Exchange Commission. Additionally, his endorsement of the encrypted messaging app Signal in 2021 caused a mix-up where investors bought shares in the unrelated company Signal Advance, inadvertently raising its market value from $55 million to over $3 billion (DeCambre, 2021). These events underscore the significant impact that influential figures on social media can have on financial markets and the investment decisions of individuals.

This research project aims to study the potential "Musk Effect" in cryptocurrency markets recently brought up by Lennart Ante in 2021 (Ante, 2021). Specifically, the research seeks to determine the degree to which Elon Musk's Twitter sentiment influences short-term Bitcoin returns and volume. Furthermore, the research aims to determine whether or not the inclusion of Elon Musk's tweet sentiments will substantially enhance the performance of models predicting Bitcoin prices.

2 Data

2.1 Data Source & Definitions

2.1.1 Dataset for Elon Musk's Tweets

The data of Elon Musk's Tweets used in this project is retrieved from a dataset posted on Kaggle, named Elon Musk Tweets (2010-2022). The dataset includes Elon Musk's Tweets from January 1, 2010, to March 5, 2022. The original dataset has 36 columns, including id, conversation_id, created_at, date, time, timezone, user_id, username, name, tweet, etc. The tweets are divided into separate CSV files by year. Specifically, there are 3115 rows from 2021-01-01 to 2021-12-31 and 1028 rows from 2022-01-01 to 2022-03-05.

2.1.2 Dataset for Bitcoin daily price

The data of Bitcoin daily price used in this project is retrieved from a dataset posted on Kaggle, named Daily Bitcoin (Stock) Price. The dataset includes all cryptocurrency prices from 2015 to 2021. The original dataset has 7 columns, including date, open, high, low, close, Adjusted Close, and Volume. There are 2713 data rows from 2014-09-17 to 2022-02-19.

2.1.3 Dataset for Macroeconomic Indicators

The data of the macroeconomic indicators is retrieved from Yahoo Finance (Python `yfinance` package). The original dataset includes the Nasdaq Composite Index, S&P 500 Index, Dow Jones Index Average, Crude Oil Price, and Gold Price. There are 252 data rows from 2021-02-01 to 2022-01-31.

2.1.4 Tweets Dataset for Model Training

The data of Tweets for model training was retrieved from Kaggle. This dataset has three columns: TextID, Text, and Sentiment (neutral, positive, negative). There are 27481 data rows in the dataset.

2.2 Data Preprocessing

The “Stock Combined” dataset was created by utilizing the “Daily Bitcoin (Stock) Price” dataset and “Python Yahoo Finance Package”. The dataset contains stock market data from February 1, 2021, to January 31, 2022. It has a total of 365 rows of data and 9 columns as shown below:

Table 1: Stock Combined Table Variables and Interpretations

Variable	Description
Date	Date of the recorded stock market data
Btc_open	Opening price of Bitcoin on that date.
Btc_close	Closing price of Bitcoin on that date.
Btc_volume	Volume of Bitcoin traded on that date
Nasdaq_close	Closing value of the NASDAQ composite index on that date
Dji_close	Closing value of the Dow Jones Industrial Average on that date
Sp_close	Closing value of the S&P 500 index on the same date
Oil_close	Closing price of oil commodities on that date
Gold_close	Closing price of gold on that date

The “Tweets Grouped” dataset was created by applying numerous preprocessing techniques (e.g., tokenization, stemming) to the “Elon Musk Tweets (2010-2022”) dataset. The dataset contains the tweets of Elon Musk from February 1, 2021, to January 31, 2022. It has a total of 352 rows of data and 2 columns:

Table 2: Tweets Grouped Table Variables and Interpretations

Variable	Description
Date	Date of creation
Tweet	Contents of tweet, tweet body

2.3 Exploratory Data Analysis (EDA)

2.3.1 EDA for Bitcoin and Stock Data

A natural question is: What other factors influence the movement of Bitcoin's closing price? Canonical work on this subject has viewed other stock indices, such as the Dow Jones Industrial Average and the S&P 500, as correlated with the price of Bitcoin. If this is indeed the case, then failing to include these variables as controls could bias our models or lead to inconsistent estimation.

To see why, consider a simple (and plausible) example: the correlation between Bitcoin's price and other stock indices is strictly positive, as is the correlation between Bitcoin's price and the sentiment scores of Elon Musk's crypto-related tweets. In this instance, the prices of other stock indices fulfill the conditions for omitted variable bias, such that OLS coefficients will be biased and inconsistent.

It is also worth measuring the extent of the correlation between Bitcoin and other stock indices. If there is an extremely strong relationship, we might be led to suspect that most variation in Bitcoin prices are explained by factors that also help explain other stock indices (in other words, Bitcoin behaves much like the rest of the stock market). On the other hand, if Bitcoin's returns exhibit a different character, this invites speculation as to other factors that could be uniquely influencing its closing prices.

Indeed, we find from an exploratory data analysis approach that Bitcoin prices and other major stock indices are highly correlated, suggesting that these variables should be included in our model as controls. We also note that the graphs are associated but far from identical. Bitcoin's price is more volatile than the indices (unsurprisingly) but also more volatile than the other commodities (oil, gold) plotted. (This is also apparent from Figure 3.) Bitcoin's price also dropped sharply during a period where none of the other indices were negatively affected. Together, these findings suggest that factors besides those influencing stock or commodities markets contribute to the price of Bitcoin in a substantial way.

We also notice when graphing Elon Musk's twitter sentiment over time (isolating to crypto tweets) that there appears to be a loose correlation between sentiment scores for Musk's Bitcoin-related tweets and Bitcoin closing prices. The sentiment for Musk's crypto tweets appears lower and more volatile during periods where Bitcoin closing prices are falling. It is also worth noting that the sentiment scores for these tweets exhibit substantial variation (though this may also be attributable to measurement error; see *Limitations,* section 4.2).

Figure 1 illustrates moving averages for Bitcoin closing price over the period February 1, 2021, to January 31, 2022. The blue line depicts the moving average of Bitcoin price for the previous 10 periods (days), the orange line depicts the moving average for the previous 20 periods, and the green line depicts the moving average for the previous 50 periods. The repeated pattern of switching between short-term and long-term moving averages, and the substantial gaps between them, indicates relatively high volatility and a high frequency of upturns and downturns.

Figure 1: Comparative Analysis of Bitcoin Closing Price Moving Averages

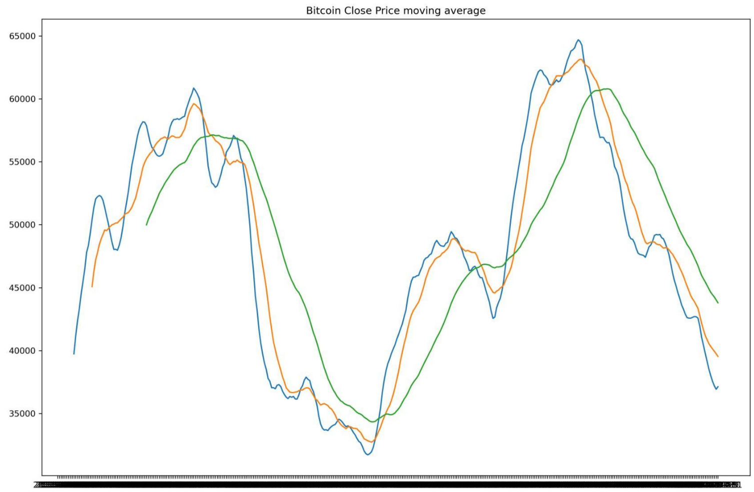


Figure 2 displays the frequency distribution for Bitcoin prices over the period of interest. The distribution appears approximately trimodal, with modes at approximately $35,000, $48,000, and $58,000. It is notable that while the data only covers slightly more than a year, the price distribution stretches from a low of approximately $30,000 to a high above $65,000. Figure 3 plots daily returns over the year, which corroborates the high levels of volatility observed in Bitcoin price. Bitcoin experienced an almost 20% return one day early in the sample period, only to lose nearly 15% of its value just a few months later.

Figure 2: Frequency Distribution Histogram with Kernel Density Estimate for Bitcoin Closing Prices

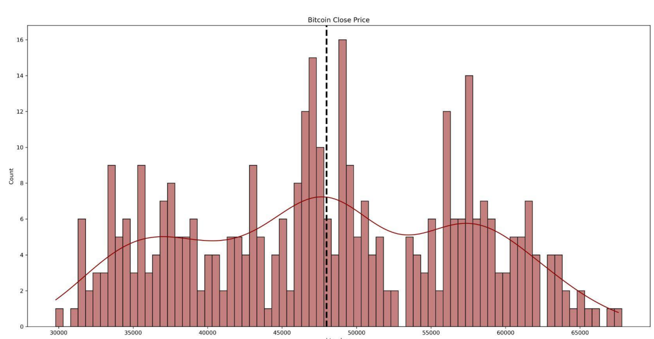


Figure 3: Time Series Plot of Bitcoin Daily Return

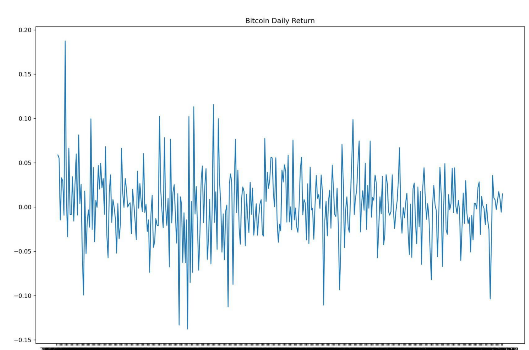


Figure 4 illustrates a trend analysis of the sentiment score (calculated via vaderSentiment) of Elon Musk's tweets. The data sample was restricted to dates on which Elon Musk posted content to Twitter related to Bitcoin or other cryptocurrencies. Clearly, there is substantial variability in this sentiment score, even day-to-day.

Figure 4: Trend Analysis of Sentiment in Elon Musk's Cryptocurrency-Related Tweets Over Time

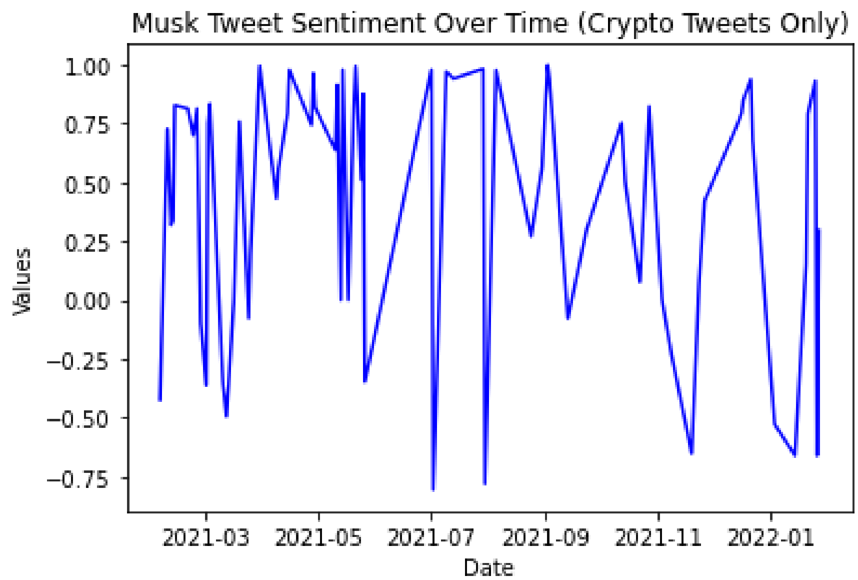
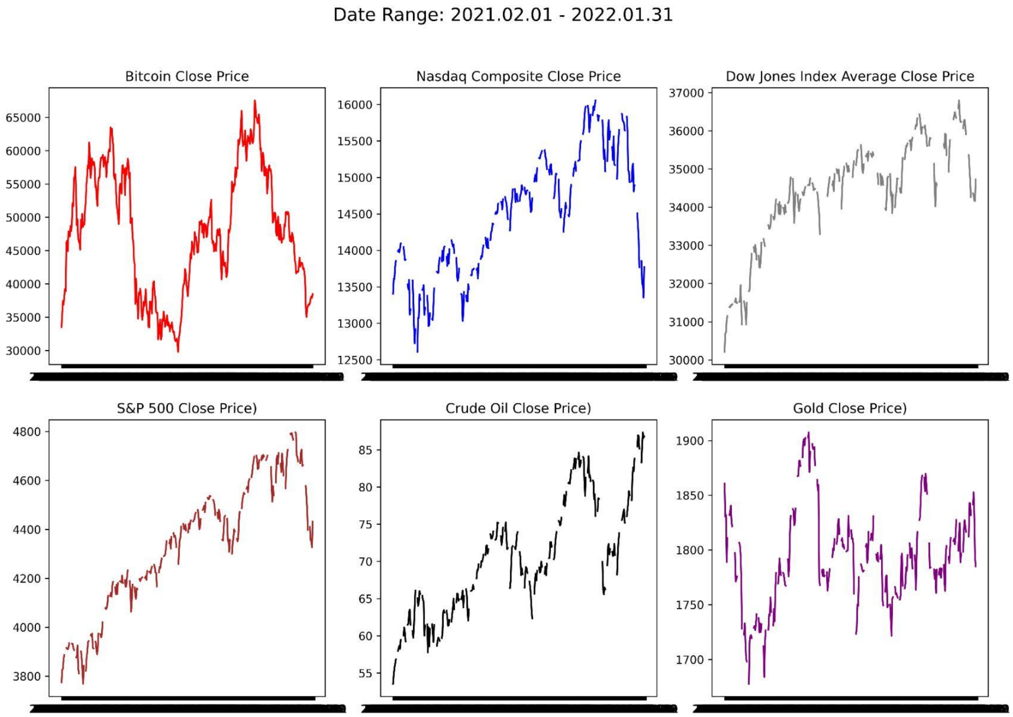


Figure 5 extends this analysis by plotting Bitcoin's closing price together with a variety of other stock indices, including the Nasdaq Composite, the Dow Jones Industrial Average, the S&P 500, and the prices of crude oil and gold. Bitcoin's closing price clearly trends together with larger movements across these indices, but other factors evidently have an influence as well; in other words, movements in major stock indices cannot completely explain movements in the price of Bitcoin. Furthermore, while the volatility in Elon Musk's Twitter sentiment makes coming to conclusive statements challenging at best, it appears that periods during which higher Bitcoin prices are observed are associated with more positive and less volatile sentiment scores.

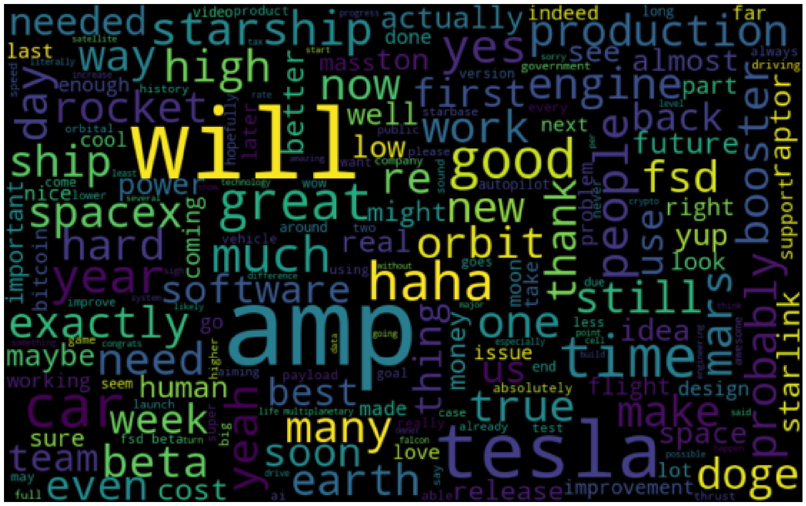
Figure 5: Comparative Closing Price Trends of Bitcoin, Major Stock Indices, and Oil from January 2021 to January 2022



In sum, Bitcoin closing prices are characterized by substantial volatility. This lends evidence to an interpretation that sentiments about Bitcoin are highly susceptible to change and that, to a greater extent than traditional currencies, Bitcoin's valuation is somewhat uncertain. The trimodal closing price distribution in Figure 2 further corroborates this evidence provided by the trading volume data. We also note that Bitcoin price appears related to other stock indices, but the overall picture painted by our analysis implies that other factors contribute to volatility in Bitcoin's price. Thus, other explanatory variables besides those impacting major stock indices may uniquely influence Bitcoin. We also present summary statistics for the control variables and the dependent variable (Bitcoin closing price) in the Appendix (Appendix A.1).

2.3.2 WordCloud for Tweets Data

Figure 6: Word Cloud for Musk's Tweets Data



We also visualized Musk's Twitter posting using a word cloud (Figure 6); larger words correspond to more frequently employed terms. "Bitcoin" and "doge" appear prominently within the cloud, suggesting that Musk indeed takes an interest in cryptocurrency. This suggests that an overall regression of Bitcoin price on Musk's Twitter sentiment may indeed pick up some of the specific influence of Musk's crypto-related tweets, though such inference remains somewhat fraught by the substantial measurement error. Our word cloud analysis also indicates that the overall valence of Elon Musk's Twitter posts skews positive.

3 Methodology

3.1 Sentiment Analysis

In our sentiment analysis, we applied a diverse set of methodologies to evaluate the sentiment of tweets. This includes the use of three types of Naive Bayesian classifiers (Multinomial, Complement, and Bernoulli), each with distinct approaches to handling text data, alongside a Decision Tree and a Random Forest classifier for more complex pattern recognition. Additionally, we employed the VADER polarity score tool to assess sentiment, which is particularly effective in analyzing social media language, like tweets. The best-performing model was then selected for further analysis.

3.1.1 Naive Bayesian Classifier (Tweets: Pos + Neg)

In this study, three Naive Bayesian models were used for sentiment analysis. (1) Multinomial Naive Bayes (MultinomialNB) used the frequency of terms in the dataset to predict classifications. This model assumes that the presence of a particular term in a class is unrelated to the presence of other terms. (2) Complement Naive Bayes (ComplementNB) is a variant of Naive Bayes suitable for imbalanced text classification tasks. It's particularly effective when dealing with unevenly distributed classes in text data. (3) Bernoulli Naive Bayes (BernoulliNB) is based on the application of the Naive Bayes theorem, it considers binary or boolean features to classify text. It works well with features that are binary-valued, where occurrences matter, not frequency. Since the Naive Bayes classifier performs better when the outcome is binary, ‘neutral’ tweets were dropped from model training. After training the model using the training data, we selected the best model for further Bitcoin price prediction.

3.1.2 Decision Tree Classifier (Tweets: Pos + Neg)

Decision Tree classifier is a tree-like graph that uses a branching method to illustrate every possible outcome of a decision. It partitions the data into subsets based on the value of attributes and aims to create branches that best separate different classes.

3.1.3 Random Forest Classifier (Tweets: Pos + Neg)

Random Forest classifier is an ensemble learning method that constructs multiple decision trees during training and merges their outputs to improve accuracy and prevent overfitting. It creates a forest of diverse trees and combines their predictions for more reliable results.

Of the three aforementioned model types, only the one that demonstrated optimal performance was selected for the next phase of our analysis.

3.1.4 VADER Polarity Score (Compound+ Pos+Neu+Neg)

For sentiment analysis on social media text, this project employed the Valence Aware Dictionary and Sentiment Reasoner (VADER) tool on "Tweets Grouped" data to classify Musk's texts in the sentiment analysis, integrated within the NLTK (Natural Language Toolkit) Python package (Hutto, 2021). This tool is adept at interpreting a range of expressions commonly found on social media, including slang, emoticons, and other non-standard language forms (Hutto, n.d.; CodeProject, n.d.). VADER's utility lies in its ability to provide a 'compound' sentiment score by synthesizing the individual positive, negative, and neutral evaluations of text data. This compound metric is scaled from a full negative expression at -1 to a fully positive one at +1, offering a spectrum of sentiment assessment ("Sentiment Analysis in 10 Minutes with Rule-Based VADER and NLTK," n.d.).

Distinct from machine learning algorithms, VADER operates on a rule-based system that utilizes a predefined lexicon of sentiment-laden words, each tagged as either positive, negative, or neutral. The system further refines its analysis by factoring in the influence of word modifiers and contextual nuances to interpret intensifiers and negations accurately (Hutto, n.d.). This obviates the need for preliminary data processing steps such as tokenization, stemming, or vectorization, which are commonly necessary in machine learning frameworks for text data preparation (“Sentiment Analysis in 10 Minutes with Rule-Based VADER and NLTK,” n.d.). Thus, the raw text is directly input into the VADER tool, which then generates sentiment scores based on its intrinsic lexicon and rule set.

To prepare the tweets for analysis using VADER, they were first converted to string format using the `.astype(str)` method. Then, each tweet in the “Tweets Grouped” data frame was processed by the `SentimentIntensityAnalyzer.polarity_scores` polarity scores function from the nltk.sentiment.vader module. This function computes sentiment scores for the text input and generates a dictionary containing scores for different categories, such as 'compound,' 'pos' (positive), 'neg' (negative), and 'neu' (neutral). Next, individual sentiment scores from the dictionaries were extracted and stored in separate columns in the “Tweets Grouped” data frame. Subsequently, a new dataset named 'sentiment_scores_comb' was created to store the sentiment scores only, and the indexes were reset. The 'date' column in the 'Tweets Grouped' data frame was further matched with the 'date' column in the 'Stock Combined' dataset, and the compound score was concatenated for prediction modeling.

3.2 Predictive Modeling

To dive deep into the relationship between Musk’s tweets and bitcoin prices, we continued on a predictive modeling study. The primary objective was to forecast the closing price of bitcoin (Y) using a set of predictors (X) that included bitcoin volume, three stock market indices, and the closing prices of oil and gold. Our methodology incorporated several advanced statistical and machine learning techniques, including Linear Regression, Random Forest, and XGBoost models. The performance of each model was evaluated using key metrics such as RMSE and R-squared for further comparisons.

3.2.1 Data Partition

The dataset is divided into two subsets: the training set and the testing set. The training set includes 80% of the data, used for building and training the model. The testing set comprises the remaining 20%, used for evaluating the model's performance. The seed of the random number generator used in splitting the dataset was fixed at 0. It helps ensure that each execution of the code yields an identical split of the dataset, thus providing uniform results for every run and maintaining consistency for different users.

3.2.2 Linear Regression

The documentation for the VaderSentiment package suggests that for most applications, simply using the "compound" score is sufficient (Hutto, 2021). From the package's readme file, the compound score is described as a "normalized, weighted composite score," making it ideal for our application: classifying overall tweet sentiment on a continuous scale.

We specify the following linear regression model:

$$btcp_t = sent_t \beta_1 + x_t' \gamma + e_t$$

where $btcp_t$ denotes Bitcoin closing price at date t, $sent_t$ is the compound score calculated via VaderSentiment for Elon Musk's tweets on date *t*, x_t' is a vector of controls and a constant observed at date *t,* and where e_t denotes the error term. The controls included are Bitcoin volume, the closing prices of the NASDAQ composite, DJI, and S&P 500 stock indices, and the closing prices for oil and gold.

We also performed a Lasso estimation procedure in an effort to determine whether our model selection could be improved, but the resulting coefficients remained statistically insignificant (Appendix A.2).

3.2.3 Random Forest

Random forest is also a powerful tool for regression tasks, leveraging the strengths of ensemble learning to predict continuous outcomes. The Random Forest approach is an ensemble learning technique that constructs multiple decision trees during training and outputs the average prediction of the individual trees for regression tasks. It makes it particularly useful when dealing with complex datasets where the relationships between variables are not well understood or are highly non-linear, and helps to avoid overfitting.

The model was fine-tuned using scikit-learn's GridSearchCV method, which performs an exhaustive search over specified parameter values. A parameter grid was defined for hyperparameter tuning, consisting of two key parameters. One is the number of trees in the forest, which is set to vary between 100, 500, and 1000, which is critical in determining the model complexity and capturing data patterns. The other is the maximum depth of each tree, being set to either unrestricted or level 5 to prevent overfitting. Negative mean squared error was used as the criterion for node splitting, and the minimum number of samples required in a leaf node was set to 10. The random forest model employed a 5-fold cross-validation approach for robust model assessment, allowing for the selection of the best parameters based on overall performance.

3.2.4 XGBoost

In our study, we adopted the XGBoost (eXtreme Gradient Boosting) method for constructing supervised regression models. XGBoost is chosen for its advanced capabilities in handling complex regression tasks. Its gradient-boosting framework is particularly adept at capturing intricate patterns in data, making it suitable for our predictive modeling needs. Specifically, the XGBRegressor function was used to construct a model with 100 sequential gradient-boosted trees, refining the errors made by its predecessor. The random number generator seed was fixed at 0 to ensure reproducibility of results. The model was fitted to the training data with these settings to optimize its learning process (Appendix A.3). Then, the prediction function was used to generate forecasts on the test data, followed by calculation of Root Mean Squared Error (RMSE) and R-squared values to evaluate the model's accuracy and predictive power.

4 Results

4.1 Sentiment Analysis

The results of the overall sentiment analysis, displaying the percentages of positive and negative labels, are presented in the following diagram (Figure 7). The models (Table 3), Multinomial, Complement, and Bernoulli Naive Bayes, demonstrated similar and high accuracy levels, scoring above 0.86. Their performance across precision, recall, and weighted average F-1 score was consistent, hovering around 0.87. Meanwhile, the

Decision Tree model achieved a lower accuracy of 0.8066, with its F-1 score, precision, and recall also around 0.81. The Random Forest model closely matched the Naive Bayes methods, scoring 0.86 across these metrics.

Figure 7: Comparative Pie Charts of Positive vs. Negative Sentiment Analysis Using Naive Bayes, Random Forest, and Decision Tree Models

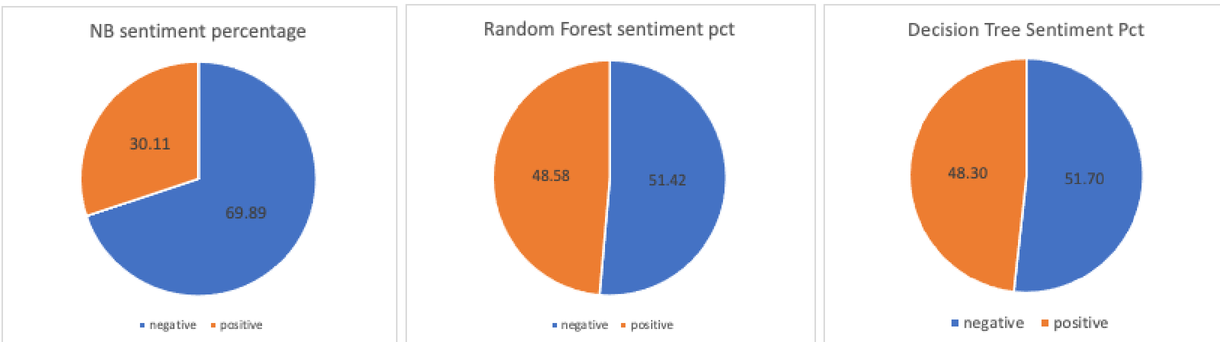


Table 3: Performance Metrics Comparison of Different Models

Measures	MultinomialNB	ComplementNB	BernoulliNB	DecisionTree	RandomForest
Accuracy	0.865	0.8671	0.8665	0.8066	0.8603
Weighted Avg (F-1)	0.87	0.87	0.87	0.81	0.86
Weighted Avg (precision)	0.87	0.87	0.87	0.81	0.86
Weighted Avg (recall)	0.86	0.87	0.87	0.81	0.86

Overall, the Naive Bayes classifiers stood out for their robust and reliable performance, outperforming the Decision Tree in most evaluated aspects (Table 4). Complement NB's better performance can be explained by its superiority in predicting unbalanced data. We will use Naive Bayes sentiment results as a feature in Bitcoin price modeling.

Table 4: Specificity and Sensitivity Comparison of Naive Bayes vs. Random Forest and Decision Tree Hybrid Models

Measure	NB x Random Forest	NB x Decision Tree	Random Forest x Decision Tree
Specificity	0.64	0.56	0.72
Sensitivity	0.77	0.69	0.77

To add on, we also briefly evaluated the association of Elon Musk's tweets on Bitcoin prices using two models (Appendix A.4): one considering all of Musk's tweets and another focusing solely on his cryptocurrency-related tweets. Our findings revealed that in both cases, the direct association between Elon Musk's tweets and Bitcoin prices is not statistically significant, with p-values exceeding the 0.05 threshold.

4.2 Predictive Models

To increase the depth of our analysis, we compared model results across three different datasets: Baseline, With Vader, and With NB. The only difference between the three versions was the selection of the independent variable. For Baseline version, we only chose the macro signals as the independent variables, while for With Vader, we chose both the macro signals and the Vader sentiment score (compound score) as the independent variables, and for With NB, we chose both the macro signals and the NB sentiment classifier results as the independent variable. This multi-dataset approach allowed us to assess the impact of different types of sentiment analysis on the accuracy of bitcoin price predictions.

Table 5: Comparison of RMSE and R^2 for Linear Modeling, Random Forest, and XGBoost

Linear Modeling				Random Forest				XGBoost			
Scores	**Baseline**	**With Vader**	**With NB**	**Scores**	**Baseline**	**With Vader**	**With NB**	**Scores**	**Baseline**	**With Vader**	**With NB**
RMSE	7830	7852	7864	**RMSE**	2967	2972	2954	**RMSE**	3282	3474	3185
R2	0.31	0.30	0.30	**R2**	0.90	0.90	0.90	**R2**	0.87	0.86	0.89

The performance of each model was evaluated using key metrics such as RMSE and R-squared (Table 5), allowing for a comprehensive and comparative analysis of their predictive capabilities in relation to bitcoin's closing price:

1. **RMSE** (Root Mean Squared Error): typical distance between the predicted values and the actual values. Lower RMSE values imply better model performance. From the table, we can see that Random Forest has the lowest RMSE overall.

2. **R-squared**: the proportion of variance in the dependent variable that is explained by the independent variables. Higher R-square values imply better model performance. From the table, we can see that Random Forest has the highest R-squared overall.

As a result (Table 5), we can see that Random Forest with Naive Bayes classifier performs the best among all versions of models. Meanwhile, comparing the With Vader and With NB with the Baseline version, we can conclude that adding an independent variable related to Musk's tweets does not have a significant effect on the model performance.

5 Limitations

To begin with, this study only focused on the daily impact of Musk's tweets and did not encompass the impacts occurring within a single day. Prior literature has shown that Musk's tweets influence the price and volume of cryptocurrency within hours. Further investigation could delve into exploring the influence of Musk's tweets on the price and volume of Bitcoin within a single day.

Secondly, according to the research done by Ante in 2021, instead of using the exact value of price the expected return is calculated over an estimation period before an unexpected event and is compared to the observed return around the event. Our study focused on the exact value and volume instead. Further investigation could delve into using the expected return as the dependent variable.

Thirdly, we did not separate tweets on the same day; rather, all tweets for the same date were concatenated. In other words, our unit of observation is dates and not tweets. This could be an issue for our study — specifically, for the regression specification in which non-crypto-tweet dates are excluded — because it implies that rather than computing the sentiment score for crypto-related tweets specifically, we are computing the overall score for all tweets on dates where crypto-related content was posted. This could lead to inconsistent estimation because our regressor of interest (Elon Musk's crypto-related tweets) was observed with measurement error. Future work could observe Musk's tweets over a longer time horizon to collect a larger sample for the statistical power necessary to test a hypothesis about the impact of Musk's crypto-related tweets on Bitcoin closing prices.

In the end, we imputed stock values for weekends to fill in missing data, which could have introduced errors.

6 Conclusions

On an overall note, the most optimal sentiment classifier is Complement Naive Bayes, and the most optimal model for predicting Bitcoin's closing prices is Random Forest Regression with the Naive Bayes classification of Elon Musk's tweets as a feature. With this model, we can achieve an R-square of 90%. However, we also found that Elon Musk's tweet sentiment does not significantly improve the performance of any model in predicting Bitcoin's closing prices, regardless of the sentiment classifier employed. The baseline model without tweet sentiment data appears to be a more significant determinant in forecasting Bitcoin price. As mentioned in the limitations part earlier, the lack of significant findings may be attributable to our focus solely on the daily influence of Musk's tweets. It's possible that the sentiments expressed in Musk's tweets have a short-term effect on Bitcoin prices that our daily analysis does not capture. More studies could be conducted to explore the short-term “Musk effect” on Bitcoin price in the future.

7 References

[1] Pavel Ciaian, Miroslava Rajcaniova & d’Artis Kancs (2016) The economics of BitCoin price formation, Applied Economics, 48:19, 1799-1815, DOI: 10.1080/00036846.2015.1109038

[2] Georgoula, Ifigeneia and Pournarakis, Demitrios and Bilanakos, Christos and Sotiropoulos, Dionisios and Sotiropoulos, Dionisios and Giaglis, George M., Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices (May 17, 2015). SSRN: http://dx.doi.org/10.2139/ssrn.2607167

[3] Meiryani, M., Tandyopranoto, C. D., Emanuel, J., Lindawati, A. S. L., Fahlevi, M., Aljuaid, M., & Hasan, F. (2022). The effect of global price movements on the energy sector commodity on bitcoin price movement during the COVID-19 pandemic. Heliyon, 8(10).

[4] Hakan Yilmazkuday (2023) COVID-19 effects on the S&P 500 index, Applied Economics Letters, 30:1, 7-13, DOI: 10.1080/13504851.2021.1971607

[5] Hsu P P, Chen Y H, Wang C H. Currency or commodity competition? Bitcoin price trends in the post-pandemic era[J]. International Journal of Applied Economics, Finance and Accounting, 2023, 15(2): 61-70.

[6] J. Luo, Q. Ji, High-frequency volatility connectedness between the US crude oil market and China’s agricultural commodity markets Energy Econ., 76 (2018), pp. 424-438

[7] Hutto, C. J. (2021). VADER (Valence Aware Dictionary and Sentiment Reasoner). Read the Docs. Retrieved from https://vadersentiment.readthedocs.io/en/latest

[8] Hutto, C. J. (n.d.). Welcome to VaderSentiment's documentation! VADER Sentiment. Retrieved from https://vadersentiment.readthedocs.io/en/latest.

[9] Muthukrishnan, R., Rohini, R (2016). LASSO: A feature selection technique in predictive modeling for machine learning. IEEE.

[10] Tibshirani, Robert (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 58, No. 1.

[11] CodeProject. (n.d.). Using Pre-trained VADER Models for NLTK Sentiment Analysis. Retrieved from https://www.codeproject.com/Articles/5269445/Using-Pre-trained-VADER-Models-for-NLTK-Sentiment

[12] Towards Data Science. (n.d.). Sentiment Analysis in 10 Minutes with Rule-Based VADER and NLTK. Retrieved from https://towardsdatascience.com/sentiment-analysis-in-10-minutes-with-rule-based-vader-and-nltk-72067970fb71

8 Appendix

A.1 Descriptive Statistics: Covariates and Bitcoin Stock Price (Table 6)

Table: Descriptive Statistics for Covariates and Bitcoin Stock price

	btc_open	btc_close	btc_volume	nasdaq_close	dji_close	sp_close	oil_close	gold_close
count	365	365	365	252	239	252	252	252
mean	47972.44	47976.48	4.3787E+10	14474.21	34382.55	4332.43	70.41	1794.09
50%	47810.69	47783.36	3.8056E+10	14535.06	34584.9	4359.74	70.02	1792.25
max	67549.73	67566.83	3.5097E+11	16057.44	36799.7	4796.56	87.35	1907.5
min	29796.29	29807.35	1.4644E+10	12609.16	30211.9	3768.47	53.55	1677.7

A.2 Lasso

While results from a Lasso specification were insignificant, we did test whether an alternative choice of loss function produces significant coefficients. Lasso estimation is useful in cases where overfitting may be present (Tibshirani, 1996). In other words, given a large number of predictors (regressors), especially relative to the number of observations, the variance of the OLS estimates increases. Additionally, because Lasso estimation penalizes the magnitude of the estimated coefficients, it is sometimes used as a feature-selection technique, since some estimates may be set to zero (Muthukrishnan, 2016).

In the context of least squares regression models, Lasso estimation minimizes the following loss function:

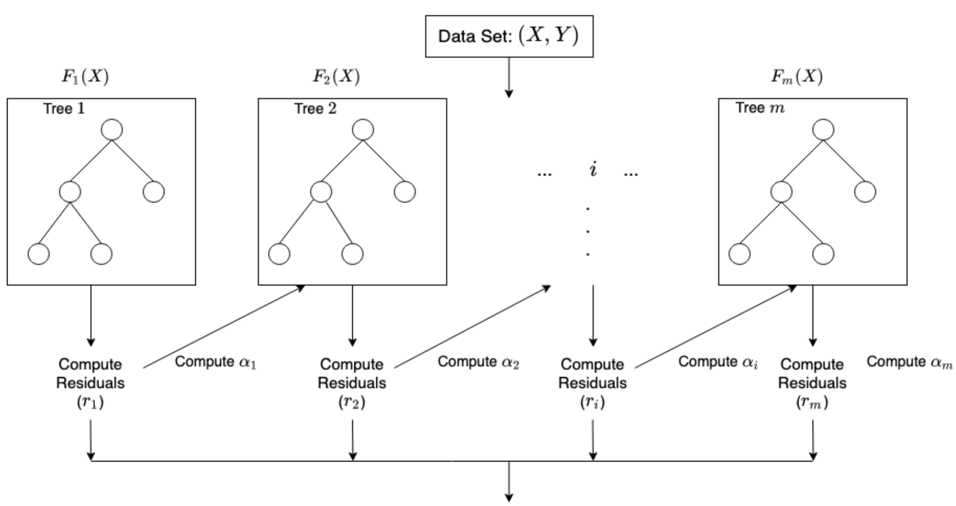
$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

λ is a hyperparameter, traditionally estimated through *k*-fold cross validation. For the purposes of our estimation, we set *k* = 5 and iteratively estimated the model for the following values of λ: 0, 0.0001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30. In each estimation, the hyperparameter value that best minimized the pseudo-out-of-sample prediction error was chosen.

Using all the dates for our data, the optimal λ was calculated as 0, which suggests that we were not overfitting the model in our original OLS specification. The Lasso and OLS estimations were thus identical.

Using only dates on which Elon Musk posted crypto-related content, we found that the optimal λ was 20, which is consistent with our smaller sample size (n = 235 in the overall sample but n = 70 in the restricted sample). This suggests we may have been overfitting the model previously in our original OLS specification. The estimated coefficient $\hat{\beta}_1$ was almost identical at 169.38 and remained statistically insignificant, with a p-value of 0.89 > 0.05.

A.3 XGBoost Graphical Interpretation



$F_m(X) = F_{m-1}(X) + \alpha_m h_m(X, r_{m-1})$,
where α_i, and r_i are the regularization parameters and residuals computed with the i^{th} tree respectfully, and h_i is a function that is trained to predict residuals, r_i using X for the i^{th} tree. To compute α_i we use the residuals computed, r_i and compute the following: $arg \min_{\alpha} = \sum_{i=1}^{m} L(Y_i, F_{i-1}(X_i) + \alpha h_i(X_i, r_{i-1}))$ where $L(Y, F(X))$ is a differentiable loss function.

A.4 Simple Regression

We first estimated the model using all dates in the timeframe of interest on which Elon Musk posted on Twitter, not just dates on which Musk's tweets contained content related to cryptocurrencies. We found that an increase in the sentiment score by 1 was associated with a $379.70 increase in Bitcoin prices. However, this coefficient is far from significant: the p-value for the hypothesis test of $H_0\colon \beta_1 = 0$ is 0.856 > 0.05, so we fail to reject the null hypothesis that the sentiment scores of Musk's tweets do not influence Bitcoin closing prices.

Figure: Scatterplot of Bitcoin Closing Prices Against Scaled Sentiment from on Musk's Tweets with Ordinary Least Squares (OLS) Regression Line

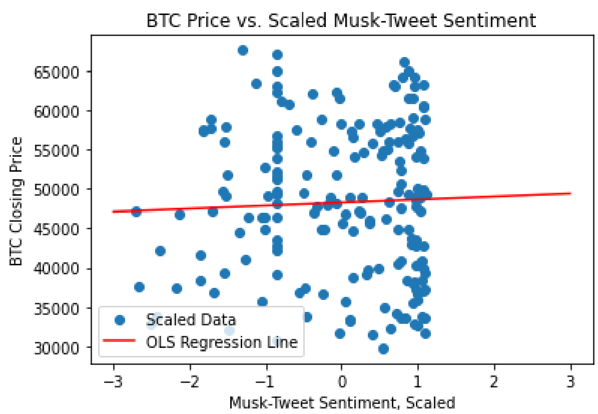
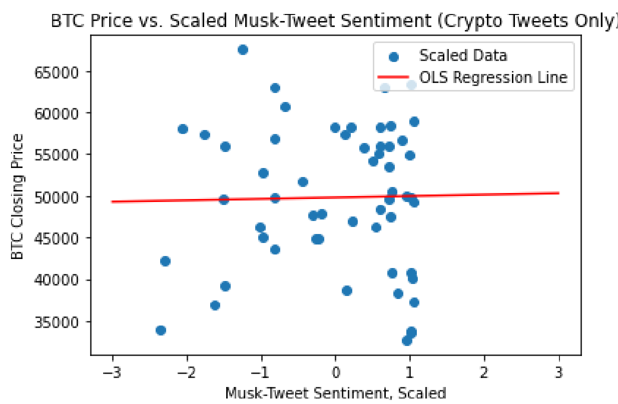


Figure: Scatterplot of Bitcoin Closing Prices Versus Scaled Sentiment of Musk's Cryptocurrency Tweets with OLS Regression Line



Next, we estimated the model using only dates on which Elon Musk's tweets contained cryptocurrency-related material. We found that an increase in the sentiment score by 1 was associated with a (smaller) $166.22 increase in Bitcoin prices. Again, this coefficient is not statistically significant with a p-value of 0.892 > 0.05.

A.5 Github Link

Please refer to our GitHub page for more information:
https://github.com/zc2691/QMSS5067_final