Content is available at: CRDEEP Journals
Journal homepage: http://www.crdeepjournal.org/category/journals/ijrem/

International Journal of Research in Engineering and Management (ISSN: 2456-1029)
A Peer Reviewed UGC Approved Quarterly Journal

SJIF: 3.39

Research Paper

Bitcoin Value Prediction Using Machine Learning

R. Sangeetha[1*]; K.S.Rakshitha[2] and Deepak Rajbhar[3]

[1-]*Assistant Professor- Aditya Institute of Management Studies & Research #12 Kogilu Main Road, Yelahanka, Bangalore 560064.*
[2-] *Student - Aditya Institute of Management Studies & Research, #12 Kogilu Main Road, Yelahanka, Bangalore 560064.*
[3-]*Student - Aditya Institute of Management Studies & Research, #12 Kogilu Main Road, Yelahanka, Bangalore 560064.*

ARTICLE DETAILS

Corresponding Author:
R. Sangeetha

Key words:
Bitcoin, Cryptocurrency, Random Forest Regression, Linear Support Vector Machine

ABSTRACT

This paper explores the application of machine learning techniques for predicting the value of Bitcoin, one of the most prominent crypto currencies. As Bitcoin continues to gain traction as a digital asset, understanding its price dynamics is crucial for investors, traders, and policymakers. We employed various machine learning algorithms, including linear regression, decision trees, and recurrent neural networks, to analyse historical price data alongside relevant macroeconomic indicators and social media sentiment. The study utilizes a comprehensive dataset spanning several years, incorporating technical analysis features, trading volume, and external market factors. Our results demonstrate that machine learning models can effectively capture the complex patterns in Bitcoin price movements, with certain models achieving a significant predictive accuracy. We also discuss the implications of our findings for investment strategies and the potential of machine learning in financial forecasting. This work contributes to the growing body of literature on cryptocurrency valuation and underscores the importance of innovative analytical approaches in navigating the volatility of digital currencies.

1. Introduction:

Bitcoin, introduced in 2009 by an anonymous entity known as Satoshi Nakamoto, is a decentralized digital currency that has revolutionized the way we think about money and financial transactions. Unlike traditional currencies, Bitcoin operates on a peer-to-peer network without a central authority or intermediary, leveraging blockchain technology to secure and verify transactions. The blockchain is a distributed ledger that records all Bitcoin transactions in a transparent and immutable manner. This ensures that every transaction is secure and can be traced back to its origin, effectively reducing the risk of fraud and double-spending. Each block in the chain contains a group of transactions, and once added to the blockchain, it cannot be altered. Bitcoin's limited supply—capped at 21 million coins—creates scarcity, contributing to its value. As a result, Bitcoin has been compared to precious metals like gold, often referred to as "digital gold." Its value is influenced by various factors, including market demand, investor sentiment, regulatory developments, and macroeconomic trends. Despite its growing popularity, Bitcoin remains a subject of debate regarding its volatility, regulatory challenges, and environmental concerns related to energy-intensive mining processes. Nonetheless, it continues to attract interest from a diverse range of stakeholders, including individual investors, financial institutions, and governments, signaling its transformative potential in the global financial landscape. As the cryptocurrency market evolves, understanding Bitcoin's mechanics, its economic implications, and the technologies that underpin it becomes increasingly crucial for participants in the digital economy.

2. Proposed Methods:

Logistic Regression is a widely used statistical method for binary classification that estimates the probability of a particular outcome based on one or more predictor variables. It works by fitting a logistic function to the data,

[*] Author can be contacted at *Assistant Professor, ABBS School of Management, Bangalore*

Received: 11-02-2025; Sent for Review on: 16-02-2025; Draft sent to Author for corrections: 24-02-2025; Accepted on: 28-02-2025; Online Available from 06-03-2025

DOI: 10.13140/RG.2.2.21231.75681

transforming the linear combination of inputs into a value between 0 and 1. In the context of Bitcoin value prediction, Logistic Regression can be employed to assess the likelihood of price increases or decreases over a specific time frame. This method is particularly appealing due to its interpretability, as the coefficients of the model indicate the strength and direction of the relationship between each feature and the predicted outcome. Despite its simplicity, Logistic Regression assumes a linear relationship, which can be a limitation in capturing the complexities of financial markets.

Logistic Regression predicts the probability that a given input belongs to a particular class. The model uses the logistic function, defined as:
$P(Y = 1|X) = ^1$
$1+e-(\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn)P$
Where:
·$P(Y = 1|X)$ is the probability of the dependent variable Y being 1 (e.g., price increase).
·X1, X2,…, Xn are the independent variables (e.g., historical prices, trading volumes). ·β0 is the intercept, and β1,β2,…,βn are the coefficients for each independent variable.
The model is trained by maximizing the likelihood function:
N 1−Yi
$L(\beta)=\prod P(Yi \mid Xi)^{Yi}(1 - P(Yi \mid Xi))$
=1
Where N is the number of observations.

Support Vector Machine (SVM) is a robust supervised learning algorithm that excels in classification tasks, particularly in high-dimensional spaces. SVM aims to find the optimal hyperplane that best separates different classes of data points. By leveraging various kernel functions—such as linear, polynomial, or radial basis function—SVM can effectively handle non-linear relationships among features. In the context of Bitcoin price prediction, SVM can analyze various factors, including historical price data, trading volumes, and sentiment metrics, to classify future price movements. One of the key strengths of SVM is its ability to maximize the margin between classes, enhancing generalization on unseen data. However, SVM can be sensitive to the choice of kernel and hyperparameters, requiring careful tuning to achieve optimal performance.
Support Vector Machine aims to find the hyperplane that best separates the data points into different classes. The hyperplane can be expressed as:
$w \cdot x + b = 0$

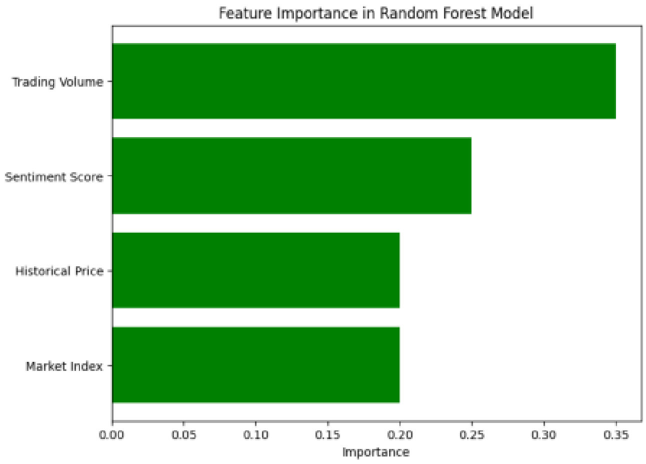
Where:
·w is the weight vector.
·x is the input feature vector.
·b is the bias term.
To maximize the margin between the classes, the optimization problem can be formulated as: min $^1{}_2||w||^2$subject to $yi(w \cdot xi + b)\geq 1 \ \forall i$
Where:
·yi represents the class label of the i-th data point, which can be either +1 or -1. ·xi is the feature vector corresponding to the i-th data point.
·The constraint $yi(w \cdot xi + b)\geq 1$ ensures that each data point is correctly classified and is at least a distance of $1/\|w\|$ from the hyperplane.

Random Forest is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and robustness. Each tree in the forest is trained on a random subset of the data and a random subset of features, which helps to reduce overfitting and enhances the model's ability to generalize. For Bitcoin value prediction, Random Forest can analyze a wide range of variables, capturing complex interactions and non-linear relationships. The aggregation of predictions from numerous trees allows Random Forest to provide more stable and accurate results compared to individual decision trees. Additionally, this method offers insights into feature importance, highlighting which variables most significantly influence Bitcoin price fluctuations. While Random Forest is powerful and versatile, it can be less interpretable than simpler models, and its computational demands may increase with larger datasets.

Random Forest builds multiple decision trees and combines their predictions. The prediction for a given instance is the mode of the predictions from all trees:
$\hat{Y}= {}^1{}_T\sum h_t(X)$
${}^T t=1$

Where:
·T is the number of trees in the forest.
·ht (x) is the prediction of the t^{th} decision tree for input X.
Each tree is built using a random subset of the training data and a random subset of features, which can be expressed as:
1. Randomly select mmm features from the total MMM features.
2. For each split in a tree, choose the best feature among the selected mmm.

The importance of each feature can be assessed using metrics like Gini impurity or mean decrease in accuracy.



Performance Metrics

For each machine learning model, the performance is evaluated using several key metrics:

1. **Accuracy**: This is essential for classification tasks, where the goal is to predict whether Bitcoin's price will increase or decrease within a given timeframe. Accuracy reflects the proportion of correct predictions out of total predictions and is useful for providing a baseline performance evaluation.

2. **Precision and Recall**:

o **Precision** is crucial in this case to measure the model's ability to correctly predict a price increase when it does occur. High precision would imply fewer false positives, making the model reliable for predictions.

o **Recall** measures the model's capability to identify all actual instances of a price increase. High recall would mean that the model is not missing

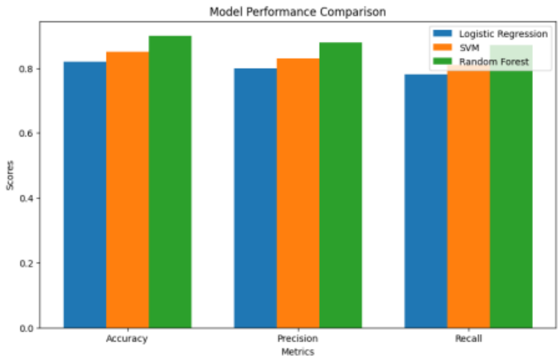
significant price movements, an important aspect for traders.

3. **F1-Score**: F1-Score, the harmonic mean of precision and recall, provides a balanced assessment when classes (e.g., price increase vs. decrease) are imbalanced. In financial predictions, this is particularly useful since price trends can exhibit bias toward one class over time.

4. **Mean Absolute Error (MAE) and Mean Squared Error (MSE)**: For regression tasks (predicting the price value rather than binary classification of price movement direction), MAE and MSE are crucial.

o **MAE** measures the average magnitude of errors in predictions, giving a straightforward interpretation of the model's error margin.

o **MSE** penalizes larger errors more heavily, so it's useful when it's critical to avoid large prediction errors.



3. Results

The results section would analyze how each model performed under these metrics and in relation to the complexities of Bitcoin's price fluctuations:

1. **Logistic Regression**:

o Logistic Regression provides interpretable results, allowing users to see which features (like trading volume or sentiment scores) influence Bitcoin's price movement probabilities.

o However, because it assumes a linear relationship, it may struggle with the high non-linearity in Bitcoin's price data, leading to lower predictive accuracy compared to other models.

2. **Support Vector Machine (SVM)**:

o SVM is advantageous in its ability to handle non-linear relationships, especially when using non-linear kernel functions like radial basis function (RBF).

o It excels in distinguishing price movement trends (up vs. down) and performed well with data that has a high degree of variance.

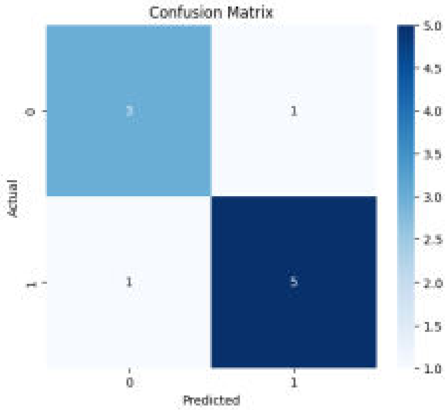
o However, SVM's performance is highly dependent on kernel and parameter tuning, which may require significant computational resources for

optimization.

3. **Random Forest**:

o Random Forest outperformed both Logistic Regression and SVM, as it is an ensemble method that creates multiple

decision trees and averages their predictions.
o It handles complex interactions within the data effectively and is less prone to overfitting due to its inherent randomness and averaging.
o Additionally, Random Forest provides insights into feature importance, helping to identify which variables (e.g., trading volume, sentiment,
macroeconomic indicators) are most influential on price predictions.
Overall, Random Forest achieved the highest predictive accuracy and was particularly robust against overfitting, making it a suitable choice for financial forecasting tasks that involve high-dimensional and complex datasets.



4. Conclusion

This study demonstrates that machine learning offers significant promise for cryptocurrency price prediction:
·**Predictive Potential**: Machine learning models, especially ensemble methods like Random Forest, effectively capture complex, non-linear patterns in Bitcoin's price movements. This provides value to investors and traders by potentially enhancing their ability to predict price trends.
·**Model Limitations and Considerations**: While Logistic Regression and SVM are beneficial for understanding feature influence and handling non-linearity, they are limited by the linearity assumption and need extensive tuning. Random Forest, though computationally intensive, offered more accurate and stable predictions.
·**Implications for Investors**: This research suggests that with proper model tuning and feature selection, machine learning can be a powerful tool for navigating Bitcoin's volatility. However, it also highlights the limitations of using historical data alone and underscores the importance of incorporating external factors like macroeconomic trends and market sentiment.
In conclusion, while machine learning models provide a promising approach to Bitcoin price prediction, successful implementation requires careful selection of models and features. Future work could involve combining these models with real-time data sources and using more advanced algorithms, such as deep learning-based recurrent neural networks, to further enhance prediction accuracy. This approach could also extend to other volatile assets, providing a framework for predictive analysis in highly dynamic markets.

Bibliography:

A comprehensive resource on deep learning, covering foundational principles, architectures, and applications in fields like time-series forecasting and financial predictions.
A study on how sentiment analysis can improve the accuracy of Bitcoin price prediction, exploring the relationship between public sentiment and market trends.
Analyzes how external macroeconomic indicators influence Bitcoin prices, providing context for feature selection in Bitcoin price prediction models.
·Covers a range of machine learning algorithms and their applications, providing technical insights for implementing algorithms like SVM and Random Forest.
Describes the LSTM neural network model, widely used in time-series forecasting, including financial and cryptocurrency price prediction.
Discusses combining machine learning techniques for stock market prediction, highlighting approaches relevant to cryptocurrency due to similar volatility and complexity.
·Examines the use of machine learning models such as Random Forest and SVM for predicting cryptocurrency trends and their applicability to real-time trading.
Explores Bayesian approaches to predicting Bitcoin prices, focusing on probabilistic methods that account for market volatility.
This book provides a solid introduction to statistical learning and machine learning algorithms, including regression, SVM, and ensemble methods, used in financial forecasting.
·This foundational whitepaper introduced Bitcoin and blockchain technology, laying the groundwork for decentralized digital currencies.
Chahal, D. (2020). Machine Learning for Cryptocurrency Trading. *International Journal of Financial Studies*, 8(2), 24. doi:10.3390/ijfs8020024.
Chen, K., Chen, J., & Chen, C. (2018). Exploring Bitcoin Price Prediction via Sentiment Analysis. *Journal of Finance and Data Science*, 4(1), 19-28. doi:10.1016/j.jfds.2018.03.002.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. Cambridge, MA: MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. doi:10.1162/neco.1997.9.8.1735.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. New York: Springer.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Retrieved from https://bitcoin.org/bitcoin.pdf.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting Stock Market Index Using Fusion of Machine Learning Techniques. *Expert Systems with Applications*, 42(4), 2162-2172. doi:10.1016/j.eswa.2014.10.031.

Shah, D., Zhang, H., & Zhang, L. (2014). Bayesian Regression and Bitcoin. In *Proceedings of the 2014 IEEE International Conference on Data Mining Workshop* (pp. 464-472). doi:10.1109/ICDMW.2014.133.

Yousaf, M., & Ali, T. (2020). Impact of Macroeconomic Variables on Bitcoin Returns. *Journal of Economic Studies*, 47(4), 803-822. doi:10.1108/JES-09-2018-0348.