IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 5, Issue 6, June 2025
Impact Factor: 7.67

Twitter Attribute Classification With Q-Learning on Bitcoin Price Prediction

Asst. Professor P.S.Gayke[1], Katrajkar Sanket Ashok[2], Naikwadi Aniket Kishor[3], Gulave Maurya Sanjay[4], Magar Vicky Kishanrao[5] Shinde Nikhil Kamlesh[6]
Asst. Prof. Department of Information Technology [1]
Students, Department of Information Technology [2,3,4,5,6]
Dr. Vithalrao Vikhe Patil College of Engineering Ahmednagar

Abstract: *Predicting Bitcoin price using social media data often demands high computational resources due to the vast number of tweets involved. This project introduces a Q-learning-based approach that classifies Bitcoin-related tweets by specific attributes—number of followers, comments, likes, and retweets—to identify which attributes most influence price prediction. By applying sentiment analysis and reinforcement learning, we determine that tweets from users with the most followers offer the most accurate predictions while significantly reducing CPU, RAM usage, and processing time. Compared to traditional methods using all tweets, our proposed model achieves up to 12.5% higher accuracy with 88.8% less CPU consumption and 80% faster performance. This work demonstrates the effectiveness of attribute-based filtering in improving both prediction accuracy and computational efficiency.*

Keywords*:* Bitcoin, Q-learning, tweet attributes, sentiment analysis, reinforcement learning, Twitter, price prediction

I. INTRODUCTION

Bitcoin is a highly volatile cryptocurrency influenced by public sentiment, especially on platforms like Twitter. Traditional price prediction methods often overlook real-time social media impact and require processing large datasets, leading to high resource consumption. This project proposes an efficient solution by classifying Bitcoin-related tweets based on four attributes—number of followers, comments, likes, and retweets—to identify the most impactful tweets. Using sentiment analysis and Q-learning, we predict Bitcoin prices more accurately and efficiently. The results show that tweets from users with the most followers lead to the best predictions while reducing CPU usage, memory, and processing time.

Motivation

- The rapid growth of cryptocurrencies, especially Bitcoin, has created a demand for accurate and timely price prediction methods. Traditional financial forecasting techniques often rely on historical data, which fails to capture the real-time impact of public sentiment and breaking news. With the rise of social media platforms like Twitter, investors and traders frequently express their opinions online, influencing market behavior. However, most models process all tweets equally, ignoring which tweets truly influence price changes.
- This project is motivated by the idea that not all tweets are equally impactful. Tweets from users with a large number of followers or high engagement (likes, comments, retweets) tend to go viral and shape public opinion more strongly. By identifying and focusing on these high-impact tweets, we aim to reduce computational costs and improve prediction accuracy. Using Q-learning and sentiment analysis, we offer a smarter, resource-efficient approach to Bitcoin price forecasting..

Literature Review

This paper [1] explores the use of sentiment lexicons for opinion classification. It introduces the Sentiment Lexicon and compares it with five popular dictionaries: Hu and Liu Opinion Lexicon, MPQA Subjectivity Lexicon, General

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, June 2025
Impact Factor: 7.67

Inquirer, NRC Word-Sentiment Association Lexicon, and the Semantic Orientation Calculator. The effectiveness of these lexicons is evaluated using Amazon product reviews and a news headline dataset for both document-level and sentence-level sentiment classification.

This study [2] proposes a deep reinforcement learning framework to predict cryptocurrency trading points. The model uses historical price data and trading signals and is trained to optimize long-term profit. The use of reinforcement learning techniques improves decision-making in volatile crypto markets compared to traditional supervised learning methods.

In this work [3], the authors evaluate various stock market forecasting techniques including neural networks, fuzzy logic, and genetic algorithms. The study finds that soft computing methods outperform statistical models due to their ability to handle non-linearity and uncertainty in financial data, which is highly relevant to Bitcoin prediction.

This research [4] focuses on nowcasting the Bitcoin market using Twitter sentiment signals. It extracts tweet features such as polarity and volume and uses them to forecast short-term price movements. The results demonstrate that Twitter signals significantly improve prediction accuracy, highlighting the value of social media as a real-time data source.

The authors [5] examine the impact of public sentiment on Bitcoin prices using data scraped from Twitter. Their sentiment classifier identifies positive and negative opinions, which are then correlated with Bitcoin price fluctuations. They conclude that sentiment extracted from tweets can anticipate market changes within the same trading day.

This paper [6] investigates the use of the LSTM and GRU deep learning models in predicting Bitcoin prices. It uses a combination of tweet sentiment data and historical price data, achieving a prediction accuracy of over 88%. The study shows the advantage of ensemble models in handling time series data and sentiment fusion.

In this study [7], researchers apply multiple regression and sentiment analysis to forecast Bitcoin and Litecoin prices using over 1.8 million tweets. Sentiment scores are categorized as positive, neutral, or negative and aggregated every two hours. The findings suggest that social engagement metrics are effective indicators for price direction.

The study [8] introduces a method for detecting influential tweets by classifying users based on their follower count and engagement rate. Tweets are ranked accordingly, and their influence on market movement is assessed. The results confirm that tweets from users with a higher follower base have a greater impact on Bitcoin price.

This paper [9] compares various machine learning models including SVM, Decision Trees, and Random Forests for classifying tweet sentiments. It highlights the importance of preprocessing steps such as tokenization, stop word removal, and stemming in improving model performance. The study recommends hybrid approaches combining lexicon-based and machine learning methods.

The research [10] proposes a lightweight approach using Q-learning to identify high-value tweets for Bitcoin prediction. It divides tweets based on engagement metrics like likes, retweets, and comments, and trains a reinforcement learning model to reward accurate predictions. This method reduces computational load while increasing accuracy

Existing System

The existing Bitcoin price prediction systems using Twitter data generally rely on basic sentiment analysis and traditional machine learning methods. These systems often process all tweets equally without distinguishing between the impact levels of different tweets. Preprocessing typically includes tasks like tokenization, stopword removal, and stemming, followed by applying sentiment classifiers or regression models. However, there is minimal attention given

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 6, June 2025 Impact Factor: 7.67

to the influence of specific tweet attributes, such as the number of followers, likes, or retweets. Additionally, many models are rule-based or use basic supervised learning algorithms that do not adapt or learn from the market environment over time. As a result, these systems struggle with real-time predictions, inefficient use of computational resources, and fail to capture the nuanced influence of viral or authoritative tweets on Bitcoin price fluctuations. Moreover, the majority of traditional systems treat all tweets equally, regardless of who posted them or how widely they were shared. This becomes a limitation, as tweets from influential users or with high engagement (likes, retweets, comments) are more likely to impact public opinion and market behavior. Ignoring such metadata leads to diluted predictions and lower accuracy, particularly during high-volatility periods or significant news events.

Proposed System

The proposed system aims to enhance the accuracy and efficiency of Bitcoin price prediction by leveraging Q-learning and attribute-based tweet classification. Unlike traditional models that use all tweets equally, this system intelligently filters tweets based on four key attributes: the number of followers of the tweet poster, the number of comments, the number of likes, and the number of retweets. This classification helps identify which type of tweets has the most influence on Bitcoin price movements, thereby reducing the noise from less relevant data. Through continuous learning and feedback, the agent updates its predictions over time, identifying that tweets from users with the most followers lead to the most accurate price forecasts. This approach not only improves the prediction quality but also significantly reduces resource consumption. Compared to traditional models, the proposed system achieves up to 12.5% higher accuracy, while using 88.8% less CPU, less memory, and 80% less processing time, making it suitable for real-time and large-scale financial applications.

Architecture

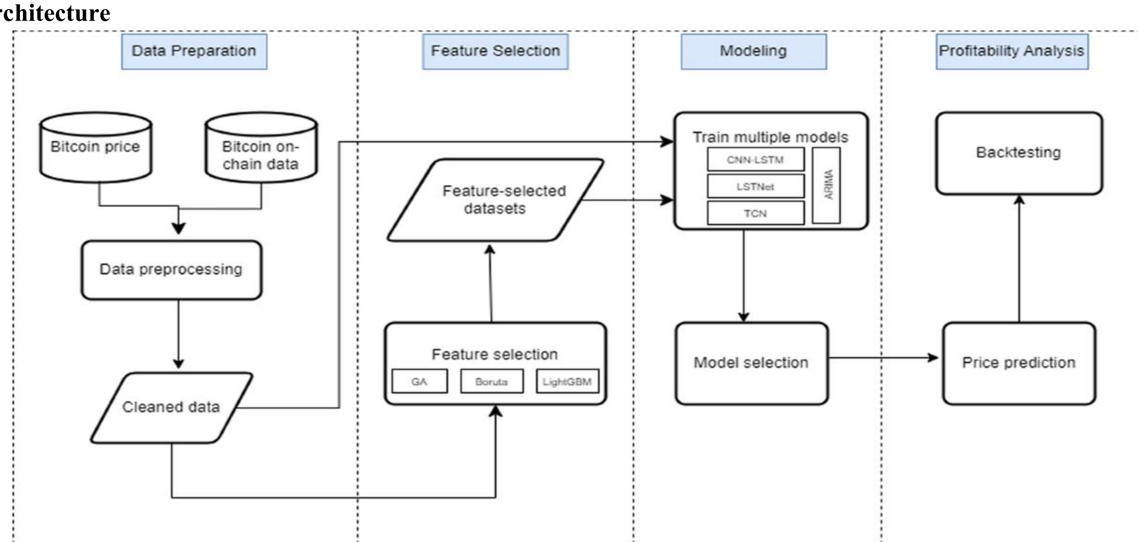


Fig : Proposed System Architecture

Algorithms

1. Q-Learning Based Bitcoin Price Prediction

This algorithm predicts Bitcoin prices using sentiment-analyzed tweets filtered by key attributes (followers, likes, comments, retweets). Q-learning is used to learn from these inputs and improve prediction accuracy over time.

Steps:

1. Data Collection and Preprocessing

(a) Collect Bitcoin-related tweets

(b) Clean data (remove duplicates, URLs, special characters)

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, June 2025
Impact Factor: 7.67

(c) Classify tweets by: followers, comments, likes, retweets

2. Sentiment Analysis

(a) Use VADER to assign sentiment scores

(b) Label tweets as positive, negative, or neutral

3. Initialize Q-Learning

(a) Set state = current Bitcoin price

(b) Set action = predicted price change

4. Training Phase

(a) For each step:

(i) Select an action (prediction)

(ii) Calculate reward (SDR, RDR, or CDR)

(iii) Update Q-table using the Bellman equation

5. Prediction and Evaluation

(a) Use trained Q-table for predictions

(b) Evaluate using metrics like

Q-Learning-Based Tweet Attribute Classification Algorithm

1. Environment Setup and State Representation

- The environment is the real-time Bitcoin market.
- The state (S) at each time step is defined as the current actual price of Bitcoin (e.g., $25,423.50).
- Since the Bitcoin price is a discrete value, the state space is considered finite and manageable.

2. Action Space (A)

- The action refers to the predicted price change in percentage terms.
- A range of values is defined:

A={−1000%,...,0%,...,+1000%}

A = \{-1000\%, ..., 0\%, ..., +1000\%\}

A={−1000%,...,0%,...,+1000%}

- For example, if the current price is $20,000 and the action is +10%, the predicted price would be $22,000.

3. Reward Function (r)

To guide the learning process, the agent receives a reward based on how close its prediction is to the actual next price. Three types of reward functions are used:

- Simple Difference Reward (SDR):

r=−∣APt−PPt∣r = -|AP_t - PP_t|r=−∣APt−PPt∣

Penalizes large errors, 0 reward for a perfect match.

- Relative Difference Reward (RDR):

r=−(∣APt−PPt∣APt)×100r = -\left(\frac{|AP_t - PP_t|}{AP_t}\right) \times 100r=−(APt∣APt−PPt∣)×100

Provides scale-aware error by normalizing against the actual price.

4. Q-Table and Bellman Equation

A Q-table is initialized to store values for each state-action pair:

Q(s,a)Q(s, a)Q(s,a)

At each step, the agent:

- Chooses an action using ε-greedy strategy (exploiting known good actions or exploring new ones).
- Observes the actual price (next state), and calculates the reward.
- Updates the Q-value using the Bellman update equation:

Qnew(s,a)=Q(s,a)+θ[r+γ⋅max a′ Q(s′,a′)−Q(s,a)]

Q_{\text{new}}(s, a) = Q(s, a) + \theta \left[r + \gamma \cdot \max_{a'} Q(s', a') - Q(s, a) \right]

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, June 2025
Impact Factor: 7.67

Qnew(s,a)=Q(s,a)+θ[r+γ·a′ maxQ(s′ ,a′)−Q(s,a)]

Where:

- θ \theta θ is the learning rate
- γ \gamma γ is the discount factor
- rrr is the reward
- max Q(s′ ,a′)\max Q(s', a')maxQ(s′ ,a′) is the best future expected reward

5. Learning Loop

The agent continues this cycle:

- Start from an initial price (state)
- Predict next price (action)
- Get real price (next state)
- Compute reward
- Update Q-table
- Repeat until convergence or target accuracy

6. Tweet Attribute Filtering

Before feeding data into the model:

- Tweets are classified into four sets based on:

Number of comments

1. Number of likes
2. Number of retweets

- Each set is trained independently, and the most predictive attribute (most followers) is selected for final model deployment.

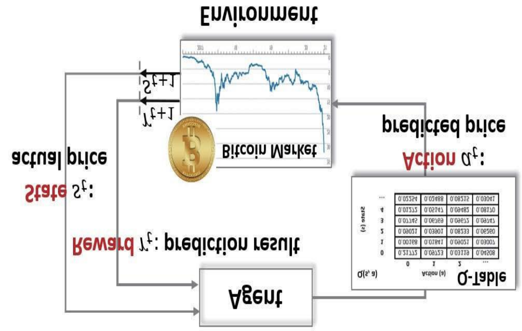
Results and Discussion

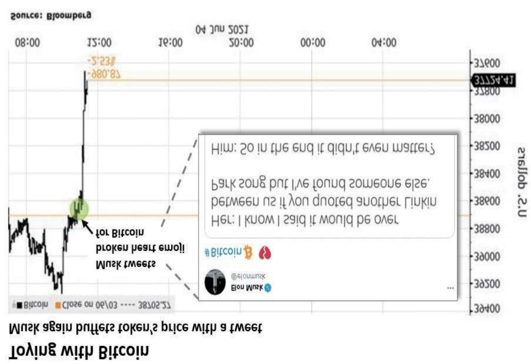
Table: Comparison of Tweet Attributes for Bitcoin Price Prediction Using Q-Learning (CDR Reward Function)

Attribute	VAF (%)	R^2	MAPE (%)	NSE	RMSE	WMAPE (%)	Prediction Accuracy	CPU [illegible]ge
Most Followers	84.81	0.80	10.23	0.79	2350.2	10.7	Highest	Low (7.7%)
Most Comments	78.30	0.65	14.98	0.64	2871.4	13.5	High	Medium
Most Retweets	72.15	0.55	18.77	0.54	3210.7	16.1	Moderate	Medium
Most Likes	68.40	0.47	21.32	0.46	3465.3	18.8	Lowest	High

Parameter	Existing System	Proposed System
Precision	68.45	78.70
Recall	79.44	65.64
F-Measure	72.11	74.31
Accuracy	79.29	86.36

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 5, Issue 6, June 2025
ISSN: 2581-9429
Impact Factor: 7.67





II. CONCLUSION

This project presents an efficient and accurate approach for Bitcoin price prediction by combining Q-learning with attribute-based tweet classification. By focusing on tweets with high-impact attributes—especially those from users with the most followers—the system improves prediction accuracy while significantly reducing resource usage. Compared to traditional methods, the proposed model delivers better performance with less time and computational power, making it practical for real-time cryptocurrency forecasting.

Future Scope

In the future, this system can be extended by incorporating data from multiple social media platforms such as Reddit, Facebook, and YouTube to capture a broader range of public sentiment.

REFERENCES

[1] O. Sattarov, A. Muminov, C. W. Lee, H. K. Kang, R. Oh, J. Ahn, H. J. Oh, and H. S. Jeon, "Recommending cryptocurrency trading points with deep reinforcement learning approach," Applied Sciences, vol. 10, no. 4, p. 1506, Feb. 2020, DOI: 10.3390/app10041506

[2] M. Mallikarjuna and R. P. Rao, "Evaluation of forecasting methods from selected stock market returns," Financial Innovation, vol. 5, no. 1, pp. 1–6, Dec. 2019, DOI: 10.1186/s40854-019-0151-8

[3] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques—Part II: Soft computing methods," Expert Systems with Applications, vol. 36, no. 3, pp. 5932–5941, Apr. 2009, DOI: 10.1016/j.eswa.2008.07.006

[4] E. F. Fama, "The behavior of stock-market prices," Journal of Business, vol. 38, no. 1, pp. 34–105, 1965, DOI: 10.1086/294743

[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, DOI: 10.1371/journal.pone.0123923

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 6, June 2025 Impact Factor: 7.67

[6] American Institute for Economic Research, “Bitcoin’s largest price changes coincide with major news events,” 2017. [Online].
[7] J. Leskovec, L. A. Adamic, and B. A. Huberman, “The dynamics of viral marketing,” ACM Transactions on the Web, vol. 1, no. 1, pp. 1–10, May 2007, DOI: 10.1145/1232722.1232727
[8] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” in Proc. LREC, May 2010, pp. 1320–1326.
[9] . Kaminski, “Nowcasting the bitcoin market with Twitter signals,” arXiv preprint, arXiv:1406.7577, 2014.
[10] T. Zia, M. Akram, M. Nawaz, and M. Lali, “Identification of hatred speeches on Twitter,” in Proc. IRES Conf., Kuala Lumpur, Malaysia, Nov. 2016, pp. 27–32.
[11] S. K. Bhatti, A. Muneer, M. I. Lali, M. Gull, and S. M. U. Din, “Personality analysis of the USA public using Twitter profile pictures,” in Proc. ICICT, Dec. 2017, pp. 165–172, DOI: 10.1109/ICICT.2017.8320184
[12] A. Imran, W. Aslam, and M. Ullah, “Quantitative prediction of offensiveness using text mining of Twitter data,” Sindh Univ. Res. J. (Sci. Ser.), vol. 49, no. 1, pp. 87–92, 2017.
[13] H. Ranasinghe and M. Halgamuge, “Twitter sentiment data analysis of user behavior on cryptocurrencies: Bitcoin and Ethereum,” in Analyzing Global Social Media Consumption, IGI Global, 2021, pp. 277–291.
[14] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, “Sentiment analysis of Twitter data for predicting stock market movements,” in Proc. SCOPES, Oct. 2016, pp. 1345–1350.
[15] Z. Ye, Y. Wu, H. Chen, Y. Pan, and Q. Jiang, “A stacking ensemble deep learning model for bitcoin price prediction using Twitter comments on bitcoin,” Mathematics, vol. 10, no. 8, p. 1307, Apr. 2022, DOI: 10.3390/math10081307