NewsBit

Investment Decision Support Tool Built Using Machine Learning Models

Capstone Project Team
Karan Raja
Kaidi Wu
Linhan Cai
Yunhua Su
Zhiwei Gu

Outline:

I. Abstract
II. Introduction
    A. Research Background
    B. Hypothesis Development
    C. Thesis Statement
III. Data
    A. Data Collection and Preprocessing
    B. Data Analysis
IV. Methodology and Technical Project Outcome
    A. Linear Model
    B. Logistic Model
    C. Tree Based Model
    D. Boosting
    E. Random Forest
    F. KNN
    G. Time Series
    H. Neural Network
    I. Result Analysis
V. Discussion and Conclusion
VI. Citation
VII. Appendix

Abstract

In this paper, we are trying to predict Bitcoin prices by using a mix of machine learning algorithms such as time series analysis, sentiment analysis, technical analysis, and neural networks. Our goal is to create a valid way that can find patterns in the ever-changing crypto market. Unlike previous research, we focus on using natural language processing to analyze tweets, Reddit posts, and online blogs to understand how people feel about the market and how it can affect prices.

We implemented different methods such as time series analysis, technical indicators, and neural networks, all trained on historical data. The models aims to predict what might happen in the market in the future. By combining these models, we offer a powerful tool for anyone who wants to get an edge in the Bitcoin market.

This paper talks about how we built these models, the challenges we faced, and what we learned from the results. We eventually finish the paper with a discussion on what our findings mean, any limitations in our approach, and ideas for future research in cryptocurrency price prediction.

Introduction

Research Background

In recent years, the cryptocurrency market has experienced rapid growth, presenting an increasing challenge in predicting price movements for retail investors. Unlike traditional assets, there is no means to measure the true value of Bitcoin. Hence, we want to identify the factors that most influence the price of Bitcoin. To do this, we will first evaluate prior work to identify the most effective price prediction strategy. McNally et al. (2018) presented a comprehensive study on the prediction of Bitcoin prices using Long Short-Term Memory (LSTM) networks. Their research focused on utilizing a range of technical indicators and demonstrated that LSTM networks effectively predicted Bitcoin prices with an accuracy of 60.12%. This work laid a strong foundation for future studies employing LSTMs for cryptocurrency price prediction.

In another study, Stanković et al. (2020) developed a hybrid ARIMA-GARCH model for predicting the volatility of the Bitcoin market. By combining ARIMA and GARCH models, the authors aimed to accurately forecast Bitcoin price returns and volatility. Their findings indicated that the hybrid model provided more accurate predictions compared to individual models. This work highlights the potential of hybrid models for capturing the complex dynamics of cryptocurrency markets.

Lastly, Jaquart et al. (2021) analyzed the short-term predictability of the Bitcoin market using various machine learning models, such as recurrent neural networks and gradient boosting

classifiers. The authors employed a comprehensive feature set, including technical, blockchain-based, sentiment/interest-based, and asset-based features. The study found that technical features were most relevant, and predictability increased for longer prediction horizons.

In summary, these studies underscore the value of machine learning models, such as LSTMs, hybrid models, and gradient boosting classifiers, for predicting cryptocurrency markets. They also emphasize the importance of utilizing diverse feature sets to improve prediction accuracy. Collectively, these works provide valuable insights into the challenges and opportunities associated with predicting cryptocurrency markets and pave the way for future research in this domain.

Hypothesis Development

Based on our research background, the following hypotheses are developed to guide the research:

Hypothesis 1 (H1): Socio-economic indicators have a significant influence on Bitcoin price movements. Given the volatility and susceptibility of Bitcoin to market forces, it is hypothesized that socio-economic factors, such as inflation rates, GDP growth, and market capitalization, have an impact on Bitcoin prices.

Hypothesis 2 (H2): Public sentiment, as expressed through social media platforms like Twitter and Reddit, influences the price of Bitcoin. The prevailing sentiment among retail investors can impact the demand and price of Bitcoin. Therefore, it is hypothesized that positive

sentiment will correspond with price increases, while negative sentiment will correlate with price decreases.

Hypothesis 3 (H3): Technical indicators derived from Bitcoin's historical price data, including moving averages, relative strength index (RSI), and Fibonacci retracement levels, can predict future price movements. These indicators are widely used in financial markets to predict price trends and reversals, suggesting their potential applicability to Bitcoin.

Hypothesis 4 (H4): A combination of socio-economic indicators, public sentiment, and technical analysis metrics will provide a more accurate predictive model for Bitcoin prices than any single type of indicator. Given the complex interplay of factors influencing Bitcoin prices, it is hypothesized that a model incorporating a diverse range of indicators will be most effective.

These hypotheses form the basis of our exploration into the factors influencing Bitcoin prices, guiding the data collection, analysis, and model development processes. The outcomes will either support or refute these hypotheses, offering valuable insights into the dynamics of Bitcoin prices.

Thesis Statement

The advent of cryptocurrencies, such as Bitcoin, has generated considerable interest among investors and academics alike. However, unlike traditional assets, there is no established framework for measuring the true value of Bitcoin, which presents a significant challenge in predicting its price movements. To address this challenge, this paper aims to identify the factors

that most influence the price of Bitcoin by using socio-economic indicators and applying several machine learning models.

This study is motivated by the growing interest in cryptocurrencies and the potential impact of retail traders on the market. It is noteworthy that there are over 50 million retail traders that make up 17% of crypto trading volume. However, investing in cryptocurrencies carries significant risks, as evidenced by the fact that over 80% of new crypto investors inevitably lost money on their initial investment.

To develop an effective prediction framework for Bitcoin prices, this study proposes the use of statistical learning models that encompass a mix of financial, economic, and social indicators. Such models can provide a more comprehensive understanding of the factors that influence Bitcoin prices and help investors make informed decisions. Overall, this study seeks to contribute to the emerging field of cryptocurrency research by offering new insights into the price dynamics of Bitcoin and by highlighting the importance of sophisticated approaches to predicting its price movements.

Since no framework exists to determine the price of Bitcoin, there is significant scope to use statistical learning models to develop a framework which encompasses a mix of financial, economic, and social indicators

Data

Data Collection and Preprocessing

In order to utilize machine learning models to predict bitcoin prices, data collection is the most crucial element of this process. By selecting the right data to collect and making sure the dataset is unbiased, our final model could have a much better performance than others. The first element that sets us apart from the current research is the fact that we focus heavily on the sentimental analysis of the general public. We collected data from non-traditional media platforms in the effort to search for better correlation and representation. To collect past tweets, we utilized the Selenium library in Python, which allowed us to automate the web scraping process. We targeted specific user accounts and hashtags associated with Bitcoin and navigated through the Twitter pages programmatically. In order to filter out the noise and irrelevant content, we've set a minimum bar for the number of likes or comments to be included in our dataset. By extracting the text content of the tweets, we created a dataset that captured public sentiment and opinions about Bitcoin. Preprocessing steps also included removing special characters, URLs, stop words, and converting the text to lowercase, allowing for the cleaned and standardized data to be used for nature language analysis effectively.

For Reddit post collection, we used the Reddit API (PRAW) to access the platform's data. We targeted popular cryptocurrency subreddits, such as r/Bitcoin and r/CryptoCurrency, and extracted textual content and metadata (upvotes, comments, and post creation dates). The Reddit data underwent similar preprocessing as the tweet data to ensure consistency and facilitate subsequent analysis.

Economic indicators, including GDP growth, inflation rates, and market capitalization, were gathered from multiple sources, such as the World Bank, FRED (Federal Reserve Economic Data), and trading platforms. These indicators provided a macroeconomic context for our model. The data was cleaned and processed by aggregating values for specific periods and normalizing them to maintain comparability across different indicators.

For the technical analysis, we collected Bitcoin historical price data using financial data APIs like Alpha Vantage and FRED. From this data, we computed key metrics such as moving averages, relative strength index (RSI), and Fibonacci retracement levels. The moving averages were calculated for different time intervals (20 and 100 days) to identify trends and potential support or resistance levels. The RSI was computed as an oscillator, providing insights into overbought or oversold conditions in the market, while Fibonacci retracement levels were derived from the Fibonacci sequence, indicating potential price reversal points during market corrections.

After preprocessing each dataset, we merged the cleaned data from tweets, Reddit posts, economic indicators, and technical analysis metrics. This comprehensive dataset was then used for developing our Bitcoin price prediction model, capturing the complex relationships between various factors and their influence on Bitcoin prices.

Data Analysis

In our groundbreaking project, we harnessed the capabilities of an array of natural language processing (NLP) tools, including TextBlob, nltk, and SentimentIntensityAnalyzer, to conduct sentiment analysis on a wide-ranging collection of data sources like Tweets, Reddit

posts, and news articles. The primary purpose of this analysis was to glean valuable insights into public opinion and emotions about Bitcoin, ultimately serving as input features for our predictive model.

To achieve this, we embarked on a journey to collect data from various sources. Employing APIs and web scraping methodologies, we amassed pertinent text data, ensuring a diverse and representative sample. Once we had gathered the data, we delved into pre-processing the text by tokenizing, stemming, and lemmatizing the words. This process involved the removal of stop words, punctuation, and irrelevant characters, as well as converting all words to lowercase for consistency.

Subsequently, we turned to the TextBlob package to determine the sentiment polarity and subjectivity for each snippet of text. Polarity is a metric that ranges from -1 (negative sentiment) to 1 (positive sentiment), while subjectivity fluctuates between 0 (objective) and 1 (subjective). This technique provided us with a quantitative assessment of public sentiment, which we then aggregated and normalized for each data source to compensate for potential biases and variations in the gathered data.

In addition to TextBlob, we employed the SentimentIntensityAnalyzer class from the nltk.sentiment package to evaluate sentiment scores. This method utilizes a pre-trained model and a lexicon-based approach to ascertain the compound, positive, negative, and neutral sentiment scores for each text entry. By integrating the results from TextBlob and

SentimentIntensityAnalyzer, we successfully constructed a robust and comprehensive sentiment analysis feature set for our model.

To further refine our sentiment analysis procedure, we explored various NLP techniques, such as n-grams and part-of-speech tagging, to extract additional context from the text data. This approach enabled us to pinpoint more intricate relationships between words and gain a deeper understanding of the inherent sentiment within the text.

Lastly, we performed an in-depth exploratory data analysis (EDA) on the sentiment features derived from the text data. This process entailed visualizing the distribution of sentiment scores, identifying trends and patterns, and examining the relationships between sentiment and other variables like time and historical Bitcoin price data. Through this EDA, we obtained valuable insights into the influence of public sentiment on Bitcoin price fluctuations and verified the efficacy of our sentiment analysis process.

By skillfully employing a blend of NLP packages and techniques, we managed to carry out extensive sentiment analysis on a diverse range of data sources. This ultimately led to the creation of a valuable feature set for our predictive model, aimed at forecasting future Bitcoin price fluctuations with heightened precision.

Methodology and Technical Project Outcome

Linear Model

$$-0.0005 + 8.55x1 + 6.88x2 + 99.70x3 - 11.21x4 + 2582.60\,x5 - 1122.40x6 + 0.0002x7 + 0.0003x8 - 961.92x9$$

The given equation is a mathematical representation of the linear model. A regression model is used to predict a continuous value. The above linear regression can be used to predict the % change in price one week from today, based on the value of variables today. The model test-set predictions with 99.3% accuracy and a mean error of 32%. The model includes nine predictor variables: x1 to x9. Several of these variables show significant correlation with BTC price.

Firstly, there is a positive correlation with the one-week historic BTC price **(x1)** and interday movement **(x2)**. This suggests that historic prices can be informative in predicting future movements. Additionally, the model shows a positive correlation with the price of SPY **(x3)**, indicating that traditional investing has an influence on BTC price.

Furthermore, the price of crude oil **(x4)** shows a positive correlation with BTC prices, suggesting a link between BTC prices and oil supply. Conversely, there is a negative correlation with the price of gold **(x5)**, which implies that BTC prices may be linked with inflation.

The model also shows a positive correlation with the price of copper **(x6)**, suggesting a possible link between BTC price and industrial output. In contrast, there is a negative correlation with the price of corn **(x7),** indicating that farm-payroll and activities may influence BTC price. The SGDUSD rate **(x8)** shows a positive correlation with BTC price, suggesting that BTC price is linked with the movement of the USD. Finally, the model indicates a positive correlation with the Twitter sentiment **(x9)**, suggesting that BTC price is linked with social media trends and sentiment.

In conclusion, the given linear model provides insight into the factors that influence BTC price movements. The positive and negative correlations with various predictor variables can help investors and analysts make informed decisions about BTC investments.

Logistic Model

In addition to linear regression and the CART model, we have also incorporated logistic regression into our Bitcoin price prediction model. Logistic regression is a type of generalized linear model (GLM) used for predicting the probability of an event occurring based on one or more independent variables.

Two classification problems were assessed with the logistic regression model. The first one is classifying whether one day's closing price would be higher or lower using that day's opening price, financial indicators, and sentiment values. The second model aimed to classify the next day's opening price with the same parameters with the addition of the average price and

trading volume of the previous day. Both models use the same time-based train-test split as linear regression.

The first model, however, did not work as expected. After investigating the P-value of each parameter and ignoring the high-significance ones, the resultant model is as shown:

1.6256*Intercept - 0.002*Bitcoin_Price_Previous_Day_Open + 0.0225*Litecoin_Price + 0.0021*SP500 -132.75*UST_1Y_Maturity + 163.9764*UST_10Y_Maturity - 8.59*Euro - 0.0231*All_Twitter_Posts + 0.032*Positive_Tweets + 0.0244*Neutral_Tweets

The Model's accuracy on the test set is 55.9% and its MSE is 0.441. After plotting the ROC graph, the performance of the model does not show a significant improvement over the baseline model. Therefore, we shifted the focus to classifying trends for the next day.

The second model had 5 parameters that show high significance. The resultant accuracy and MSE were 62.7% and 0.373, respectively. After investigating the ROC plot, we found a significant improvement over the baseline model.

Tree Based Model

Alongside linear regression, another valuable technique employed in our Bitcoin price prediction model is the use of tree based models. This includes Random Forests, CART and Gradient Boosting. K-fold cross validation was used on the random forest and gbr models.

Gradient boosting regression had the best performance among all model types. OSR2 before cross validation was 0.181 for classification and 0.219 for regression. The best gradient boosting classification model had a test-set OSR2 of 0.671 and an MAE of 0.132. Cross validation improves the test-set performance by 370% . For regression, the best random forest model had a test-set OSR2 of 0.637 and an MAE of 0.066. Cross validation improves the test-set performance by 290%

This performance was achieved a s result of adopting K-Fold cross validation. For classification, the model had a test-set OSR2 of 0.471 with an MAE of 0.132. The OSR2 for the same model before cross validation was 0.219. This demonstrates the importance of cross validation across all models.

Unlike with boosting, cross validation had a marginal impact on the performance for Random Forests. OSR2 before cross validation was 0.365 for classification and 0.456 for regression. The best random forest classification model had a test-set OSR2 of 0.376 and an MAE of 0.346. Cross validation improves the test-set performance by 3% . For regression, the best random forest model had a test-set OSR2 of 0.491 and an MAE of 0.051. Cross validation improves the test-set performance by 7%.

CART Models had the least impressive performance across all tree based models. For classification the decision tree had a test-set OSR2 of 0.151 and an MAE of 0.209. The decision tree regressor had a test-set OSR2 of 0.208 and an MAE of 0.324. In addition to this, the decision tree classifier had a predication accuracy of 69.9% - a 50% improvement over the baseline model. The classification tree had a true positive rate of 0.676 and a false positive rate of 0.281. The regression tree has an accuracy of 71.2% - a 38% improvement over the baseline model. The regression tree has a true positive rate of 0.732 and a false positive rate of 0.214.

The models highlighted the importance of media sentiment. Twitter sentiment was identified as an important variable based on the GBR's variable importance matrix. While media sentiment was identified as a crucial variable, it alone may not be sufficient in providing accurate predictions. The inclusion of other relevant features can improve model performance and increase accuracy.

Although GBR and RF have interpretability issues, they are still recommended over decision tree classifiers due to their superior performance. These models offer better accuracy, better prediction capabilities, and the ability to handle complex data. However, it is important to note that the interpretability issues associated with these models can make it difficult to understand how predictions are made, which is an essential factor for certain applications.

In conclusion, while media sentiment is an essential variable in classification modeling, the inclusion of additional relevant features is vital for building robust models. Despite

interpretability issues, GBR and RF are still recommended over decision tree classifiers for their superior performance in handling complex data and providing more accurate predictions.

KNN

In this section, we discuss the process of fine-tuning our k-Nearest Neighbors (KNN) model to achieve optimal performance for predicting Bitcoin prices. Our goal was to find the best set of hyperparameters that produced the most accurate predictions, while maintaining a balance between model complexity and generalization. The choice of k has a significant impact on the performance of our KNN model. A small k value can lead to overfitting, whereas a large k value might result in underfitting. To determine the optimal k, we employed a grid search method, iterating through various k values and assessing the model's performance on a validation set using evaluation metrics such as accuracy, precision, recall, and F1-score. We selected the k value = 4 which yielded the best balance between these metrics.

To reduce the model's complexity and improve its generalization capabilities, we performed feature selection to identify the most relevant input features for predicting Bitcoin prices. We used Recursive Feature Elimination (RFE) to rank the features based on their importance. We then iteratively removed the least significant features and assessed the impact on the model's performance. This process allowed us to identify a subset of features that resulted in the best balance between model complexity and predictive power. To ensure that our fine-tuned KNN model generalizes well to unseen data, we conducted a thorough validation process. We employed technique k-fold cross-validation to estimate the model's performance on multiple subsets of the data. This process allowed us to verify the robustness of our model and provided

confidence in its predictive capabilities. By carefully fine-tuning the hyperparameters and preprocessing steps, we were able to achieve an optimal KNN model for predicting Bitcoin prices.

Time Series

In our study, we tried some time series models to predict Bitcoin prices. We test how well they could predict using out-of-sample R-squared (OSR2). We looked at four popular models: Random Walk, Auto-Regressive (AR) models, Random Forest, and ARIMA. We found that Random Walk isn't complex enough to predict Bitcoin prices accurately, and it has overfitting problems, making it not good for real use.

On the other hand, the AR model did well, with an impressive OSR2 of 0.987. This means it can find short-term patterns in the data well. But, the AR model does have an overfitting problem, which is common with time series data. This overfitting happens because AR models use past Bitcoin prices to predict future values. Even with this problem, the AR model is strong for short-term predictions, especially for regular investors.

For the ARIMA model, it didn't give accurate Bitcoin price predictions because it needs stationary data. Financial data, like cryptocurrency data, is usually not stationary. So, the ARIMA model isn't good for this prediction.

The Random Forest model gave pretty good results, with an OSR2 of 0.704. This shows it can predict patterns without overfitting. It might not be as exact as the AR model for

short-term predictions, but the Random Forest model is useful for predicting long-term trends. This helps both regular and big investors.

In conclusion, the AR and Random Forest models were the best at predicting Bitcoin prices among all the time series models. They can be used well in both retail and institutional situations. But, it's important to remember that for big investors, the data might not be detailed enough (like at the second-scale) to be as helpful as it is for regular investors.

Neural Network

Inspired by biological neural networks in the human brain, a Neural Network (NN) is a data-driven, computational model that can recognize patterns and relationships in raw data. It consists of interconnected layers of nodes or "neurons" that work in unison to analyze and learn from data. Such capabilities make NNs ideal for various tasks, including image recognition, natural language processing, and predictive analytics.

One type of NN is the Feed-Forward Neural Network (FFNN), characterized by the absence of cycles or loops. In this model, information flows in one direction, from input to output. Despite its relative simplicity, an FFNN can solve complex problems when properly trained.

An FFNN model was employed to predict Bitcoin prices, which typically range between 32,000 and 52,000 units. This model featured four layers using Rectified Linear Unit (ReLU) and purelin activation functions. To improve performance, the input features were scaled using

StandardScaler. A learning rate of 0.001 was set, and the Mean Squared Error (MSE) was used as the loss function.

The model's effectiveness is evident in its results: a Root Mean Squared Error (RMSE) of 4426.97, a Mean Absolute Error (MAE) of 3487.60, and an R^2 score of 0.91. Despite the seemingly high MAE, it's deemed reasonably small given the Bitcoin price range. The R^2 score indicates that the model can explain approximately 91% of the variance in Bitcoin prices, confirming a good fit.

Both training and validation losses decreased and converged to a similar value, signifying that the model is performing well. Despite its simplicity, this FFNN model demonstrated fair performance, underscoring the relevance of the selected features.

Result Analysis

In our study, we deployed and evaluated the performance of various prediction models, including random forest, CART, linear regression, and a baseline model, to predict Bitcoin prices. To assess the performance of these models, the standard is to use several metrics, such as precision, accuracy, and true positive rate (TPR). Due to the nature of Bitcoin price prediction, we considered accuracy as the most critical metric, as the purpose of our model is to be able to successfully capture the fluctuation in the Bitcoin price. Table 1 below summarizes the performance of each model using the metrics such as PRE, ACC, and TPR:

Model	Precision	Accuracy	TPR
Baseline	N/A	0.5	N/A
Random Forest	0.75	0.82	0.8
CART	0.68	0.72	0.74
Linear Regression	0.70	0.75	0.76

The random forest model outperformed the other models and the baseline, achieving the highest accuracy of 0.82. Compared to the baseline model, our model has greatly maximized the chance of successfully predicting bitcoin price movement. Therefore, we selected the random forest model for further analysis.

We then evaluated the significance of each data input on the performance of the random forest model

Data Input	Accuracy with Input	Change in Accuracy
Twitter Sentiment	0.86	+0.04
Reddit Sentiment	0.85	+0.03
GDP Growth	0.84	+0.02
Inflation Rate	0.83	+0.01
Market Capitalization	0.82	No change
Trade Volume	0.82	No change
Moving Average	0.81	-0.01

RSI	0.80	-0.02
MACD	0.79	-0.03
Commodity Channel Index	0.78	-0.04

The results from Table 2 indicate that Twitter sentiment had the most significant positive impact on the model's accuracy, followed by Reddit sentiment, GDP growth, and the inflation rate. Market capitalization and trade volume did not have a noticeable impact on the model's accuracy. Given the market cap and trade volume movement of bitcoin has become placid, we predict that those two data inputs will no longer dominate or take part in the price movement.

On the other hand, some technical analysis features, such as moving average, RSI, MACD, and commodity channel index, negatively impacted the model's accuracy. The reduced accuracy may be attributed to the complex and volatile nature of the Bitcoin market, which can be influenced by various factors that are not necessarily captured by traditional technical indicators. Moreover, the inclusion of too many features can lead to overfitting, causing the model to perform poorly on unseen data.

Discussion and Conclusion

In conclusion, our study builds upon secondary research on cryptocurrency price prediction and leverages advanced AI algorithms and predictive modeling techniques. The data collected from tweets, Reddit posts, and online blogs offer valuable insights into market sentiment and trends. Our methodologies, including time series analysis, sentiment analysis, technical analysis, and neural networks, have proven effective in accurately predicting cryptocurrency price movements. These preliminary results demonstrate the potential of AI algorithms in this field and provide insight for future research and development. A prospective team continuing this project should consider refining the models, incorporating additional data sources, and exploring novel techniques to enhance prediction accuracy further.

Citation

1. Chen, Junwei. 2023. "Analysis of Bitcoin Price Prediction Using Machine Learning" *Journal of Risk and Financial Management* 16, no. 1: 51 https://doi.org/10.3390/jrfm16010051
2. Chen, Junwei. 2023. "Analysis of Bitcoin Price Prediction Using Machine Learning" *Journal of Risk and Financial Management* 16, no. 1: 51. https://doi.org/10.3390/jrfm16010051
3. Patrick Jaquart, David Dann, Christof Weinhardt, “Short-term bitcoin market prediction

via machine learning", The Journal of Finance and Data Science, Volume 7, 2021, Pages 45-66, ISSN 2405-9188, https://doi.org/10.1016/j.jfds.2021.03.001.

4. Barr, K. (2022, November 15). *Vast majority of people who invest in bitcoin inevitably lose money, study shows*. Gizmodo. Retrieved February 15, 2023, from https://gizmodo.com/bitcoin-crypto-bank-of-international-settlements-1849784466
5. Coulter, Kelly Ann. "The Impact of News Media on Bitcoin Prices: Modelling Data Driven Discourses in the Crypto-Economy with Natural Language Processing." Royal Society Open Science 9, no. 4 (April 20, 2022). https://doi.org/10.1098/rsos.220276.
6. Gurrib, Ikhlaas, Firuz Kamalov, and Linda Smail. "Bitcoin Price Forecasting: Linear Discriminant Analysis with Sentiment Evaluation." *ArabWIC 2021: The 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research, Sharjah, UAE*, 2021. https://doi.org/10.1145/3485557.3485561.
7. Kaastra, Iebeling, and Milton Boyd. "Designing a Neural Network for Forecasting Financial and Economic Time Series." *Neurocomputing* 10, no. 3 (April 1996): 215–36. https://doi.org/10.1016/0925-2312(95)00039-9.
8. Kaman, Sweta. "News Sentiment Analysis by Using Deep Learning Framework." *News Sentiment Analysis By Using Deep Learning Framework*, 2020. https://doi.org/10.14293/s2199-1006.1.sor-.ppcv5ia.v1
9. *Mai, Feng, Zhe Shan, Qing Bai, Xin (Shane) Wang, and Roger H.L. Chiang. "How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis." Journal of Management Information Systems 35, no. 1 (2018): 19–52*
10. Mittal, Aditi, Vipasha Dhiman, Ashi Singh, and Chandra Prakash. "Short-Term Bitcoin

Price Fluctuation Prediction Using Social Media and ...” IEEE Xplore, 2019. https://ieeexplore.ieee.org/document/8844899.

11. Philippas, Dionisis, Hatem Rjiba, Khaled Guesmi, and Stéphane Goutte. “Media Attention and Bitcoin Prices.” Finance Research Letters 30 (September 2019): 37–43. https://doi.org/10.1016/j.frl.2019.03.031.
12. Riveroll, F. (2020, March 14). *Predicting bitcoin price with business news (python)*. Medium. Retrieved February 15, 2023, from https://medium.com/swlh/predicting-bitcoin-price-with-business-news-python-f3bcf60f5818
13. Sattarov, Otabek, Heung Seok Jeon, Ryumduck Oh, and Jun Dong Lee. “Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis.” *2020 International Conference on Information Science and Communications Technologies (ICISCT)*, 2020. https://doi.org/10.1109/icisct50599.2020.9351527.
14. Steinert, Lars, and Christian Herff. “Predicting Altcoin Returns Using Social Media.” PLOS ONE 13, no. 12 (2018).
15. *Tandon, Chahat, Sanjana Revankar, Hemant Palivela, and Sidharth Singh Parihar. “How Can We Predict the Impact of Social Media Messages on the Value of Cryptocurrency? Insights from Big Data Analytics.” International Journal of Information Management Data Insights 1, no. 2 (2021): 100035.*
16. Throuvalas, A. (2022, December 20). *Portion of bitcoin supply held by retail investors reaches all-time high: Glassnode*. Decrypt. Retrieved February 15, 2023, from https://decrypt.co/117685/portion-bitcoin-supply-held-retail-reaches-all-time-high-glassnode

17. Yang, C. (2019, February 3). *How to use news articles to predict BTC price changes*. Medium. Retrieved February 15, 2023, from https://towardsdatascience.com/how-to-use-news-articles-to-predict-btc-price-changes-c37e305a97f
18. Brown, E. (2012) *Will twitter make you a better investor? A look at sentiment, user reputation and their effect on the stock market* [Preprint].
19. *Mary C. Lacity & Marius A. Janson (1994) Understanding Qualitative Data: A Framework of Text Analysis Methods, Journal of Management Information Systems, 11:2, 137-155, DOI: 10.1080/07421222.1994.11518043*
20. Atanasov, Todor. “What Is R2 Value in Excel?” keysdirect.us. keysdirect.us, January 29, 2023. https://keys.direct/blogs/blog/what-is-r2-value-in-excel