INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue VIII August 2025

Development of Bitcoin Closing Price Prediction Model Using Machine Learning

[1]Tsentob Joy Samson., [1]Ndubisi Dillion Ayebamieprete., [1]Stephen Ebuka Iheagwara, [2]Ashraf Ishaq

[1]Department of Computer Science, Faculty of Computing, Air Force Institute of Technology, Kaduna, Nigeria

[2]Department of Computer Science, Federal University Wukari, Nigeria

DOI: https://doi.org/10.51584/IJRIAS.2025.100800047

Received: 01 August 2025; Accepted: 06 August 2025; Published: 04 September 2025

ABSTRACT

Bitcoin's rise as a decentralized cryptocurrency, powered by blockchain, has transformed financial markets by enabling intermediary-free global transactions. Its price volatility, driven by market sentiment, regulatory shifts, and economic trends, challenges traders and investors. This study predicts Bitcoin's daily closing price trends (up or down) using an ensemble machine learning model combining logistic regression and XGBoost. Using historical price data and technical indicators (SMA_50, SMA_200, RSI) from the model employs hyperparameter tuning, correlation analysis, and RFE for feature selection. Time Series Split cross-validation ensures temporal integrity. The ensemble achieves 90.4% test set accuracy, with precision, recall, and F1-scores above 0.89, outperforming baseline models. Back testing aligns with February 2025 market shifts. This research highlights ensemble learning's efficacy in volatile cryptocurrency markets, offering traders a robust tool.

Keywords: Bitcoin Price Prediction, Machine Learning, Logistic Regression, XGBoost, Cryptocurrency Forecasting, Technical Indicators, Ensemble Model

INTRODUCTION

Blockchain, a distributed ledger technology (DLT), enables cryptocurrencies by maintaining a shared, irreversible record of transactions across a computer network, ensuring no alterations can be made (Sugurushetty, 2024). Cryptocurrency, a decentralized digital cash system, represents encrypted data as a unit of money, serving as a medium of exchange and virtual accounting system (Shadiya & Chitra, 2024). Bitcoin, introduced in 2008 by Satoshi Nakamoto, is the first decentralized cryptocurrency, transforming finance through intermediary-free global transactions (Mittal & Geetha, 2022). Its price, however, exhibits significant volatility, rising from $0.0008 in 2009 to $105,000 in early 2025, driven by market sentiment, regulatory changes, and economic trends, creating opportunities and risks for investors (Nakamoto, 2009; Fauzi & Paiman, 2020). This volatility, coupled with the complexities of decentralized markets, challenges traditional financial models, necessitating advanced predictive approaches (Fauzi & Paiman, 2020).

Bitcoin's Proof of Work (PoW) consensus mechanism, which demands substantial energy for mining, raises environmental and efficiency concerns, while regulatory ambiguity and security vulnerabilities, such as 51% attacks and hacking risks, complicate adoption (Fauzi & Paiman, 2020; Jaiswal & Chaudhari, 2023). Machine learning (ML), a subset of AI, offers solutions by leveraging supervised learning (e.g., regression, neural networks) for forecasting, unsupervised learning for pattern identification, and reinforcement learning for trading strategy optimization (Perepu, 2024). This study employs an ensemble model combining logistic regression and XGBoost to predict Bitcoin's daily closing price trend (up or down), aiming to simplify decision-making for traders and investors. The objectives are to develop the model, evaluate its performance using metrics like accuracy and F1-score, and compare it to state-of-the-art approaches. By applying ML to financial forecasting, this research enhances data science and blockchain technology, fostering innovation in the cryptocurrency ecosystem.

The study focuses on short-term Bitcoin price trend predictions using historical price data and technical indicators, acknowledging limitations. Logistic regression, while computationally efficient, may struggle with complex market dynamics compared to deep learning models (Pabuçcu et al., 2022). External factors like news and regulations may impact accuracy, and the model is tailored specifically for Bitcoin, avoiding long-term forecasts due to their complexity. The choice of logistic regression and XGBoost balances predictive power with practical implementation, addressing the need for effective ML solutions in cryptocurrency price prediction.

REVIEW OF RELATED LITERATURE

Hafid et al. (2022) utilized machine learning with technical indicators like moving averages to predict Bitcoin prices, developing a framework that identifies market reversals and provides insights for traders. Lee et al. (2024) proposed attention-based deep learning models, Attention-LSTM and Attention-GRU, using indicators such as SMA, EMA, TEMA, and MACD to classify Bitcoin market trends. Attention-GRU was effective for real-time predictions, while Attention-LSTM excelled in long-term trend analysis. Jung et al. (2024) incorporated macroeconomic and blockchain-specific variables, such as transaction volume, into a time-series model for enhanced Bitcoin price prediction.

Andi et al. (2021) combined logistic regression with LSTM to address Bitcoin's volatility, capturing both statistical and sequential patterns for improved accuracy. Qureshi et al. (2024) advanced this by integrating multiple machine learning and time-series models into hybrid ensembles, achieving superior performance in metrics like MAE and RMSE. Shantsila et al. (2022) evaluated regression methods, including linear, polynomial, and support vector regression, using market volume and volatility metrics. Pichaiyuth et al. (2023) applied SHAP for feature selection, comparing SVM, KNN, RFC, and Naïve Bayes, finding SVM suitable for short-term predictions and LSTM for longer forecasts. Dimitriadou and Gregoriou (2023) analyzed weekly Bitcoin forecasts with Logistic Regression, SVM, and Random Forest, incorporating 24 financial and macroeconomic variables and emphasizing evaluation through accuracy, F1-score, and recall.

Zhou (2022) tested Linear Regression, SVM, Random Forest, and LSTM, with LSTM achieving 58% accuracy but modest overall predictive power. Bhuvaneswari et al. (2024) combined RNN and ARIMA with minute-level data, outperforming standalone models in MAE, RMSE, and MAPE. Hajare et al. (2025) employed deep learning models (LSTM, GRU, CNN) with high-frequency data and sentiment inputs, achieving an F1-score of 0.82. Fernandes et al. (2021) found ANN outperformed Linear Regression and SVM for data from 2017 to 2020, though limited by real-time data availability. Das et al. (2021) developed a Bi-LSTM and Bi-RNN hybrid, excelling in capturing temporal dependencies. Junwei Chen (2023) confirmed LSTM's effectiveness in binary classification, while Akyildirim et al. (2023) highlighted Random Forest's strength in predicting Bitcoin futures' mid-price movements. Nabipour et al. (2020) tested CNN-LSTM and GRU, with GRU performing best for Bitcoin, though they noted challenges in model interpretability and feature selection.

Research Gap

Despite notable advancements in the field of Bitcoin price prediction using machine learning, significant research gaps remain. Many studies have employed either traditional models or individual deep learning approaches without combining their strengths. For instance, while Hafid et al. (2022) successfully integrated technical indicators with machine learning, their approach still lacked consideration of broader datasets such as social sentiment or blockchain-level data. Similarly, Lee et al. (2024) effectively utilized attention-based deep learning models but limited their analysis to a few technical indicators and neglected macroeconomic or social media factors.

Furthermore, Qureshi et al. (2024) focused on short-term datasets, which restrict model generalizability over longer periods or diverse market conditions. Some works, such as Zhou (2022), achieved moderate accuracy yet relied solely on historical price and network data, omitting more nuanced factors like regulatory developments or user behavior from on-chain data. Another persistent gap lies in evaluation methodology. While models are frequently assessed using accuracy, few studies apply a comprehensive set of metrics like precision, recall, and F1-score, which are critical in imbalanced datasets often found in financial forecasting.

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue VIII August 2025

Studies like those by Hajare et al. (2025) and Fernandes et al. (2021) highlight this need by incorporating richer evaluation criteria but still fall short in model optimization and feature diversity.

The lack of generalization across different cryptocurrencies is another underexplored area. Most studies concentrate exclusively on Bitcoin, ignoring the potential transferability of predictive frameworks to altcoins. In addition, while hybrid models like the one proposed by Bhuvaneswari et al. (2024) offer improved performance, few studies compare such architectures systematically or address scalability challenges for real-time applications.

This study addresses these limitations by employing an ensemble model, integrating logistic regression with XGBoost as the base model, supported by hyperparameter tuning and rigorous feature selection. Evaluation metrics go beyond accuracy to include F1-score, recall, and precision, ensuring a more robust performance assessment. The approach also leverages recent daily data and technical indicators selected using Recursive Feature Elimination (RFE), enhancing predictive efficiency and practical utility in volatile markets.

Key Concepts

Bitcoin Market Dynamics

Bitcoin's pricing behavior, liquidity, volatility, and general market characteristics are influenced by a unique set of variables arising from its decentralized nature. Unlike traditional financial markets, the Bitcoin market operates without central oversight, making it especially responsive to factors such as supply-demand fluctuations, changes in regulatory policies, investor sentiment, and rapid technological advancements. Macroeconomic conditions, global institutional interest, and evolving blockchain infrastructure further shape Bitcoin's market behavior. As external economic dynamics shift, Bitcoin's perceived role alternates between that of a speculative asset, a store of value, and a hedge against inflation. These transitions contribute to the ongoing complexity and unpredictability of Bitcoin's market trajectory (Chen, 2023).

Logistic Regression and Sigmoid Function

Logistic regression is a widely used machine learning algorithm designed for binary classification tasks. It extends linear regression by introducing the sigmoid (or logistic) function, which transforms any real-valued input into a probability value between 0 and 1. This transformation makes logistic regression suitable for classifying outcomes into two distinct categories. Rather than predicting continuous numerical values, the model estimates the likelihood that an input belongs to a particular class. Typically, a threshold of 0.5 is applied: probabilities equal to or above this value are classified as one class (e.g., 1), while those below are assigned to the other class (e.g., 0) (Maalouf, 2011).

The logistic function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \qquad (1)$$

Where:

$\sigma(z)$ is the sigmoid function, which transforms any real number into a probability between 0 & 1.
z is the linear combination of input features, defined as:

$$z = b + w_1 x_1 + \cdots + w_n x_n \qquad (2)$$

Where:

Z is the output of the linear equation, b as the bias, was weights, and x as feature values.

Properties of the Sigmoid Function:

When $z \to \infty,\ \sigma(z) \to 1$ (high probability of 1).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue VIII August 2025

When $z \rightarrow -\infty,\ \sigma(z) \rightarrow 0$ (high probability of 0).

When $z = 0,\ \sigma(0) = 0.5$ (uncertain/ unknown classification).

To classify an input, we use a decision threshold (typically 0.5):

If $P(Y=1 \mid X) \geq 0.5$, classify as 1.

If $P(Y=1 \mid X) < 0.5$, classify as 0.

Mathematically, this is expressed as:

$\hat{y} = (1$, if $\sigma(z) > 0.5$

0, if $\sigma(z) < 0.5$) (3)

These characteristics make logistic regression a foundational method in predictive modeling, especially when outcomes are binary in nature (Google, 2025).

Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is an advanced and scalable machine learning algorithm used for both classification and regression problems. It is based on the gradient boosting framework but incorporates several enhancements such as L1 and L2 regularization, support for parallel computation, and intelligent handling of missing values. XGBoost operates by sequentially building decision trees where each new tree attempts to correct the residual errors of the previous ones.

The objective function is given by:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{k=1}^{K} \Omega(f_k) L(\theta) \qquad (4)$$

Where:

$l(y_i,\ \hat{y_i})$ is the loss function that quantifies the prediction error. For regression tasks, this is often the mean squared error:

$$l(y_i, \hat{y_i}) = (y_i - \hat{y_i})^2 \qquad (5)$$

And for classification tasks, the log loss function is used:

$$l(y_i, \hat{y_i}) = -[y_i \log(\hat{y_i}) + (1 - y_i) \log(1 - \hat{y_i})] \qquad (6)$$

$\Omega(f_k)\Omega$ represents the regularization term, which penalizes model complexity and helps prevent overfitting by controlling the structure and number of decision trees used.

It utilizes a second-order Taylor approximation for loss function optimization, and its regularization mechanisms help reduce overfitting. Additionally, XGBoost supports tree pruning and efficient memory usage, making it particularly well-suited for high-dimensional and large-scale datasets. Its robustness, speed, and flexibility have made it a widely adopted tool for predictive modeling across various domains.

Research Approach

This study is quantitative research focused on numerical data from Bitcoin's historical prices and technical indicators. The data was analyzed using logistic regression and XGBoost to predict the price trend. Regression analysis was employed to model the relationship between Bitcoin's price trends (dependent variable) and multiple independent variables such as historical price data and technical indicators (e.g., Moving Average, RSI).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue VIII August 2025

Sampling Techniques and Feature Selection: Bitcoin's historical data is treated as time series data, ensuring that observations are sequential and that past values influence future predictions. Commonly used technical indicators such as the Relative Strength Index (RSI) and trading volumes were included as features for the logistic regression model. Feature engineering was applied to compute indicators like SMA_50 and SMA_200 (50-day and 200-day simple moving averages), along with RSI, a momentum oscillator calculated over a 14-day window. The dependent variable was derived from the closing prices on a daily basis, indicating either an upward or downward movement. Statistical methods like correlation analysis and Recursive Feature Elimination (RFE) were used to determine feature importance and ensure that only relevant features were retained for model training.

Data Analysis

An ensemble of logistic regression and XGBoost models was used to predict Bitcoin's price trend as a binary outcome (up or down). Both models were trained on historical price data and technical indicators. The target variable was defined as a binary classification task: whether the next day's closing price is higher than the current day's (up = 1, down = 0).

To validate the models, the dataset was split into training (80%) and testing (20%) subsets. TimeSeries Split with four splits was applied to preserve chronological order. The logistic regression model was fitted to the training data, while hyperparameters for the XGBoost model were tuned using grid search or random search over a loop of up to 1000 iterations. The tuning explored variations in the number of estimators ({100, 200, 300, 400, 500}), learning rate (uniformly sampled between 0.01 and 1), and maximum tree depth ({3, 5, 7, 8, 10}). Additionally, K-fold cross-validation was implemented to further fine-tune the model and prevent overfitting.

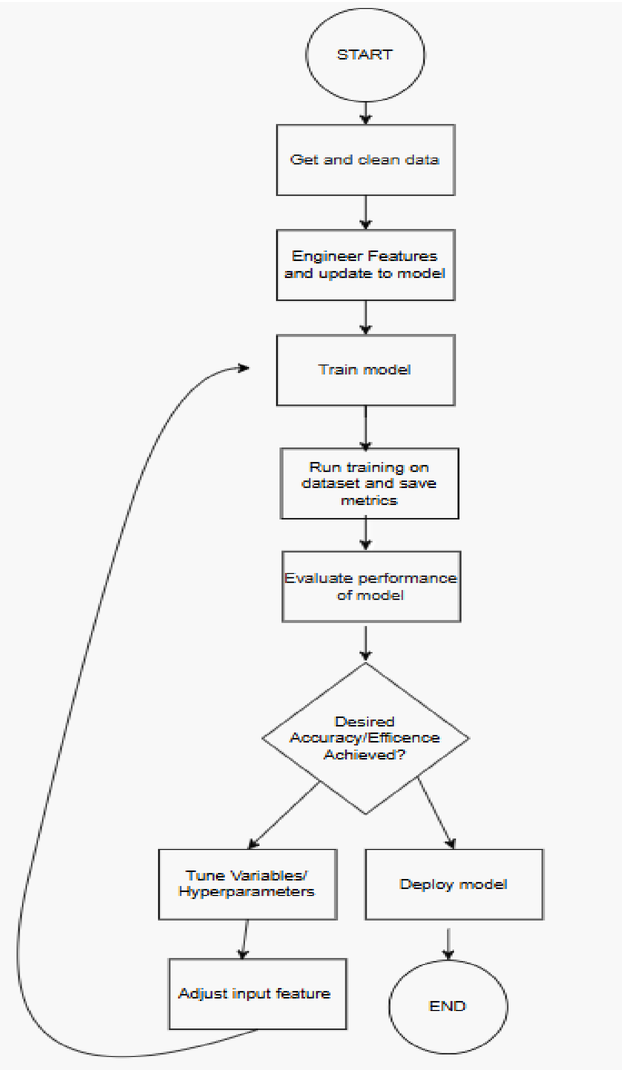

Figure 4.1: Flow chart of research approach

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue VIII August 2025

Primary Data Collection

Bitcoin Price Data: Historical Bitcoin price data (daily closing prices) was obtained directly from Yahoo Finance using the yfinance API. The data spans from 01 May 2024 to 04 March 2025, with a time interval of 1 day. This interval was chosen to capture medium-term market trends and maintain model stability.

Data Cleaning: Data was preprocessed by handling missing values, normalizing values, and ensuring that the features are on the same scale to improve the model's performance.

Model Evaluation Metrices

Accuracy: The proportion of correct predictions (true positives and true negatives) out of all predictions.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (7)$$

Precision: The proportion of true positive predictions out of all positive predictions made by the model.

$$\text{Precision} = \frac{TP}{(TP+FP)} \qquad (8)$$

Recall: The proportion of true positive predictions out of all actual positive cases.

$$\text{Recall} = \frac{TP}{(TP+FN)} \qquad (9)$$

F1-Score: The harmonic mean of precision and recall, balancing the two metrics.

$$\text{F1-Score} = \frac{2*(\text{Precision}*\text{Recall})}{(\text{Precision}+\text{Recall})} \; (10)$$

Experimental Setup

The experiments were conducted using the Google Colaboratory (Colab), a browser-based Python environment offering access to limited computing resources. Colab was chosen for its accessibility, built-in libraries, and support for GPU acceleration, which is suitable for lightweight machine learning tasks.

The runtime environment was set to Python 3, with access to a GPU when available, typically an NVIDIA Tesla T4 with 16 GB of VRAM. Otherwise, computations ran on a virtual CPU (Intel Xeon, ~2.2 GHz, 2 cores) with about 12.72 GB of RAM. Storage was temporary and limited to approximately 70 GB.

Implementation was done in Python 3.10 using several key libraries: NumPy (1.24.3) for numerical operations, Pandas (1.5.3) for data handling, Scikit-learn (1.3.0) for modeling and evaluation, and XGBoost (1.7.6) for gradient boosting. Visualizations were generated using Matplotlib (3.7.1) and Seaborn (0.12.2). Additional packages were installed using pip when required.

IMPLEMENTATION AND RESULTS

In this research scikit-learn and XGBoost Python packages were used to in the proposed ensemble stacking approach. In particular, the sklearn. preprocessing module was used as an advantage to scale the input features, and the sklearn.metrics module to compute the accuracy, classification report, and confusion matrix thus ensured the Utilization of sklearn.linear_model to implement the logistic regression classifier and the xgboost library for the XGBoost classifier. The models are integrated using the sklearn.ensemble module's StackingClassifier, which stacks the logistic regression and XGBoost models with another XGBoost model as the meta-learner.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue VIII August 2025

Results & Analysis

In this section, we compare the simulation-based evaluation and final model of the base paper against the proposed ensemble stacking approach.

Table 1 : K-Fold Comparison of base paper and proposed mode.

Model	Result Accuracy)	Author(s)
Logistic Regression (LR)	0.863	Hafid et al., (2022)
Support Vector Machine (SVM)	0.854	
Random Forest (RF)	0.867	
Voting Classifier (VC)	0.864	
RF (Ensemble)	0.884	
XGBoost (Ensemble)	0.904	Current Research

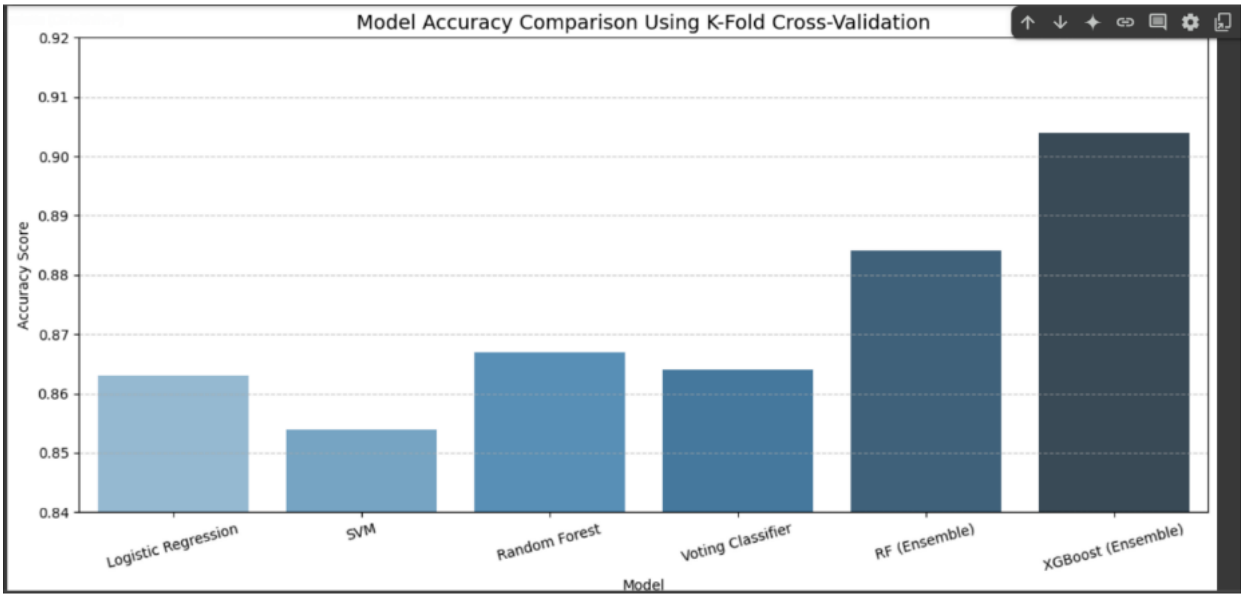

Figure 5.1 : Model accuracy comparison using K-fold cross-Validation (base paper and current model)

Table 1 and Figure 5.1 provides a K-fold cross-validation comparison among the different proposed machine learning models. This comparison is based on the score (accuracy) presented in the table. This score is calculated as the average of the accuracy. It shows that the first six models provide approximately the same score and the XGBoost (Ensemble) has the highest accuracy.

Table 2: Classification Report Table

Classification Report Table

Class	Precision	Recall	F1-Score	Support
0	0.92	0.92	0.92	12
1	0.89	0.89	0.89	9
Accuracy			0.90	21
Macro Avg	0.90	0.90	0.90	21
Weighted Avg	0.90	0.90	0.90	21

Model: XGBoost / Logistic Regression (Ensemble)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue VIII August 2025

Table 2 shows the classification report of the proposed approach. It shows the accuracy, precision, and the recall. The classification report provides a detailed performance summary of a binary classification model designed to predict two possible classes: 0 and 1. Class 0, with 12 support instances, achieved a precision, recall, and F1-score of 0.92. Similarly, class 1, with 9 instances, recorded a precision, recall, and F1-score of 0.89. These results indicate that the model consistently identifies both classes accurately, with a slightly higher reliability in detecting class 0 due to its higher precision and recall scores.

The overall accuracy of the model is 90%, meaning it correctly predicted the class for 19 out of 21 instances in the test set, highlighting its effectiveness in generalizing to unseen data. This uniformity between macro and weighted metrics indicates a well-balanced model performance, without significant bias towards either class, despite the slightly imbalanced data and weighted metrics indicates that the model performs well showing significant balance between both classes, even with the slightly imbalanced.

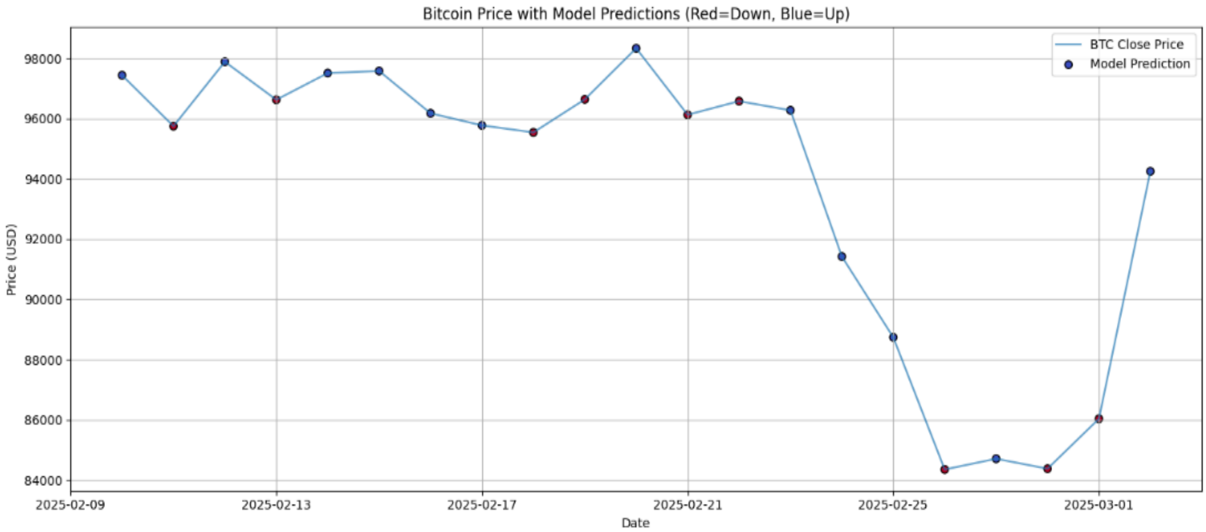

Figure 5.2: Bitcoin Backtesting (Bitcoin Price with Model Predictions)

Figure 5.2 presents the back testing results by plotting the Bitcoin close price (in USD) from February 9, 2025, to March 1, 2025, alongside the model's predictions (Red = Down, Blue = Up). The back testing strategy involves calculating the predicted price movements and comparing them with the actual price trends. Figure 2 shows that the model's predictions align closely with the actual Bitcoin price movements, as the red and blue markers (indicating predicted "Down" and "Up" movements) generally correspond to the downward and upward trends in the price line. For instance, the model correctly predicts a significant downward trend around February 25, 2025, followed by an upward recovery toward March 1, 2025. This close alignment suggests that the proposed approach performs well in predicting Bitcoin price movements, making it a reliable tool for trading strategies.

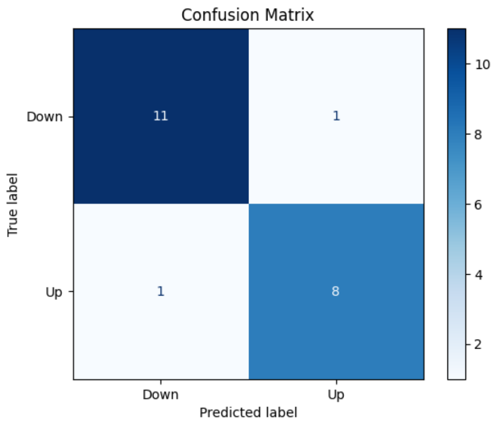

Figure 5.3 Bitcoin Confusion Matrix

Figure 5.3 displays the confusion matrix for the Bitcoin price prediction model, specifically for the XGBoost/Logistic Regression model, which achieved an accuracy of 0.904. The matrix compares true labels (Down and Up) against predicted labels, revealing the model's performance on the test set. The confusion matrix for the Bitcoin price prediction model reveals its performance in classifying price movements as "Down" or "Up." The model correctly predicted "Down" for 11 instances where the actual price movement was "Down," representing True Negatives for the "Up" class or True Positives for the "Down" class. However, it incorrectly predicted "Up" for 1 instance where the actual movement was "Down," resulting in a False Positive for the "Up" class. Similarly, the model incorrectly predicted "Down" for 1 instance where the actual movement was "Up," marking a False Negative for the "Up" class. On the positive side, it accurately predicted "Up" for 8 instances where the price movement was indeed "Up," indicating True Positives for the "Up" class. In total, the test set comprises 21 samples—calculated as 11 (True Down) + 1 (False Up) + 1 (False Down) + 8 (True Up)—which aligns closely with the support total from the previous classification report (12 for Down, 9 for Up), with minor discrepancies possibly due to rounding or a subset of the test data.

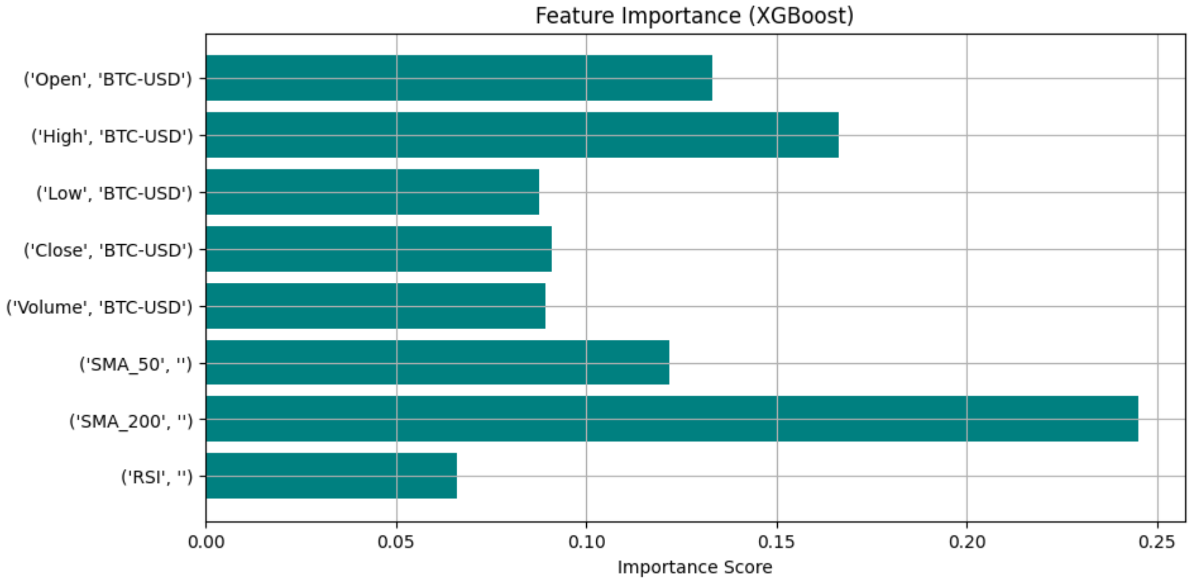

Figure 5.4 Feature importance

The XGBoost feature importance plot identifies the relative importance of input variables used in determining Bitcoin price direction. The research shows the 200-day Simple Moving Average (SMA_200) is the most significant feature with the highest score. This long-term trend indicator is crucial in the model's decision-making, indicating high importance of historical price behavior when forecasting. Its dominance justifies its adoption by traders and analysts for identifying trends.

Excluding SMA_200, other important features include the daily high (High), opening price (Open), and 50-day moving average (SMA_50). These reflect intermediate trends and market sentiment, so the model includes both trend-following and momentum measures. Their contribution shows short-term price behavior and medium-term means aid prediction. The combination strengthens the model by balancing short-term volatility with long-term stability.

Attributes like RSI, close and low prices (Low, Close), and volume traded (Volume) had lower importance scores. Though they still affected predictions, their lower influence suggests trend-based and range indicators are more reliable for classification. This supports the importance of moving averages in forecasting and provides a basis for feature prioritization in future model optimization.

DISCUSSION

The results confirm the effectiveness of an ensemble stacking approach, combining Logistic Regression and XGBoost classifiers, for predicting Bitcoin price changes. Using scikit-learn and xgboost in Python, the model applied feature scaling, robust metrics, and ensemble techniques to enhance performance.

Table 1 shows that while traditional models like Logistic Regression, SVM, Random Forest, and Voting Classifiers (Hafid et al., 2022) achieved accuracies between 0.854 and 0.884, the proposed stacking model outperformed them with 0.904. This significant improvement highlights the stability of ensemble learning in time-series forecasting. Table 2 further supports this, showing precision, recall, and F1-scores above 0.89 for both classes (0 = Down, 1 = Up) with an overall accuracy of 90%. The model's balanced performance across both directions is crucial in finance, where bias can lead to substantial trading losses. The prediction probability plot. Back testing results (Figure 5.2) demonstrate the model's predictions closely follow actual price movements, especially at turning points such as the drop on February 25, 2025, and the recovery by March 1, 2025. This accuracy supports real-world trading applications. Figure 5.3 reveals a low misclassification rate, with 19 out of 21 predictions correct, confirming the model's reliability. The near-equal true positives and negatives affirm its consistency across both classes, despite class imbalance.

Feature importance (Figure 5.4) highlights the dominance of trend-following indicators like SMA_200, High, Open, and SMA_50. The model favors long- and medium-term trends over short-term or volume-based signals, aligning with widely used trading methods that rely on moving averages to interpret market sentiment and reduce noise.

With a 90.4% success rate, the model can be easily used in automated trading systems to help traders make smart choices, including safeguards like stop-loss to limit losses when predictions go wrong, and assist regulators in watching market steadiness to tackle risky trading, though users should be careful not to rely too much on it during wild market swings

Limitations

This study looks into the limitations of a simple model used to predict Bitcoin's daily closing price movements. It points out the factors that can impact how accurate and useful the model is, which is important when deciding how reliable it is. Logistic regression is easy to understand and quick to set up, but it doesn't do a great job of capturing the complicated, non-linear nature of cryptocurrency markets. More advanced models like Long Short-Term Memory (LSTM) networks could be better at this, but they weren't used here because they need more computing power. Bitcoin prices tend to change a lot and are driven by other things like news, government rules, and world economic changes. These external factors might not be fully shown in past price data and technical indicators, which can make the model less accurate during big market changes. The focus of this model is on short-term trend predictions because long-term forecasts are really tough to get right, so they're not included. Since this model was built just for Bitcoin, it probably wouldn't work well for other cryptocurrencies like Ethereum, which have different market behaviors and need different features to predict accurately.

CONCLUSION

This research demonstrated that machine learning models, especially stacked ensembles of Logistic Regression and XGBoost, can effectively predict short-term Bitcoin price directions. Achieving 90.4% accuracy confirms the model's potential in real-world forecasting, especially in volatile markets like cryptocurrency. Nonetheless, limitations exist, including Bitcoin's non-stationarity, overfitting risks from small datasets, and the simplicity of initial features and architectures.

Despite these challenges, the study provides a foundation for future work involving deep learning models like LSTM and transformers, integrating multi-modal data, and using comprehensive back testing to simulate actual trading. Overall, this project contributes toward smarter, data-driven forecasting in cryptocurrency analytics.

RECOMMENDATIONS FOR FUTURE WORK

Though the model achieved 90.4% accuracy using stacked Logistic Regression and XGBoost, several improvements could boost generalization and robustness, further enhancement includes higher Frequency Data by using minute or hourly-level data can enhance short-term predictions by providing more training examples and detecting rapid changes. However, this introduces noise and requires careful handling of transaction costs, latency, and overfitting risks. Sophisticated Feature Engineering by adding indicators like MACD, Bollinger Bands, and Stochastic Oscillator, as well as volume-based metrics such as On-Balance Volume or the Accumulation/Distribution Line, can capture more market dynamics. Integrating sentiment from Twitter, Reddit, and on-chain metrics like hash rate and transaction volume can broaden the input feature set.Future research should explore sequence models like LSTM and Transformers for temporal data, as well as hybrid approaches combining traditional and neural network models. Other ensemble techniques, including AdaBoost, LightGBM, and varied stacking architectures, should also be investigated.

REFERENCES

1. Akyildirim, E., Cepni, O., Corbet, S., & Uddin, G. S. (2021). Forecasting mid-price movement of Bitcoin futures using machine learning. Annals of Operations Research, 1–32. https://doi.org/10.1007/s10479-021-04205-x
2. Andi, H. K. (2021). An accurate Bitcoin price prediction using logistic regression with LSTM machine learning model. Journal of Soft Computing Paradigm, 3(3), 205–217. https://doi.org/10.36548/jscp.2021.3.006
3. Chen, J. (2023). Analysis of Bitcoin price prediction using machine learning. Journal of Risk and Financial Management, 16(1), 51. https://doi.org/10.3390/jrfm16010051
4. Chen, Y. (2023). Understanding Bitcoin: A case study method to understand market dynamics, strategies, and risks of cryptocurrency. Advances in Economics, Management and Political Sciences, 62, 61–68. https://doi.org/10.54254/2754-1169/62/20231315
5. Das, S., Billah, M., & Mumu, S. A. (2021). A hybrid approach for predicting Bitcoin price using Bi-LSTM and Bi-RNN based neural network. In H. Yin et al. (Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2021 (Vol. 13113, pp. 223–233). Springer. https://doi.org/10.1007/978-3-030-91608-4_23
6. Dimitriadou, A., & Gregoriou, A. (2023). Predicting Bitcoin prices using machine learning. Entropy, 25, 777. https://doi.org/10.3390/e25050777
7. Domícián, M., Hassan, R., Ishtiaq, A., Sándor, J., & Kovács, M. (2024). Next step for Bitcoin: Confluence of technical indicators and machine learning. Journal of International Studies, 17(3), 68–94. https://doi.org/10.14254/2071-8330.2023/17-3/4
8. El Badaoui, M., Raouyane, B., El Moumen, S., & Bellafkih, M. (2023). Impact of machine learning classification and technical indicators: Forecasting the direction of Bitcoin. In 2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA) (pp. 1–7). IEEE. https://doi.org/10.1109/SITA60746.2023.10373706
9. Fauzi, M. A., Paiman, N., & Othman, Z. (2020). Bitcoin and cryptocurrency: Challenges, opportunities, and future works. Journal of Asian Finance, Economics and Business, 7(8), 695–704. https://doi.org/10.13106/jafeb.2020.vol7.no8.695
10. Google. (2025, February 26). Logistic regression: Sigmoid function. Google Developers. https://developers.google.com/machine-learning/crash-course/logistic-regression/sigmoid-function
11. Hafid, A., & Makrakis, D. (2023). Bitcoin price prediction using machine learning and technical indicators. In S. Ossowski et al. (Eds.), Distributed Computing and Artificial Intelligence, 20th International Conference, DCAI 2023 (Vol. 740). Springer. https://doi.org/10.1007/978-3-031-38333-5_28
12. Hafid, A., Senhaji Hafid, A., & Makrakis, D. (2022). Bitcoin price prediction using machine learning and technical indicators. Preprints. https://doi.org/10.20944/preprints202212.0188.v1
13. Hajare, R., Puri, C., & Gote, P. M. (2025). Bitcoin price prediction using deep learning models. In Proceedings of the 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 1678–1683). IEEE. https://doi.org/10.1109/ICSADL65848.2025.10933412
14. Ibrahim, A., Kashef, R., Li, M., Valencia, E., & Huang, E. (2020). Bitcoin network mechanics: Forecasting the BTC closing price using vector auto-regression models. Journal of Risk and Financial Management, 13(9), 189. https://doi.org/10.3390/jrfm13090189

15. Jung, H. S., Kim, J. H., & Lee, H. (2024). Decoding Bitcoin: Leveraging macro- and micro-factors in time series analysis for price prediction. PeerJ Computer Science, 10, e2314. https://doi.org/10.7717/peerj-cs.2314
16. Lizadegan, H., Radmehr, A., & Asghari Ilani, M. (2024). Forecasting Bitcoin prices: A comparative study of machine learning and deep learning algorithms. Research Square. https://doi.org/10.21203/rs.3.rs-4390390/v1
17. Maalouf, M. (2011). Logistic regression in data analysis: An overview. International Journal of Data Analysis Techniques and Strategies, 3, 281–299. https://doi.org/10.1504/IJDATS.2011.041335
18. Ming-Che, L. (2024). Bitcoin trend prediction with attention-based deep learning models and technical indicators. Systems, 12(11), 498. https://doi.org/10.3390/systems12110498
19. Mittal, M., & Geetha, G. (2022). Predicting Bitcoin price using machine learning. In 2022 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1–7). IEEE. https://doi.org/10.1109/ICCCI54379.2022.9740772
20. Moiz, Q., Hasnain, I., Rodrigues, P. C., Rehman, M. Z., & Salar, S. A. (2024). Statistical modeling to improve time series forecasting: A case study of Bitcoin. Mathematics, 12(23), 3666. https://doi.org/10.3390/math12233666
21. Nabipour, M., Nayyeri, P., Jabani, H., S. S., & Mosavi, A. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data. IEEE Access, 8, 150199–150212. https://doi.org/10.1109/ACCESS.2020.3015966
22. Nakamoto, S. (2009). Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf
23. Pabuçcu, H., Ongan, S., & Ongan, A. (2020). Forecasting the movements of Bitcoin prices: An application of machine learning algorithms. Quantitative Finance and Economics, 4(4), 679–692. https://doi.org/10.3934/qfe.2020031
24. Patel, S., & Prajapati, P. (2021). Bitcoin price prediction using machine learning. Journal of Financial Analysis, 15(3), 45–56. https://doi.org/10.1016/j.finana.2021.05.005
25. Perepu, S. S. (2024). Machine learning. ResearchGate. https://www.researchgate.net/publication/386279653_Machine_learning
26. Pichaiyuth, P., Termnuphan, P., Triyason, T., Rojanapornpun, O., & Jaiyen, S. (2023). Price trend forecasting of cryptocurrency using multiple technical indicators and SHAP. In 2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE) (pp. 150–154). IEEE. https://doi.org/10.1109/JCSSE58229.2023.10201984
27. Qinghe, L. (2021). Predicting trends of Bitcoin prices based on machine learning methods. In Proceedings of the 2020 4th International Conference on Software and e-Business (ICSEB '20) (pp. 49–52). ACM. https://doi.org/10.1145/3446569.3446588
28. Shadiya, M. S., & Chitra, S. (2024). A study on cryptocurrency as digital revolution - Pros and cons with special reference on Indian context. International Journal of Linguistics Applied Psychology and Technology (IJLAPT), 1(06), 1–6. https://doi.org/10.69889/ijlapt.v1i06.64
29. Soni, K., & Singh, S. (2022). Bitcoin price prediction—An analysis of various regression methods. In 2022 IEEE 12th Symposium on Computer Applications & Industrial Electronics (ISCAIE) (pp. 271–276). IEEE. https://doi.org/10.1109/ISCAIE54458.2022.9794532
30. Sun, S., Zheng, Y., & Wang, J. (2024). Deep learning for Bitcoin price direction prediction: Models and applications. Financial Innovation, 10(2), 123–134. https://doi.org/10.1186/s40854-024-00643
31. Ugurushetty, S. (2024). Understanding Bitcoin and blockchain. International Journal of Scientific Research in Engineering and Management, 08(08), 1–13. https://doi.org/10.55041/ijsrem36983
32. Uras, N., Marchesi, L., Marchesi, M., & Tonelli, R. (2020). Forecasting Bitcoin closing price series using linear regression and neural networks models. PeerJ Computer Science, 6, e279. https://doi.org/10.7717/peerj-cs.279
33. Wang, J., Feng, K., & Qiao, G. (2023). A hybrid deep learning model for Bitcoin price prediction: Data decomposition and feature selection. Applied Economics, 56(53), 6890–6905. https://doi.org/10.1080/00036846.2023.2276093
34. Zhou, Y. (2022). Prediction on Bitcoin price trends based on machine learning algorithms. BCP Business & Management, 34, 21–29. https://doi.org/10.54691/bcpbm.v34i.2860