Bitcoin, Ethereum and Binance Price Prediction using Social Media Sentiments

Sri Vishnuvardhan Reddy Akepati[1] , L. Bhavya[2]

[1]B.Tech in Computer Science and Engineering, JNTUA College of Engineering, Pulivendula, INDIA

(Email id: srivishnuakepati@gmail.com)

[2]Assistant Professor, Department of CSE, JNTUA College of Engineering Pulivendula, INDIA

(Email id:. bhavyalevadala@gmail.com)

ABSTRACT: In this investigation, Twitter data and Reddit posts are used to develop a strategy for projecting the price movements of Bitcoin, Ethereum, and Binance. Bitcoin, Ethereum, and Binance are the three top cryptocurrencies in terms of capitalization, representing over $1,275 billion in combined value. Understanding the relationship between public mood and cryptocurrency prices can thus lead to more accurate price predictions. A method is introduced for real-time and adaptive cryptocurrency price prediction based on Twitter sentiments, and an exploratory analysis is conducted on a network of Reddit communities related to these cryptocurrencies to analyze Bitcoin, Ethereum, and Binance price movements. This research includes both long-term analysis with time series analysis and short-term analysis with social media sentiments. Time series features were collected from submissions and comments on selected Twitter and Reddit postings, tested against cryptocurrency prices, and then used to forecast cryptocurrency price changes using classification models. Price fluctuations for Bitcoin, Ethereum, and Binance are forecasted

Keywords: Cryptocurrency, Bitcoin, Ethereum, Binance, Twitter, Reddit, LSTM, Sentiment Analyzer.

1. Introduction:

Bitcoin is a decentralized electronic money system that, following Satoshi Nakamoto's development, ushered in a major revolution in the financial system. It indicates a shift in technology toward peer-to-peer networks and encryption technologies. Bitcoin, Ethereum, and Binance are not governed or controlled by any government or bank due to their decentralized nature and technological processes. Due to its frequent media mentions and diffusion, cryptocurrency has grown in popularity and attracted a large number of users. Because of its popularity, the price of cryptocurrencies, which fluctuates in real-time like a stock market, is highly intriguing to develop a model that can predict cryptocurrency prices in real-time using online social media data. It can be valuable for investors, businesspeople, banks, organizations, and anyone who uses cryptocurrency for transactions if it can be forecasted with an acceptable degree of precision.

With the introduction of communication networks such as Twitter, Facebook, Instagram, and blogs, communicating information and experiences has become straightforward. Every day, hundreds of thousands of Twitter users, for example, generate enormous amounts of Bitcoin-related tweet data. By utilizing technologies such as Machine Learning, Natural Language Processing, Time Series Analysis, and others, this massive data can be used to explore Bitcoin patterns using continually generated real-time data from social networking sites. Since Bitcoin is a new phenomenon, it is still novel to use social media data to apply machine learning and deep learning with improved accuracy and speed. Although there is a lot of study on the application of various machine learning approaches for predicting time series, there isn't much research on cryptocurrencies. This research could bridge the gap by establishing a model based on cryptocurrency price history and Twitter sentiment for successful cryptocurrency price prediction.

We tackle this issue by obtaining cryptocurrency price history and collecting Bitcoin, Ethereum, and Binance-related tweets, then conducting sentiment analysis on them. We assessed the association between tweet sentiment and price direction before applying average sentiment analysis scores to predict the direction of price for cryptocurrencies. We tested the categorization model using a variety of inputs. We can then improve the accuracy of model by using the outputs. Methodology and analyzing part of the proposed module will be discussed in Section III and Section IV, respectively. Finally, Section IV concludes our job.

2. Related Work:

The term 'efficiency' comes up initially when discussing financial asset forecasting. If the price contains all forms of information, such as prospective prices, previous movements, and so on, the market is said to be efficient. In this situation, market investors will be unable to foresee it because all investing strategies are already factored into the price, with only independent and unexpected circumstances moving the price [Fama, 2021]. Many scholars have attempted to investigate the market's efficiency ([Sensoy, 2019], [Mensi et al., 2019], [Urquhart, 2016]), concluding that it is not yet efficient but is on its way to becoming so. Based on this, there is still room for prediction, as seen by the use of prediction tools. Akyildirim et al. (2021) used a variety of machine learning algorithms to predict the values of twelve cryptocurrencies, with promising results. The researchers found that the market is inefficient because historical price movements assist in anticipating future prices. According to Chevallier et al. (2021), cutting-edge investment techniques such as Buy and Hold outperform Machine Learning approaches.

The field of bitcoin price prediction is still relatively new, and we might discover research that contradicts one another in the literature. Even still, many of them did not see social media's influence as a factor in their decision. Some papers have previously attempted to examine the relationship between social media and the bitcoin market. Verma and Sharma (2020) used an ELMo embedding model and an SVM (Support Vector Machines) classifier to try to predict the Bitcoin price using sentiment from the Twitter social media, from which a main observation for the current project is that Bitcoin prices are not affected by investor sentiment, at least on Twitter, a conclusion that is contradicted by Jahjah and Rajab (2020), who found a strong relationship between Twitter sentiment and the Bitcoin price. More research is discovered, with viable applications built using various machine learning and deep learning approaches ([Pant et al., 2018], [Pathak and Kakkar, 2020]). Reddit, which is within the scope of this project, has also been used to predict cryptocurrency price swings ([Ortu et al., 2021]), such as pump-and-dump schemes ([Nghiem et al., 2021]). Another example of using social media as a data source is Google Trends, which was utilized by Jethin et al. (2018) to anticipate Bitcoin and Ethereum prices, along with Twitter sentiment."

3. Methodology

3.1 Data :

The pricing of the Binance cryptocurrency exchange is used. The following criteria are used to rank and assess cryptocurrency exchanges: Date, Open, High, Low, Close, Adj Close, and Confidence that the volume recorded by an exchange is accurate. Extracting postings from Twitter and Reddit is essential for sentiment analysis. Aside from that, pertinent data about them is gathered, such as the number of likes, comments, and shares. The posts are sorted by cryptocurrency name, with Bitcoin, Ethereum, and Binance as keywords. Twitter is a social media network that receives 500 million tweets every day, implying that the tags in question will generate a large number of tweets. Because the Twitter API (Application Programming Interfaces) only allows for a search of 450 tweets in a 15-minute window, this rate limit imposes a constraint on the current project, which necessitates the usage of the 'snscrape' tool. Snscrape is a scraper for social media platforms (SNS). It scrapes tweets for information such as usernames, hashtags, text, likes, and retweets. Reddit produces 303.4 million posts each month [Hutchinson, 2020], which is comparable to Twitter. Because the Reddit API has search constraints, we used the pushshift method to retrieve Reddit content. Our code attempts to extract the text of postings made on these social networking platforms. Data is available from January 1, 2022, to March 30, 2022.

3.2 Data Preparation:

The data preparation process begins by removing emoticons, special characters, exclamation marks, and question marks from the associated posts. This article attempts to apply sentiment analysis approaches to Reddit posts and Twitter data related to cryptocurrencies. The goal of this research is to investigate the relationship between sentiment analysis and bitcoin market movement, with the lead net returns as our target variable. Sentiment analysis is performed on Twitter and Reddit data using TextBlob, Flair, and Vader. VADER (Valence Aware Dictionary and Sentiment Reasoner) is a Python library that specializes in social media sentiment analysis. It includes an algorithm that adjusts sentiment intensity based on punctuation, slang, emoticons, and acronyms. VADER generates four different scores: compound, negative, neutral, and positive. The TEXTBLOB command returns the polarity and subjectivity of a sentence. Polarity is defined as the difference between negative and positive sentiment. The polarity is reversed with negative terms. The sentiment classifier in Flair is based on a character-level LSTM neural network that predicts based on letter and word sequences.

3.3 Model Building:

Initially, Twitter data for Bitcoin, Binance, and Ethereum is extracted from the Twitter API. The Twitter API (Application Programming Interfaces) only allows for a search of 450 tweets in a 15-minute window, which imposes a constraint on the present project. Therefore, Snscrape is utilized. Sentiment analysis is then applied to the Twitter data and Reddit data using TextBlob, Flair, and Vader. After applying sentiment analysis and processing the Twitter data and Reddit data related to these cryptocurrencies, a prediction of lead net returns is conducted.

4. Implementation of the prototype

The RMSE values of Bitcoin, Ethereum, and Binance are then obtained by passing these intermediate net values via LSTM. The LSTM (long short-term memory network) is a communication memory network. The LSTM, or long short-term memory network, is a recurrent neural network that addresses the problem of fading gradients. This type of recurrent neural network is used in profound learning because it can teach extremely large designs. LSTM allows the network to learn more about many different time steps by keeping a more constant error. As a result, the network can gradually develop mutual trust. An LSTM cell's forget and remember gates allow it to decide whether to transmit or block data based on its strength and relevance. Weak signals that keep the gradient from fading are thus suppressed. The performance of the RNN and LSTM networks is examined to determine the model's efficiency.

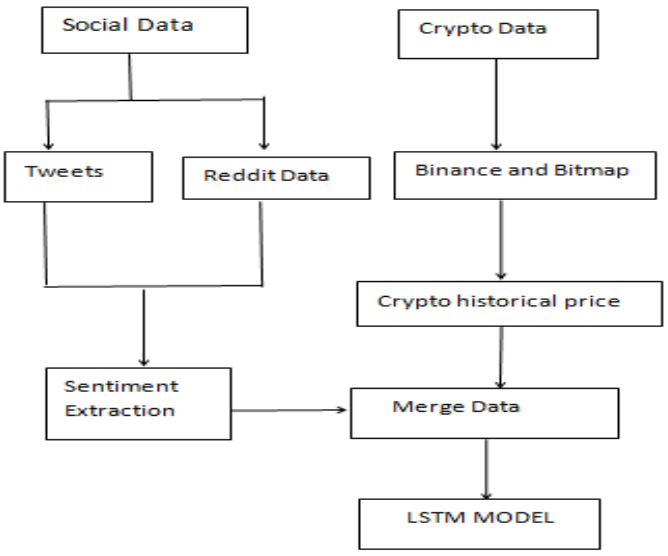

Fig 1: Prototype of the methodology

4.1 Long Term Investments:

LSTM time series models are created without sentiments using cryptocurrency data, such as open, close, high, low, and volume, for three crypto currencies from the time of their market introduction in this work.

4.2 Results:

The fig 1.1 shows the RMSE values of Bitcoin, Ethereum and Binance using twitter and reddit sentiment analysis. The fig 1.2 shows the MAE values of Bitcoin, Ethereum and Binance using twitter and reddit sentiment analysis.

Crypto Currency	Twitter		Reddit	
	Train Score (RMSE)	Test Score (RMSE)	Train Score (RMSE)	Test Score (RMSE)
Bitcoin	2251.2410	5638.555	652.38	1047.8
Ethereum	310.4340	229.2685	70.70	69.78
Binance	40.1644	31.8748	6.55	7.04

Fig 1.1: RMSE values

Crypto Currency	Twitter		Reddit	
	Train Score (MAE)	Test Score (MAE)	Train Score (MAE)	Test Score (MAE)
Bitcoin	1548.0	5516.7	337.80	482.61
Ethereum	253.02	192.60	52.83	54.39
Binance	34.114	28.283	4.89	5.50

Fig 1.2 : MAE values

The below fig 1.3 shows the RMSE and MAE of Bitcoin, Ethereum and Binance LSTM time series model developed in long term Investments

Crypto Currency	RMSE		MAE	
	Train Score	Test Score	Train Score	Test Score
Bitcoin	698.4	13427.4	341.70	11643.6
Ethereum	29.97	2034.45	18.42	1584.9
Binance	20.52	18.59	5.77	8.01

Fig 1.3: RMSE and MAE for long term investments

5. Conclusion:

The association between social media factors and cryptocurrency markets (Bitcoin, Ethereum, and Binance) was explored. The dynamics of these variables, their linkages, and their implications for the next day's returns were investigated using various statistical tests (lead net returns). According to the findings, while social media variables provide some information for predicting future cryptocurrency price movements, they are insufficient; more data is needed to determine whether these price movements are influenced by other factors such as macroeconomic variables, political landscapes, and so on.

References:

1. [Abraham et al., 2018] Abraham, J., Higdon, D., Nelson, J., and Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis.

2. [Akyildirim et al., 2021] Akyildirim, E., Goncu, A., and Sensoy, A. (2021). Prediction of cryptocurrency returns using machine learning. Annals of Operations Research, 297(1):3–36. [Ante, 2021] Ante, L. (2021). How elon musk's twitter activity moves cryptocurrency markets.

3.[Anwar and Yaqub, 2020] Anwar, A. and Yaqub, U. (2020). Bot detection in twitter landscape using unsupervised learning. In The 21st Annual International Conference on Digital Government Research, pages 329– 330.

4. [Berentsen and Schar, 2018] Berentsen, A. and Schar, F. (2018). A short introduction to the world of cryptocurrencies.

5.[Brito and Castillo, 2013] Brito, J. and Castillo, A. (2013). Bitcoin: A primer for policymakers. Mercatus Center at George Mason University.

6.[Chaum, 1983] Chaum, D. (1983). Blind signatures for untraceable payments. In Advances in cryptology, pages 199–203. Springer.

7. [Chong et al., 2017] Chong, E., Han, C., and Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83:187–205.

8. [Choueiry, 2021] Choueiry, G. (2021). What is an acceptable value for vif? https://quantifyinghealth.com/vif-threshold/ . Accessed: 2021-05-20.

9. [Di Muzio, 2021] Di Muzio, T. (2021). Gamestop capitalism. wall street vs. the reddit rally (part i).

10. [Diggle et al., 2002] Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., et al. (2002). Analysis of longitudinal data. Oxford University Press.

11. [Elder, 1993] Elder, A. (1993). Trading for a living: psychology, trading tactics, money management, volume 31. John Wiley & Sons.

12.[Essaidi et al., 2020] Essaidi, A., Zaidouni, D., and Bellafkih, M. (2020). New method to measure the influence of twitter users. In 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), pages 1–5. IEEE.

13. [Fama, 2021] Fama, E. F. (2021). Efficient capital markets a review of theory and empirical work. The Fama Portfolio, pages 76–121.

14.[Huang et al., 2018] Huang, Y.-P., Hlongwane, N., and Kao, L.-J. (2018). Using sentiment analysis to determine users' likes on twitter. In 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), pages 1068–1073. IEEE