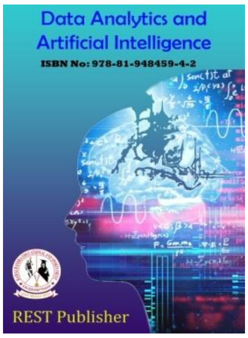
Data Analytics and Artificial Intelligence

Vol: 3(1), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/

DOI:https://doi.org/10.46632/daai/3/1/19



Bit coin Sentiment Analysis

***N. Sharmila Banu, P.P. Ram Prasad, A. Prathosh, K. Venkatesh Prasad**

Veltech high-tech Dr.Rangarajan Dr, Sakunthala Engineering College, Chennai, Tamil Nadu, India.

*Corresponding Author Email: jarmi.it@gmail.com

Abstract: *Twitter sentiment has proven useful in predicting whether the Bit coin price will rise or fall. Increase/decrease. In this article, we seek to rely on technical status not only to predict direction, but also to predict bullish/bearish levels. We use not only present the results of tests that explore the relationship between sentiment and futures prices at different levels of time granularity, with the aim of discovering the optimal time period at which sentiment behaves. Become a reliable indicator of price changes.*

Keywords: *Bit coin, Crypto currency, NLP*

1. INTRODUCTION

Bitcoin (Nakamoto 2009), since its introduction in 2008/9, has attracted considerable attention for a variety of reasons from various stakeholders with differing opinions on the usefulness and its potential. Despite the fact that it and its underlying blockchain technology have been repeatedly declared "dead", Footnote1's recent record price and research suggest the opposite (Ellul 2021). Since such opinions can and do influence the interest of potential investors, it is our desire to investigate whether better methods can be developed to support such opinions. Investors or not. In fact, not only can public opinion influence the interest of potential investors, but public sentiment itself can be used to empower investors to make informed decisions. more about predicting future prices. There is no doubt that sentiment affects the price of an asset - and as Baker and Wurgler put it: "The question is no longer, as it was a few decades ago, whether investor sentiment has an effect. affect stock prices or not, but rather how to measure investor sentiment and quantify its influence" (Baker and Wurgler 2007). Although this statement refers to a well-established literature on the application of sentiment analysis to traditional markets (Gunter et al. 2014; Rao and Srivastava 2012; Li et al. 2014; Mittal et al. Goel 2012), sentiment analysis can also be used to predict cryptocurrency prices, as recent research demonstrates (Valencia et al 2019; Kraaijeveld andDe Smedt 2020; Abraham et al 2018; Stenqvist and Lönnö 2017; Pant 2018; Galeshchuk et al. 2018; Climate in 2020; Naem and associates. Year 2020; Serafin et al. Year 2020; Tourism 2019; Balfagih and Keselj 2019; Mohapatra et al. Year 2020). (Kraaijeveld et De Smedt 2020; Abraham et al. 2018; Stenqvist et Lönnö 2017; Pant 2018; Galeshchuk et al. 2018; Kilimci 2020; Naeem et al. 2020; Balfagih et al. 2019; Mohapatra et al. 2020). Although the state of the art has achieved encouraging results, further research efforts are needed to overcome some of the problems. Many of the following problems are common in price prediction models based on sentiment analysis of Twitter data: (i) valuations are typically based on minimal historical data (Pant 2018; Valencia et al. 2019 Stenqvist and Lönnö 2017; Kilimci 2020); (ii) forecasts tend to be later when predicted prices are actually observed in the market (Serafini et al. 2020; iii) models are often limited to predicting direction (up or down) (Kilimci 2020; Valencia et al 2019; Galeshchuk et al. 2018), although several studies have produced accurate price predictions with limited success (Pant 2018; Li and Dai 2020; Serafini et al. 2020). Some of the problems described above can be attributed to the following specific challenges. The direction and magnitude of price changes are often non-linear and therefore difficult to solve (Kimoto et al. 1990). Additionally, tweets are often copied for marketing purposes and can also be automated by tweet bots (Valencia et al. 2019). Tweets also often contain features that lead to noise (related to sentiment analysis), including hashtags, bio mentions, and URLs (Kraaijeveld and De Smedt 2020). At the same time, the use of sarcasm in tweets can skew sentiment predictions (Rosenthal et al. 2014). Therefore, before solving the general problem of sentiment extraction, it is advisable to conduct pre-processing of tweets to reduce this noise. In this article we look at predicting price changes (beyond simple direction) and to our knowledge this is the first article offered to do so. We present a comprehensive review and concluding results for several models. Furthermore, the proposed models overcome the late prediction problem encountered in the modern state. Furthermore, we investigate the predictive relationship between Twitter sentiment and related price changes based on different time lags. Through this, we address the question of which period of time between the expression of sentiment and the price change yields the best results. An extensive study was performed to determine how different types of neural networks and features used could affect accuracy, where each model studied was evaluated. Prices are based on different combinations of features used as well as different time lags introduced between sentiment and price changes. The rest of this article is organized as follows. The following section provides an overview of sensitivity analysis, followed by recent studies on Bit coin and other crypto price predictions. Then, the classification problems are solved and the methods used for data preprocessing, feature extraction and the proposed neural models are presented. The results are then presented and discussed. Finally, concluding with some orientations for future work.

2. BACKGROUND ON SENTIMENT ANALYSIS

In its raw form, natural language text means nothing to a computer, much more than encoded bytes. Over the past decade, much progress has been made in the field of natural language processing (NLP) with the goal of enabling computer systems to better reason about natural languages. Sentiment analysis, as the name suggests, analyzes and extracts sentiment, opinion, subjectivity, and polarity from the text. Sentiment analysis use cases abound, including but not limited to analyzing product markets and automatically flagging positive/negative/probably harmful feedback on websites and platforms platform of social media. Since the introduction of sentiment analysis in the NLP community (Pang and Lee 2008), several techniques have been proposed to associate polarity with a piece of text. It can be presented as a classification problem, where a piece of text is classified as positive, negative, or neutral. Some approaches also specify a value that reflects the confidence level associated with the respective pole. Vocabulary-based approaches use a vocabulary (a set of words) and associated sentiment scores to compare with the text being classified to determine the ultimate endpoint. A widely used lexical-based implementation, VADER (Valence Aware Dictionary and Sentiment Reasoner) (Hutto and Gilbert 2015), also uses rule matching, which attempts to determine polarity based on the input text at language models in use. VADER combines selected features from three validated Footnote4 vocabularies as well as tweet intensity rules extracted from parsing, grammatical terms, and valence values of 800 tweets (Stenqvist and Lönnö 2017). Over the past decade, sentiment analysis has been widely applied to Twitter data, which indeed “poses new and different challenges” (Agarwal et al. Twitter app. More traditional sentiment analysis (Hussein 2018) and a wide range of approaches (Medhat et al. 2014). In addition, the international SemEvalFootnote5 workshop helped facilitate further research by presenting a series of shared challenges to the community. Particularly relevant to the context of this work, since the 2013 workshop (Nakov et al. 2013), a shared mission focused on sentiment analysis on Twitter was released every year.

3. Related work

1. The aim of this study was to measure the interaction between media sentiment and Bit coin price. Because some researchers have argued that the value of Bit coin is also determined by the perception of users and investors, this article will look at how.

2. This paper targets to become aware of changes at the tweet textual content for the duration of pre processing in order that the ensuing sentiment rankings excellent correlated with Bit coin's last prices. We created distinctive approaches of pre-processing textual content for VADER scoring and examined them on truncated and full-period tweets.

3. The goal of this paper is to decide the predictable charge route of Bit coin in USD with the aid of using system mastering strategies and sentiment evaluation. Twitter and Reedit have attracted a super deal of interest from researchers to take a look at public sentiment. We have implemented sentiment evaluation and supervised system mastering concepts to the extracted tweets from Twitter and Red dit posts, and we examine the correlation among bit coin charge moves and sentiments in tweets.

4. Specifically, the authors train an LDA-based classifier that uses current BTC price information and BTC news announcement headlines to predict the next day's BTC price direction. The authors compare the results with the support vector machine model (SVM) and stochastic estimation method. Using BTC price information and crypto- related news alerts allows us to assess the importance of these different sources and types of information.

5. The motive of this examine is to forecast the route of Bit coin fee via way of means of analysing consumer evaluations in social media along with Twitter. To our knowledge, that is the first actual try which estimates the route of Bit coin fee fluctuations via way of means of the usage of deep gaining knowledge of and phrase embedding fashions with inside the contemporary studies. For the motive of estimating the route of Bit coin, convolution neural networks

(CNNs), recurrent neural networks (RNNs), and long-brief time period reminiscence networks (LSTMs) are used as deep gaining knowledge of architectures and Word2Vec, GloVe, and Fast Text are hired as phrase embedding fashions with inside the experiments. In order to illustrate the contribution of our work, experiments are done on English Twitter dataset.

4. EXISTING SYSTEM

Prognosis is a procedure for forecasting time series data based on an additive model when nonlinear trends are adjusted for annual, weekly, and daily seasonality, as well as the influence of holidays. This works best with time series with strong seasonal effects and multi-season historical data. Prophet is very solid when it comes to missing data and changing trends, and usually handles outliers well. Prophet is an open source software released under the Face book Core Data Science term. It is available for download on CRAN and Py PI.

5. PROPOSED SYSTEM

Opinion mining, or sentiment analysis, is a text analysis technique that uses computational linguistics and natural language processing to automatically identify and extract an emotion or opinion in a text. a text (positive, negative, neutral, etc.). It allows you to get inside your customers' minds and figure out what they like and don't like and why, so you can create products and services that meet their needs. When you have the right tools, you can perform opinion research automatically, on almost any form of unstructured text, with very little human intervention.

6. ARCHITECTURE DIAGRAM

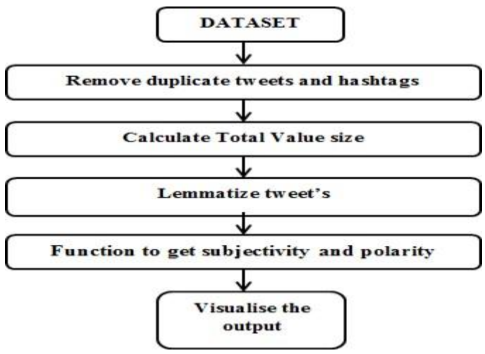


FIGURE 1. Architecture Diagram

Implementation Of Crypto currency Prediction Model: This section illustrates the step-by-step implementation of the project. It describes the whole process of data preparation, data pre processing, emotion score calculation, correlation analysis, and performing predictive deep learning model to predict bit coin price from calculated sentiment score Maths.

(1) ***Data Extraction:*** The data required for the implementation of the bit coin price forecasting project is extracted from two open sources.

(2) ***Twitter:*** Twitter is a famous social media web website online wherein tens of thousands and thousands of humans explicit themselves through" tweets." Likes and re-posts are used to deliver aid or disapproval of the tweets. The information changed into extracted the usage of the Twitter API and the important thing terms" bit coin" and" BTC" to take a look at and understand humans' attitudes in the direction of bit coin. To use python libraries to hook up with the Twitter API, a top class developer account changed into first received to get a excessive get right of entry to price and get right of entry to antique tweets. The usage of the "twy thon" library, which changed into created for extracting tweets the usage of the Twitter API and authentication keys – Access token and Access token mystery – that had been furnished. Id, Text, Username, Number of Followers, Number of Tweets, Number of Likes, and Tweet advent date and time info had been accrued for every tweet. Twitter authenticates with python to get right of entry to its API, the usage of the function" Twython" from the python package" twy thon" and oauthversion as 2. After efficiently authenticating with the Access token and Access token mystery furnished through Twitter, the tweets may be retrieved the usage of the streaming API or through specifying time frames for preceding tweets. The extracted information is then translated to CSV layout and used with inside the information pre- processing.

(3) ***Bit coin Time-Series Data*****:** Bit coin is the world's first decentralized digital currency because it works without the use of a central bank or a single administrator. The system was created to function as a peer-to-peer network, where ransactions are performed directly between users without the use of an intermediary. Bit coin price historical data is obtained from an API, cry ton compare 1. Anyone can access and use bit coin historical price data for free. Hourly bit coin price history for the respective time period of tweets retrieved using the API. Closing price, high price, low price, open price, and timestamp are attributes of the data set. The current implementation is said to be able to predict the hourly closing price

(4) ***Data Pre-Processing:*** Raw Twitter Collection contains various unwanted characters, photos, videos and has htags. In the pre- processing stage, they will be removed as they affect the sentiment scoring for each tweet.

(5) Removal of URLS: Users tend to include hyperlinks in their articles. Extracted tweets with URLs will contribute the least to the sentiment score calculation. In addition, a lot of redundant data will hinder the calculation speed and accuracy of the analysis.

(6) ***Removal of Has htags*****:** Users have a tendency to apply has htags to specific their opinion in brief or to suggest the tweet is associated with a selected topic. Although the tweets with" BTC" are extracted, the person"." might now no longer make contributions anymore submit the instruction of datasets for the calculation of sentiment score. So, the hash person is eliminated from the extracted tweets data.

(7) ***Removal of Username*****:** Users can tag other users in tweets. Normally this is used when one user intends to pass it on to another user, but it is not necessary. As this will not contribute to the sentiment of the tweet, the username in the tweet text is identified with the "@" character and is removed from the text.

(8) ***Removal of Special Characters*****:** Computational speed and analysis performance suffer as the volume of data increases. This is handled by removing redundant data that does not contribute to the analysis. Because special characters like question mark ("?"), exclamation mark ("!"), semicolon (";") and "@" only increase the volume of data to analyze and do not polarity, these characters will be removed during pre processing.

7. SENTIMENT SCORE CALCULATION

As the evaluation is aimed to realize the effect of the emotions of the humans opinion at the charge of the bit coin and the in addition effect of a tweet may be calculated primarily based totally at the fans and likes of the tweets, those elements are taken into consideration for the duration of the calculation of the very last sentiment rating. The rating decided is equal to the compound rating of the tweet expanded through the person fans and the likes attained through the tweet. A well-known persona or an influencer are maximum probably to have greater fans and may get greater like in evaluation to an regular individual that's a sensible scenario. This is taken into consideration for the duration of the calculation of the sentiment rating in opposition to every tweet. User fans depend and likes of the tweet The calculated sentiment rating is then grouped primarily based totally at the time stamp of the tweet. This facts is then transformed right into a csv document which has the attributes time and sentiment rating that became calculated in opposition to every tweet. Closing charge of the hour is selected to be the reaction variable for the prediction of bit coin charge. All different attributes with inside the bit coin charge dataset High charge, Low Price and Opening charge are dropped. As the charge prediction is designed to be made on an hourly basis, the facts in each the datasets associated with bit coin and twitter are aggregated primarily based totally on hour. These data frames are then merged the usage of the time that's not unusual place characteristic in each datasets. A very last pre processed data frame with the attributes time, sentiment rating and bit coin charge in USD is created for the prediction of bit coin charge for the subsequent hour

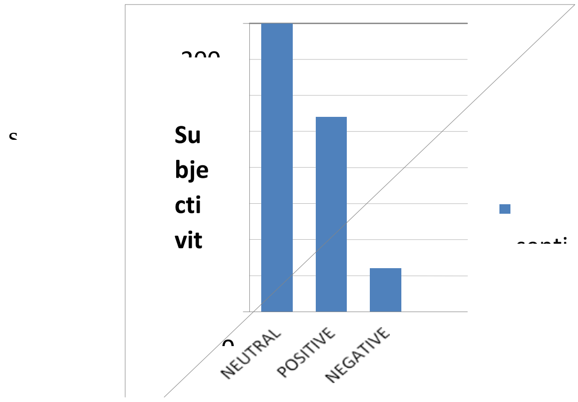


FIGURE 2. Prediction of Bit coin Sentimental Analysis

8. CONCLUSION

In this report, a functional Bit coin sentiment analysis model was developed. We achieved a final accuracy of 89.0% on our dataset. This article compared the performance of several different neural models in predicting cryptocurrency price movements from Twitter tweet data. The underlying hypothesis of this work is that opinions expressed on social networks can act as useful predictors of these fluctuations, especially when they incorporate characteristics such as feelings and attitudes.

REFERENCES

[1] Al-Ghuribi, S. M., Mohd Noah, S. A. and Tiun, S. (2020). Unsupervised Semantic Approach of Aspect- Based Sentiment Analysis for Large-Scale User Reviews, IEEE Access 8: 218592– 218613.

[2] Balfagih, A. M. and Keselj, V. (2019). Evaluating Sentiment C1assifiers for Bitcoin Tweets in Price Prediction Task, 2019 IEEE International Conference on Big Data (Big Data), pp. 5499– 5506.

[3] Bibi, S., Hussain, S. and Faisal, M. I. (2019). Public Perception Based Recommendation System for Cryptocurrency, 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 661–665. ISSN: 2151-1411.

[4] Gualberto, E. S., Sousa, R. T. D., Vieira, T. P. D. B., Costa, J. P. C. L. D. and Duque, C. G. (2020). From Feature Engineering and Topics Models to Enhanced Prediction Rates in Phishing Detection, IEEE Access 8: 76368–76385. Conference Name: IEEE Access.

[5] Han, W., Tian, Z., Zhu, C., Huang, Z., Jia, Y. and Guizani, M. (2021). A Topic Representation Model for Online Social Networks Based on Hybrid Human–Artificial Intelligence, IEEE Transactions on Computational Social Systems 8(1): 191–200. Conference Name: IEEE Transactions on Computational Social Systems.

[6] Martin, K., Alsmadi, I., Rahouti, M. and Ayyash, M. (2020). Combining Blockchain and Machine Learning to Forecast Cryptocurrency Prices, 2020 Second International Conference on Blockchain Computing and Applications (BCCA), pp. 52–58.

[7] Narman, H. S. and Uulu, A. D. (2020). Impacts of Positive and Negative Comments of Social Media Users to Cryptocurrency, 2020 International Conference on Computing, Networking and Communications (ICNC), pp. 187–192. ISSN: 2325- 2626

[8] Saadah, S. and Whafa, A. A. A. (2020). Monitoring Financial Stability Based on Prediction of Cryptocurrencies Price Using Intelligent Algorithm, 2020 International Conference on Data Science and Its Applications (ICoDSA), pp. 1–10.

[9] Sun, J., Zhou, Y. and Lin, J. (2019). Using machine learning for cryptocurrency trading, 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS), pp. 647–652.

[10] Yao, W., Xu, K. and Li, Q. (2019). Exploring the Influence of News Articles on Bitcoin Price with Machine Learning, 2019 IEEE Symposium on Computers and Communications (ISCC), pp. 1–6. ISSN: 2642-7389.

[11] Yiying, W. and Yeze, Z. (2019). Cryptocurrency Price Analysis with Artificial Intelligence, 2019 5th International Conference on Information Management (ICIM), pp. 97– 101.

[12] Zhang, W., Yoshida, T. and Tang, X. (2008). TFIDF, LSI and multi-word in information retrieval and text categorization, 2008 IEEE International Conference on Systems, Man and Cybernetics, pp. 108–113. ISSN: 1062-922X.

[13] Georgoula, D. Pournarakis, C. Bilanakos, "Using time-series and sentiment analysis to detect the determinants of bitcoin prices." Available at SSRN 2607167 (2015).

[14] D. Garcia, F. Schweitzer. "Social signals and algorithmic trading of Bitcoin." Royal Society open science 2.9 (2015): 150288.
[15] V. Pagolu, K. Challa, G. Panda, “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", International conference on signal processing.