



# International Journal of Advanced Research in Arts, Science, Engineering & Management

Volume 10, Issue 3, May 2023



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 6.551**

# Twitter Attribute Classification with Q- Learning on Bitcoin Price Prediction

Dr. Shilpashree S<sup>1</sup>, Ashreetha B<sup>2</sup>, Praneeth Venkat Sai E<sup>3</sup>, G V Vidyashree<sup>4</sup> and Geetha Priya K<sup>5</sup>  
Guide, Associate Professor, Dept. of CSE, Nagarjuna College of Engineering and Technology, Bengaluru, India<sup>1</sup>  
U.G. Student, Dept. of CSE, Nagarjuna College of Engineering and Technology, Bengaluru, India<sup>2,3,4,5</sup>

**ABSTRACT:** Bitcoin price prediction based on people's opinions on Twitter usually requires millions of tweets, using different text mining techniques, and developing a machine learning model to perform the prediction. These attempts lead to the employment of a significant amount of computer power, central processing unit (CPU) utilization, random access memory (RAM) usage, and time. To address this issue, in this project, we consider a classification of tweet attributes that effects price changes and computer resource usage levels while obtaining an accurate price prediction. To classify tweet attributes having a high effect on price movement, we collect all Bitcoin-related tweets posted in a certain period and divide them into four categories based on the following tweet attributes: (i) the number of followers of the tweet poster, (ii) the number of comments on the tweet, (iii) the number of likes, and (iv) the number of retweets. We separately train and test by using the Q-learning model with the above four categorized sets of tweets and the best accurate prediction among them. We compare our approach with a classic approach where all Bitcoin-related tweets are used as input data for the model, by analyzing the CPU workloads, RAM usage, memory, time, and prediction accuracy. The results show that tweets posted by users with the most followers have the most influence on a future price, and their utilization leads to spending 80% less time, 88.8% less CPU consumption, and 12.5% more accurate predictions compared with the classic approach.

## I.INTRODUCTION

Earlier stock market forecasting research relied on past stock values. Most studies have discovered that analyzing previous prices is not sufficient to anticipate stock market changes because stock market prices are highly volatile. According to the efficient market hypothesis, financial market movements are influenced by news, current events, and product releases, all of which have a substantial impact on a company's stock value. As a large stock market, Bitcoin has no central controlling authority and is regulated solely by the public. As a result, Bitcoin is viewed as a volatile cryptocurrency and its value is influenced by public ideas.

Bitcoin (BTC), the decentralized cryptographic currency, is like most commonly known currencies in the sense that it is affected by socially constructed opinions, whether those opinions have basis in facts, or not. Since Bitcoin was revealed to the world, in 2009, it quickly gained interest as an alternative to regular currencies. As such, like most things, opinions and information about Bitcoin are prevalent throughout the Social Media sphere.

"Bitcoin is not a currency it's an asset," Pavan Sukhdev, the president of environmental advocacy group WWF International and a former managing director at Deutsche Bank, told Yahoo Finance in a recent interview. He pointed to the extreme volatility and lack of backing value as reasons for its illegitimacy.

## II.RELATED WORK

In this section, we classify the related research into the following three categories:

- (1) Bitcoin price prediction with public opinion,
- (2) Striving for accurate prediction, and
- (3) Resource usage minimization.

Sentiment analysis is an important field for researchers, as people's thoughts and emotions have become popular and an acceptable technique for examining and analyzing public opinion. Twitter, Facebook, and Instagram are examples of social media platforms used to collect sentiment data for research. The major goal of adopting these approaches is to identify and extract emotions in spoken or written language using natural language processing techniques. Among other social media platforms, Twitter has recently attracted interest from a wide range of academic disciplines, as it is considered useful for analyzing economic and social datasets. The employment of machine learning algorithms on the data extracted from Twitter has opened widely opportunities including identification of hatred speeches, analyzing personalities based on profile pictures, prediction on offensiveness in tweets, etc.

### III.METHODOLOGY

For our project, the chosen programming language is Python. Python is widely recognized as an easy-to-learn and powerful programming language.

#### Libraries:

**PyTorch:** PyTorch is a popular open-source machine learning library widely used for developing and training neural networks. which allows us flexible and efficient execution of our neural network models.

**Matplotlib:** It offers a wide variety of plot types, including line plots, scatter plots, bar charts, histograms, heatmaps, and Seaborn to enhance the visual appeal of the plots generated with Matplotlib. These visualizations help us analyze and present data related to twitter tweet attribute.

**Numpy:** It allows us to work with large amounts of data efficiently and perform various operations on arrays. It utilizes optimized C code under the hood, making it significantly faster than traditional Python lists for numerical computations. This speed advantage is crucial for processing large amounts of data in real-time.

**Working with q-learning algorithm:** In this section, we introduce our approach to predicting Bitcoin prices based on Twitter data. For this, we adopt simple reinforcement learning, in which the environment was the Bitcoin market. For building the model, we used Machine Learning & Reinforcement Learning Algorithm [Q Learning].

### IV.PERFORMANCE EVALUATION

Here we evaluate the performance of each model such as F1 score, Precision, Recall and Accuracy of the model is calculated.

#### Working Model:

**A.DATA COLLECTION:** We use a total of 1690 days' data that is in the time from April 1, 2014, to November 14, 2018, in the Bitcoin price market as real data to predict the Bitcoin price because it was observed that the Bitcoin price fluctuated substantially during this period. This motivates us to verify the effectiveness of the proposed method during this period.

**BITCOIN TWEET DATA:** We use Tweepy and Twitter's streaming API for the Bitcoin-related tweet data. Tweepy is a Python-based open-source framework, makes it easier to gather tweets using Twitter API.

**Yahoo Finance Dataset:** We can view historical price, dividend, and split data for most quotes in Yahoo Finance to forecast the future of a company or gain market insight. **Historical data can be downloaded as a CSV file to be used offline, which you can open with Excel or a similar program.**

**B.SENTIMENT ANALYSIS:** As a final step, we apply sentiment analysis to determine 358 subjective emotions or views expressed in the tweets on Bitcoin. We perform sentiment analysis by categorizing textual views into categories such as ``positive,' ``negative,' or ``neutral.'

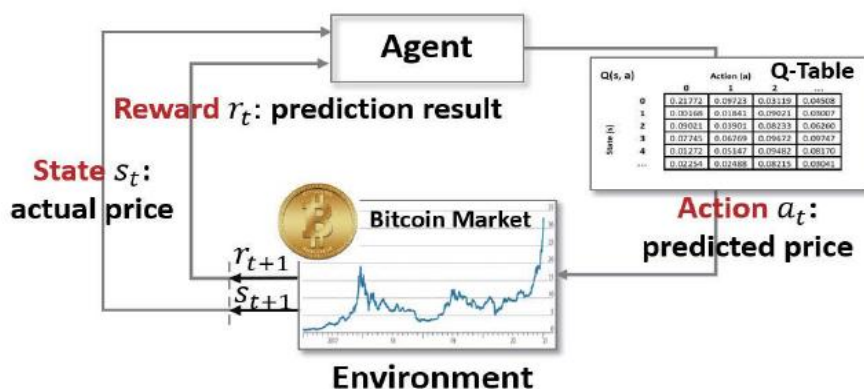


Fig. 1. Working with q-learning

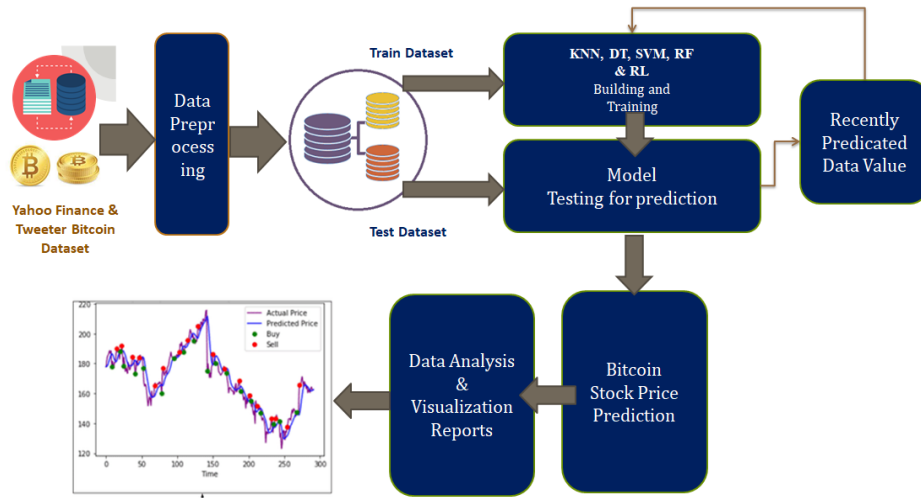


Fig 2 system architecture

## V.EXPERIMENTAL RESULTS

### Write Your Tweet

Enter Tweet

Predict

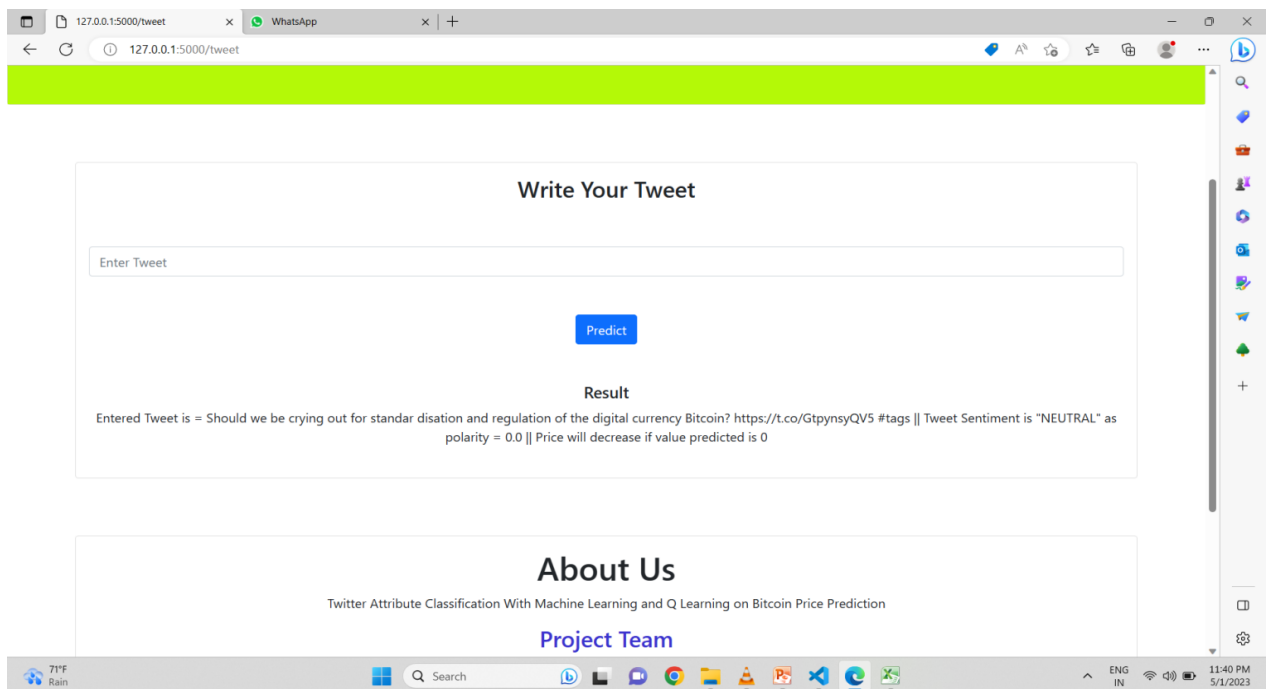
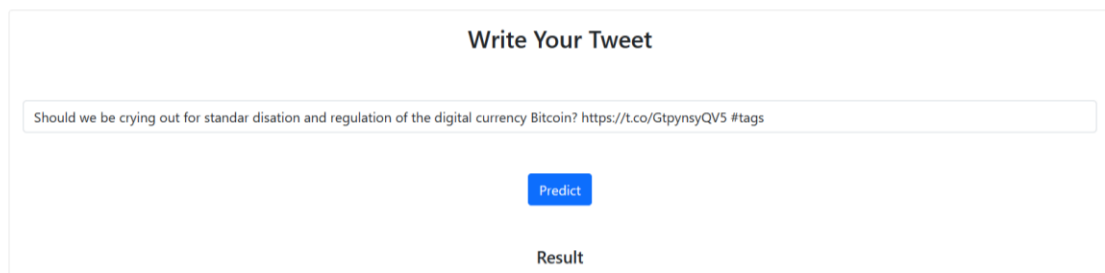
#### Result

Entered Tweet is = Whatever one believes about crypto (many don't yet see the potential/need) major US policy decisions should be made by legislatures and regulatory process, not unelected officials. Process exists in part to avoid throwing the baby out with the bathwater || Tweet Sentiment is "POSITIVE" as polarity = 0.0625 || Price will decrease if value predicted is = 0

### About Us

Twitter Attribute Classification With Machine Learning and Q Learning on Bitcoin Price Prediction

Project Team



Figures shows the results of analysis of bitcoin price prediction



## **VI. CONCLUSION**

The data is collected from Yahoo Finance online server. Another dataset on Bitcoin Tweets are scrapped from twitter by proving the Authentication using API keys. We applied various ML (Random Forest, Support Vector Machines, K – Nearest Neighbor and Decision Tree) Q-Learning Algorithms (Deep Q-Learning). We have compared the ML and QL models at the end based on the accuracy level. Finally, we have developed the Web Application, Flask based, which is working on the Real-time tweets from twitter website. The final model deployed in Q-Learning with 100% accuracy in training and 70% in testing the Data.

## **REFERENCES**

- [1] O. Sattarov, A. Muminov, C.W. Lee, H. K. Kang, R. Oh, J. Ahn, H. J. Oh, 877 and H. S. Jeon, "Recommending cryptocurrency trading points with deep reinforcement learning approach," Appl. Sci., vol. 10, no. 4, p. 1506, 879 Feb. 2020.
- [2] M. Mallikarjuna and R. P. Rao, "Evaluation of forecasting methods from selected stock market returns," Financial Innov., vol. 5, no. 1, pp. 1\_6, Dec. 2019.
- [3] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques\_Part II: Soft computing methods," Exp. Syst. Appl., vol. 36, no. 3, pp. 5932\_5941, Apr. 2009. 885
- [4] E. F. Fama, "The behavior of stock-market prices," J. Bus. vol. 38, no. 1, pp. 34\_105, 1965. [Online]. Available: <https://www.jstor.org/stable/2350752> 888
- [5] L. Kristoufek, "What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis," PLoS ONE, vol. 10, no. 4, Apr. 2015, Art. no. e0123923. 891
- [6] American Institute for Economic Research. Accessed: May 24, 2022. [Online]. Available: <https://www.aier.org/article/bitcoins-largest-price-changes-coincide-with-major-news-events-about-the-cryptocurrency/> 894
- [7] Twitter. Accessed: Jun. 20, 2022. [Online]. Available: <https://itzone.com>. 895 vn/en/article/elon-musk-tweets-alluding-to-break-up-with-bitcoin/ 89
- [8] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," ACM Trans. Web, vol. 1, no. 1, pp. 1\_10, 2007. 898
- [9] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proc. 7th Int. Conf. Lang. Resour. Eval. (LREC), vol. 10, May 2010, pp. 1320\_1326.
- [10] A. Nagar and M. Hahsler, "Using text and data mining techniques to extract stock market sentiment from live news streams," in Proc. Int. Conf. Comput. Technol. Science. vol. 47, 2012, pp. 91\_95. 920



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| Mobile No: +91-9940572462 | Whatsapp: +91-9940572462 | [ijarasem@gmail.com](mailto:ijarasem@gmail.com) |

[www.ijarasem.com](http://www.ijarasem.com)