

7151CEM - Computing Individual Research Project

CW1 Project Brief

Module Guide: DR Robert Jinks

Project Guide: Simmon Billings

SID: 10864627

Mail: muchuramiv@uni.coventry.ac.uk

LEVERAGING TWITTER SENTIMENT ANALYSIS TO ANTICIPATE BITCOIN VALUE

Section A – Ethics Application:

<input checked="" type="checkbox"/>	I submitted my ethics application and my application has been approved. I include my ethics certificate in the appendix as evidence.
<input type="checkbox"/>	I submitted my ethics application and my application is currently under review.
<input type="checkbox"/>	I have not submitted my ethics application.

Section B – Project Proposal:

1. Research Question, Problem Statement or Topic for Investigation

Aim:

To anticipate value fluctuations by Twitter sentiment analysis on Bitcoin.

Objectives:

- Using Twitter data of users to analyse the impact on Bitcoin price over time.
- Utilizing natural language processing to find negations, contractions, punctuation, and caps to give scores between -1 and 1.
- Calculate the score to resemble the importance of each tweet by different variables depending on weights and number of retweets, likes and the number of users that follow the tweet.
- Perform EDA on the data to delete duplicates, normalize and check for correlation.
- Create a model to identify words in the data using python libraries with the help of the nltk module.

Research Question:

- How Twitter users are reacting to the Bitcoin prices over time by performing sentimental analysis.
- To determine and understand if user sentiment on Twitter influences the Bitcoin price to rise or fall.

Problem Statement:

As technology advances in the 21st century, the currency is changing to decentralized networks. Bitcoin is the cryptocurrency that came into existence to effectively challenge and help people with transactions around the world without any hassle. Bitcoin is the type of currency that helps people to store and transact their wealth without any 3rd party. Though

many investors and businesses are backing the crypto assets it is considered to be highly volatile despite their contribution to society. Influencers and Billionaires like Elon Musk, Kevin O'Leary, Gary Vee etc., through their social media accounts, might be responsible for the Bitcoin value fluctuations (Theseus.fi. et al., 2022).

2. Intended User or group and their requirements:

In this thesis, by recognising the issue using social media like Twitter how people's opinions on Bitcoin price are increasing by the force of only a single entity and/or the huge number of people boosting the price. This will help key investors' behaviour in investing or selling the stock (Raviv, P M et al., 2018). The vision of this project is to use statistical analysis and data mining techniques to conclude social media influence and bitcoin price.

3. Systems requirements, project deliverables and final project outcome:

As mentioned in the research question, this study's goal is to find the social media influence on bitcoin using Data Mining, and Statistical Analysis.

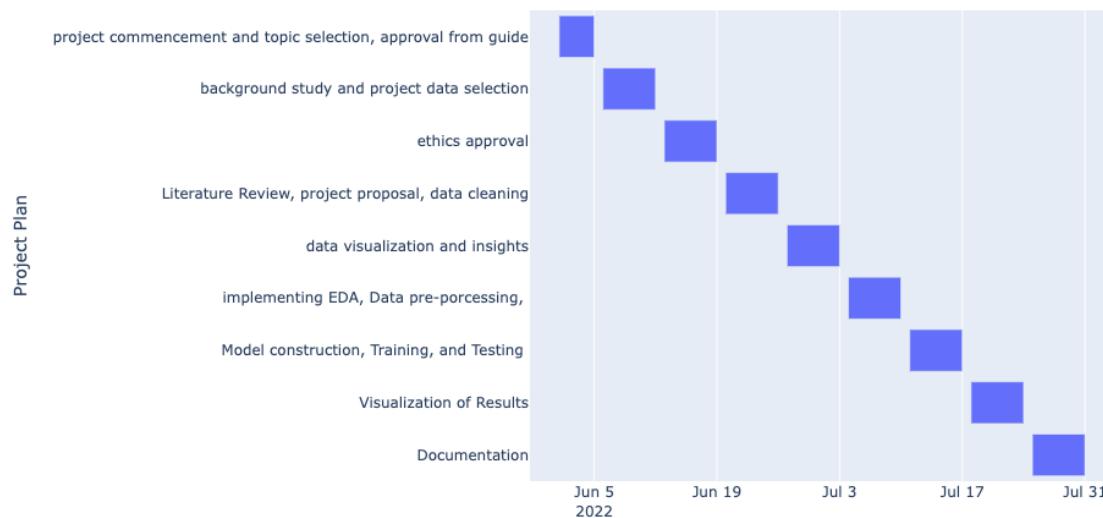
1. Downloading the data from Kaggle
<https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets>.
2. Analysing, cleansing, and pre-processing it.
3. Finding the subjective and polarity of the data.
4. Calculating the sentiment score of tweets.
5. Applying data mining techniques.
6. Evaluating the developed project.
7. Concluding the research.

To deliver the project, MacOS, Brave Browser, Web IDE - Google Colab (Google Colab or Collaboratory allows the user to create and execute Python in the browser with no setup required, free GPU access, and convenient publishing.), Python programming language, Python libraries (vaderSentiment, Pandas, NumPy, Matplotlib, seaborn, Scikit-learn, Keras, TensorFlow) will be imported.

4. Primary Research Plan:

This primary research for the project requires 3 phases.

1. **Phase 1** includes selecting the project topic and dataset, doing a background search on outcomes, creating an ethics form, submit a proposal document.
2. **Phase-2** includes code, like creating the dataframe of the dataset, cleaning, data pre-processing, splitting the data for modelling, training, testing, and finding the accuracy.
3. **Phase-3** includes documentation. Creating documentation on Introduction, Literature Review, Methodology, Requirements, Analysis, Design, Implementation, Risk Management, Critical Appraisal, Conclusion, And References.



5. Initial/Mini Literature Review:

This section involves the past reviews and background or related works of the project, the cryptocurrencies are discussed on the basis followed by Twitter sentiment analysis and its role in the financial markets.

The authors of “Prediction of Bitcoin Price using Deep Learning Model” utilized yahoo finance data of 7 years to get this value, they used deep learning models like Recurrent Neural Networks(RNN) (S. E. Freeda, T. C. E. Selvan and I. G. Hemanandhini et al., 2021). They had compared with various machine learning models such as svm, gaussian naïve bayes, random forest, svm and K-Nearest neighbours. They had got 76.99 percent when compared to other models. While the accuracy got better but the log loss was 7.18 percent.

A study on “Bitcoin price prediction” focused on predicting the Bitcoin on certain criteria or attributes that impact the bitcoin price (L. Felizardo, R. Oliveira, E. Del-Moral-Hernandez, and F. Cozman, et al., 2019). The researchers conducted an extensive study on other factors that affect the accuracy the price prediction and now they are selecting factors that impact the price. They used support vector machines to analyse the data in an 80:20 ration. The main advantage of svm is that they have a high space complexity and a low time complexity.

A study on “Using time-series and sentiment analysis to detect the determinants of bitcoin prices” (Kraaijeveld, O., & De Smedt, J. et al., 2020) twitter sentiment analysis by georgoula used svm regression models to forecasted the fluctuations in their study (Georgoula et al., 2015). They achieved 89.6 percent of accuracy and discovered just a short-term association between twitter sentiment is positive on bitcoin. Another study (Garcia and Schweitzer, et al., 2015) found that twitter sentiment polarity follows price changes using a lexicon-based method with granger-causality testing and vector autoregressive(VAR) model. Another study (Mai et al. 2015) combined intraday analysis and demonstrated that tweets may be used to anticipate bitcoin prices.

Similar methodologies have been used in research that forecasted, where a sentiment analysis methodology is frequently paired with a (Granger_) causality test and/or a regression model to clean their twitter data, researchers used sophisticated pre-processing techniques such as stop-word removal, stemming, tokenisation, and filtering out non-English tweets. Many of the publications that adopt a lexicon-based approach make use of the Loughran and McDonald financial corpus(Mai et al. 2015, Li et al., 2014, Karalevicius et al., 2018).

However, this study was constrained by the fact that the bitcoin prices was obtained by the form of only one exchange or is the social media like twitter is making a part to go prices to go up and down.

6. Bibliography:

1. Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis. (2020, November 4). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9351527/?sessionid=ietzsfngSFOKoX3MQy_PMriPYyilt76ZfNbE7tHi-5Vy1G_wdCu!-508165331?tp=&arnumber=9351527
2. Raviv, P. (2018, June 9). Blockchain Makes Sentiment Analysis Made Affordable to All. Bitcoinist.Com.<https://bitcoinist.com/blockchain-makes-sentiment-analysis-made-affordable/>
3. Theseus.fi. 2022. The impact of Twitter user sentiment on Bitcoin pricing value. [online] Available at: <<https://www.theseus.fi/bitstream/handle/10024/510978/Final%20Thesis%20Report.pdf?sequence=2>> [Accessed 17 June 2022].
4. Xu, J., Bai, W., Hu, M., Tian, H., & Wu, D. (2020). "Bitcoin miners: Exploring a covert community in the Bitcoin ecosystem." ("Bitcoin miners: Exploring a covert community in the Bitcoin ecosystem ...") ("Sci-Hub | Bitcoin miners: Exploring a covert community in the Bitcoin ...") Peer-to-Peer Networking and Applications, 14(2), 644–654. <https://doi.org/10.1007/s12083-020-01021-1>
5. S. E. Freeda, T. C. E. Selvan and I. G. Hemanandhini, "Prediction of Bitcoin Price using DeepLearning Model," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2021, pp. 1702-1706, doi: 10.1109/ICECA52323.2021.9676048
6. Georgoula et al., 2015 Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D.N., Giaglis, G.M., 2015. "Using time-series and sentiment analysis to detect the determinants of bitcoin prices." ("(PDF) Using Time-Series and Sentiment Analysis to Detect the ...") SSRN Electron. J.
7. Garcia and Schweitzer, 2015 D. Garcia, F. Schweitzer Social signals, and algorithmic trading of Bitcoin Roy. Soc. Open Sci., 2 (9) (2015), p. 150288
8. Mai et al., 2015 F. Mai, Q. Bai, Z. Shan, X.S. Wang, R.H. Chiang From Bitcoin to big coin: the impacts of social media on Bitcoin performance SSRN Electron. J. (2015), pp. 1-16
9. Li et al., 2014 X. Li, H. Xie, L. Chen, J. Wang, X. Deng News impact on stock price return via sentiment analysis Knowl.-Based Syst., 69 (1) (2014), pp. 14-23
10. Karalevicius et al., 2018 V. Karalevicius, N. Degrande, J. De Weerd Using sentiment analysis to predict interday Bitcoin price movements J. Risk Financ., 19 (1) (2018), pp. 56-75

11. Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, 65, 101188. doi: <https://doi.org/10.1016/j.intfin.2020.101188>

Appendix:

LEVERAGING TWITTER SENTIMENT ANALYSIS TO ANTICIPATE BITCOIN VALUE

P137852



Certificate of Ethical Approval

Applicant: Vamsidhar Muchurami
Project Title: LEVERAGING TWITTER SENTIMENT ANALYSIS TO ANTICIPATE BITCOIN VALUE

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval: 15 Jun 2022
Project Reference Number: P137852