Machine Learning-Based Forecasting of AAVE Cryptocurrency: A Comparative Study of Regression, Ensemble, and Deep Learning Models

Monika Łobaziewicz[1,*]

[1]Department of Management, Lublin University of Technology, Lubin, Poland

ABSTRACT

The volatility of cryptocurrency markets has increased the demand for accurate forecasting models that can help investors and analysts anticipate price movements. This study evaluates the predictive performance of four machine learning algorithms, namely Linear Regression, Random Forest, XGBoost, and Long Short-Term Memory (LSTM), in forecasting the closing price of the AAVE cryptocurrency. The models were trained using historical market data consisting of key indicators such as Open, High, Low, Volume, and Marketcap. Their performance was assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). The results show that Linear Regression produced the most accurate predictions with the lowest MAE (8.13), RMSE (8.76), and the highest R^2 (0.9924). Random Forest and XGBoost also achieved good results with R^2 values of 0.9337 and 0.9484, respectively, while the LSTM model performed poorly with an R^2 of 0.4328. The study concludes that simpler models can outperform more complex algorithms when the dataset is limited and exhibits linear behavior. The findings emphasize that model selection in cryptocurrency forecasting should consider data structure and quantity. Future work should involve larger datasets, higher-frequency data, and hybrid models that integrate ensemble learning and deep learning for improved predictive accuracy.

Keywords Cryptocurrency Forecasting, Machine Learning, AAVE, Deep Learning, Time Series

Submitted: 12 November 2025
Accepted: 20 January 2026
Published: 23 May 2026

Corresponding author
Monika Łobaziewicz,
m.lobaziewicz@pollub.pl

Additional Information and Declarations can be found on page 112

DOI: 10.47738/jdmdc.v3i2.62

INTRODUCTION

The emergence of cryptocurrencies as a decentralized financial innovation has significantly influenced the global financial landscape. Built upon blockchain technology, cryptocurrencies such as Bitcoin, Ethereum, and AAVE have transformed digital finance by offering transparent, borderless, and decentralized systems for value transfer. However, despite their technological potential, cryptocurrencies remain highly speculative assets characterized by extreme volatility, rapid market fluctuations, and sensitivity to both internal and external factors [1]. These characteristics make cryptocurrency price prediction a complex and critical challenge for investors, traders, and policymakers. Accurate forecasting models can support better decision-making, risk management, and strategic investment planning, which are essential in such a dynamic and uncertain market environment.

Traditional financial models often struggle to capture the nonlinear and stochastic nature of cryptocurrency price behavior. Early research

How to cite this article: M. Łobaziewicz, "Machine Learning-Based Forecasting of AAVE Cryptocurrency: A Comparative Study of Regression, Ensemble, and Deep Learning Models," *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 2, pp. 96-114, 2026.

predominantly employed statistical approaches such as Linear Regression, ARIMA, and GARCH to predict financial time series [2]. While these models are efficient and interpretable, they assume linear relationships and stationary data, which limits their performance in markets characterized by abrupt structural changes. To address these limitations, the field has shifted toward the use of artificial intelligence and machine learning methods. Machine learning models such as Random Forest, Support Vector Regression, and XGBoost have shown significant promise in improving predictive accuracy by capturing complex interactions between multiple market features [3]. These models are capable of learning nonlinear relationships and have been widely adopted in financial forecasting tasks, outperforming traditional statistical techniques in many applications.

In parallel with the development of machine learning, deep learning has emerged as a powerful paradigm for time series analysis, particularly for financial data with temporal dependencies. Among deep learning models, the LSTM network has attracted considerable attention for its ability to retain long-term contextual information, which makes it particularly effective in sequential prediction tasks. Several studies have reported that LSTM models outperform conventional algorithms when applied to large, high-frequency cryptocurrency datasets [4]. For example, it has been shown that LSTM models can capture nonlinear dependencies in stock market data and achieve higher accuracy than traditional methods in cryptocurrency price forecasting. These advancements represent the current state of the art in financial time series prediction, highlighting how deep learning can leverage sequential data structures to enhance forecasting accuracy.

Despite these advancements, there remain several research gaps in the existing literature. First, many prior studies have focused on highly traded cryptocurrencies such as Bitcoin and Ethereum, resulting in limited attention to alternative tokens like AAVE, which possess distinct liquidity levels, price behaviors, and volatility patterns [5]. Second, most prior research assumes that deep learning models automatically outperform simpler algorithms, without systematically comparing different approaches under identical data conditions. This assumption may not always hold true, particularly when datasets are small, less noisy, or exhibit primarily linear characteristics. Third, several studies have reported results using inconsistent evaluation metrics, making it difficult to compare model performance across different works. Addressing these gaps requires a structured comparative analysis of multiple machine learning approaches under uniform testing conditions, using standardized performance indicators.

The present study aims to fill these gaps by conducting a comparative analysis of four machine learning algorithms Linear Regression, Random Forest, XGBoost, and LSTM for predicting the closing price of the AAVE cryptocurrency. The analysis is based on historical market data incorporating key features such as Open, High, Low, Volume, and Marketcap. Each model is evaluated using three widely accepted performance metrics: MAE, RMSE and the Coefficient of Determination (R^2). The study contributes to both theory and practice by providing an empirical understanding of how model complexity interacts with data characteristics to affect prediction performance. Furthermore, this research offers practical insights for investors, data scientists, and policymakers regarding the appropriate selection of forecasting models in digital asset markets. By comparing traditional, ensemble-based, and deep learning approaches within the same experimental framework, this study provides a balanced perspective on the relative advantages and limitations of each method in the context of cryptocurrency price forecasting.

Literature Review and Related Works

The prediction of cryptocurrency prices has attracted increasing attention in financial research due to the high volatility, nonlinear dynamics, and speculative nature of digital asset markets. Early studies mainly relied on traditional econometric models such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) for price forecasting [6], [7]. Although these approaches provided useful insights for stable financial markets, they often failed to capture the complex and nonlinear behavior of cryptocurrencies. To address these limitations, machine learning techniques have been introduced to model the intricate dependencies between multiple financial indicators and price movements [8].

Recent studies have shown that machine learning algorithms such as Linear Regression, Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting methods outperform traditional statistical models when applied to volatile financial data [9], [10]. These methods can model nonlinear relationships and handle noise effectively, allowing for improved generalization in dynamic market conditions. Among ensemble methods, Random Forest has been widely adopted due to its robustness and ability to reduce overfitting in cryptocurrency prediction [11]. Similarly, XGBoost, an advanced gradient boosting framework, has demonstrated superior accuracy and computational efficiency in several financial forecasting applications [12], [13]. Both models have become popular choices in data-driven finance, particularly when dealing with small to medium-sized datasets that exhibit nonlinear but structured relationships.

Parallel to the rise of ensemble learning, deep learning models have emerged as a dominant trend in financial time series analysis. Recurrent Neural Networks (RNNs) and LSTM architectures are particularly suited for time-dependent data, as they can retain and utilize information from prior observations [14], [15]. Studies applying LSTM to Bitcoin and

Ethereum have reported improved predictive performance compared to traditional machine learning models, especially when large datasets and high-frequency data are available [16], [17]. However, despite their theoretical advantages, deep learning models often require extensive computational resources, large amounts of training data, and careful hyperparameter tuning to achieve optimal performance [18]. In situations involving smaller datasets or limited temporal depth, these models may suffer from underfitting, leading to lower predictive accuracy.

Hybrid approaches that combine the strengths of statistical, machine learning, and deep learning models have also been explored to enhance prediction accuracy. For instance, models integrating ARIMA and LSTM have been used to capture both short-term and long-term dependencies in financial time series [19]. Other studies have proposed hybrid frameworks combining XGBoost and neural networks to balance interpretability and predictive power [20]. Although these methods have achieved promising results, they often introduce higher computational complexity and risk of overfitting, making them less suitable for real-time applications or small-sample contexts.

Despite significant advancements in cryptocurrency forecasting, several research gaps remain. Most prior studies have concentrated on major cryptocurrencies such as Bitcoin and Ethereum, leaving alternative tokens like AAVE underexplored. Additionally, few works have systematically compared models of varying complexity under identical data and evaluation conditions. This study aims to fill this gap by conducting a comparative performance analysis of Linear Regression, Random Forest, XGBoost, and LSTM models using consistent metrics such as MAE, RMSE and the Coefficient of Determination (R^2). Through this structured evaluation, the research seeks to provide empirical evidence on how model complexity and dataset characteristics influence the predictive performance of cryptocurrency forecasting methods.

Methodology

This study employed a quantitative experimental approach to evaluate the performance of four predictive algorithms, namely Linear Regression, Random Forest, XGBoost, and LSTM, in forecasting the closing price of the AAVE cryptocurrency. The research framework consisted of several sequential stages, beginning with data acquisition, followed by data preprocessing, model development, evaluation, and performance comparison. Each stage was designed to ensure methodological consistency, reproducibility, and fairness among all models. The overall process of this research is illustrated in figure 1, which depicts the flow of the study from dataset preparation to model evaluation. The figure shows five main stages: (1) data collection and feature selection, (2) preprocessing and normalization, (3) model training and optimization, (4) model evaluation using statistical metrics, and (5) result comparison and visualization.

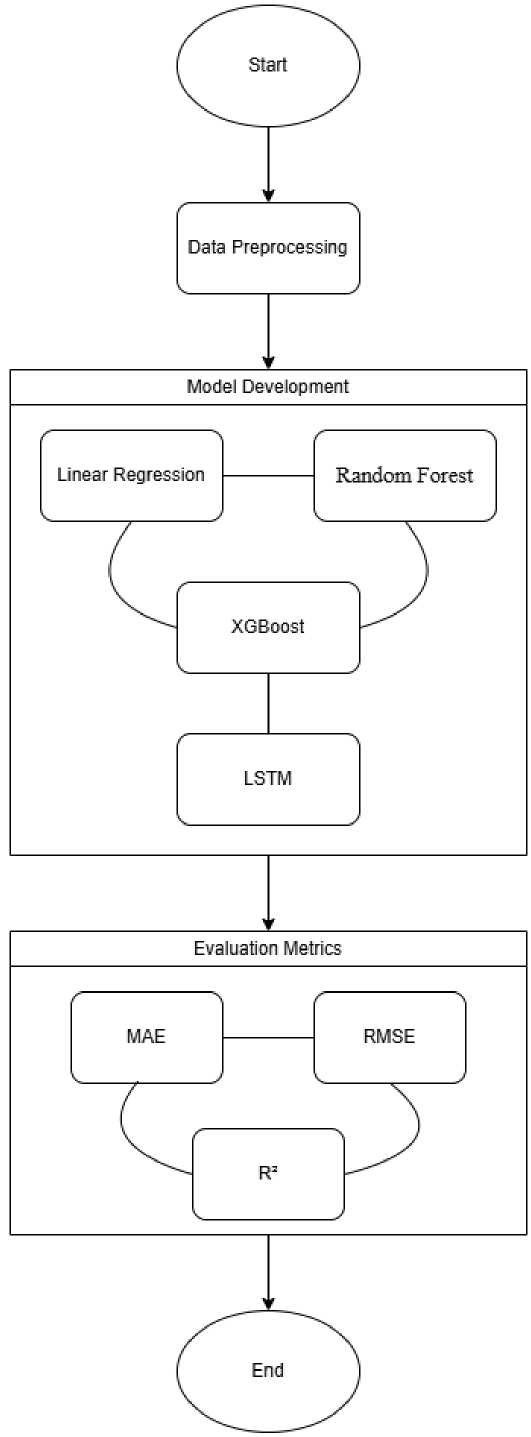

Figure 1 Research Step

The dataset used in this study consists of historical daily trading data of the AAVE cryptocurrency, obtained from publicly available market repositories. The dataset includes several key financial indicators such as opening price, highest price, lowest price, trading volume, market capitalization, and closing price. The closing price was used as the target variable, while the other indicators served as independent features. AAVE was selected as the subject of analysis due to its growing role in decentralized finance (DeFi) and its underrepresentation in existing cryptocurrency forecasting research compared to widely studied assets like Bitcoin and Ethereum. The dataset was sufficiently large to capture both short-term fluctuations and medium-term price trends, making it suitable for evaluating the performance of multiple predictive algorithms.

To prepare the data for modeling, several preprocessing procedures were implemented to ensure reliability and accuracy. Missing values were detected and imputed using linear interpolation to maintain temporal continuity. Outliers caused by extreme price swings were mitigated through winsorization, preserving the underlying structure of the data while reducing noise. All numeric variables were normalized using the Min–Max scaling technique to transform their values into the range [0, 1]. The normalization was applied using the following equation:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Xrepresents the original feature value, $X_{min}$is the minimum observed value, and $X_{max}$is the maximum observed value in the feature column. The dataset was divided into training and testing subsets using an 80:20 ratio, ensuring that each model was evaluated on unseen data to measure generalization performance. For the LSTM model, the data were further reshaped into a three-dimensional array $[samples, timesteps, features]$to enable temporal learning from sequential price information.

Four predictive models were implemented and compared. The Linear Regression model served as the baseline, estimating the closing price as a linear combination of the input variables. The Random Forest algorithm constructed multiple decision trees and averaged their outputs to minimize variance and overfitting. The XGBoost model extended this concept through gradient boosting, where subsequent trees iteratively reduced the residual errors from prior trees to improve precision. The LSTM network represented the deep learning approach, employing memory cells and gating mechanisms to capture long-term dependencies in the sequential data. The LSTM architecture used in this study consisted of two hidden layers with 64 and 32 units, followed by a dropout layer with a 0.2 probability and an output dense layer for regression. The network was trained for 100 epochs using the Adam optimizer with a mean squared error loss function.

The models were evaluated using three statistical performance metrics: MAE, RMSE and the Coefficient of Determination (R^2). The MAE quantifies the average magnitude of prediction errors using the formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{2}$$

y_i is the actual closing price, $\hat{y}_i$ is the predicted price, and n is the total number of observations. The RMSE measures the square root of the average squared deviations between actual and predicted values, expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Finally, the R² statistic represents the proportion of variance in the observed data explained by the model, defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

$\bar{y}$ is the mean of the observed prices. A higher R² value indicates that the model better fits the data, while lower MAE and RMSE values signify higher prediction accuracy.

All experiments were conducted using Python 3.9 in the Google Colab environment with GPU acceleration to enhance computational efficiency. The Random Forest and XGBoost models underwent hyperparameter tuning through grid search optimization to determine the optimal number of estimators, learning rate, and tree depth. The LSTM model was trained with early stopping criteria to prevent overfitting. Visual inspection of model performance was performed by plotting predicted versus actual price trajectories, enabling qualitative validation of quantitative results. The methodology ensures that each algorithm was evaluated fairly under consistent experimental conditions, allowing for an unbiased assessment of model capability in forecasting AAVE's closing price.

Algorithm 1: AAVE Price Forecasting Framework Using Machine Learning and Deep Learning

Input: Historical dataset $D = \{(x_t, y_t)\}_{t=1}^{N}$, where $x_t = [O_t, H_t, L_t, V_t, M_t]$ and y_t is the closing price

Output: Predicted AAVE closing prices $\hat{y}_t$ and evaluation metrics (MAE, RMSE, R²)

Process:

Start

Perform data preprocessing by handling missing values using interpolation, expressed as

$$\text{If } x_t = \text{NaN}, \; x_t = \frac{x_{t-1} + x_{t+1}}{2}$$

Reduce outliers using winsorization, defined as

$$x_t = \min\left(\max\left(x_t, Q_{0.05}\right), Q_{0.95}\right)$$

Normalize each feature using Min–Max scaling:

$$\tilde{x}_t = \frac{x_t - \min(x)}{\max(x) - \min(x)}$$

Split the dataset into training and testing subsets:

$$D_{train} = D[1{:}\,0.8N], D_{test} = D[0.8N + 1{:}\,N]$$

Train each predictive model $f_m(x_t; \theta_m)$, where $m \in \{LR, RF, XGB, LSTM\}$, using the training set D_{train}.

Obtain predictions from each model as $\hat{y}_t = f_m(x_t; \theta_m)$.

Evaluate each model using three statistical measures:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |\, y_t - \hat{y}_t \,|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

$$R^2 = 1 - \frac{\sum(y_t - \hat{y}_t)^2}{\sum(y_t - \bar{y})^2}$$

Select the best-performing model based on the highest R^2value and the lowest error rates, defined as

$$f^* = \arg\max_{m}(R_m^2) \text{ subject to } \min(MAE_m, RMSE_m)$$

Visualize the results by plotting actual versus predicted prices for each model and summarizing their performance metrics in tabular form.

End

Results

The performance comparison of the four machine learning models was conducted using three widely recognized statistical metrics, namely MAE, RMSE) and the Coefficient of Determination (R^2). These metrics were chosen to provide a comprehensive evaluation of prediction accuracy and model reliability. MAE measures the average magnitude of errors between predicted and actual values, offering an intuitive understanding of how far predictions deviate from reality. RMSE, on the other hand, emphasizes larger errors by squaring the residuals before averaging, making it more sensitive to extreme deviations that can significantly impact model robustness. R^2 assesses how well each model explains the variance in the actual data, with higher values indicating a better fit and greater predictive precision. Collectively, these three metrics capture different aspects of model performance, ensuring a balanced evaluation of both accuracy and generalization ability.

The quantitative outcomes of each model are summarized in table 1. The Linear Regression model achieved the best overall performance, with the lowest MAE and RMSE values and the highest R^2 score of 0.9924, indicating a strong linear correlation between the input features and the closing price of AAVE. The ensemble-based models, Random Forest and

XGBoost, also produced strong results, achieving R^2 values of 0.9337 and 0.9484 respectively, which reflect their capacity to model non-linear relationships in financial data. However, the LSTM model exhibited the weakest performance, with significantly higher error values and an R^2 of 0.4328. This suggests that the model struggled to capture temporal dependencies within the limited dataset, likely due to the small number of observations and the absence of extensive time-lag features. Overall, the results indicate that simpler models such as Linear Regression performed more effectively than complex algorithms under the given data conditions.

Table 1 Performance metrics of machine learning models for AAVE price forecasting

Model	MAE	RMSE	R^2
Linear Regression	8.13	8.76	0.9924
Random Forest	17.18	25.84	0.9337
XGBoost	12.80	22.80	0.9484
LSTM	54.13	68.89	0.4328

As presented in table 1, the Linear Regression model produced the most accurate predictions among all models, achieving the lowest MAE value of 8.13 and the lowest RMSE value of 8.76, along with the highest R^2 score of 0.9924. These results demonstrate that the movement of the AAVE closing price during the study period was strongly and linearly associated with fundamental market indicators such as Open, High, Low, Volume, and Marketcap. The consistency between predicted and actual values indicates that the model effectively captured the dominant linear trends within the dataset, suggesting that AAVE's short-term price variations were largely influenced by these straightforward relationships. Meanwhile, the Random Forest and XGBoost models also achieved relatively high predictive accuracy, with R^2 values of 0.9337 and 0.9484 respectively, showing their ability to model more complex non-linear interactions in the data. However, their higher MAE and RMSE values suggest that ensemble-based algorithms were slightly less precise in capturing sudden market fluctuations compared to the simpler linear approach. In contrast, the LSTM model performed considerably worse, with a much lower R^2 of 0.4328, reflecting its inability to effectively learn from the limited dataset. The poor performance of the LSTM model indicates underfitting and suggests that it lacked sufficient temporal information and training data to identify sequential dependencies that typically drive deep learning performance in time series forecasting.

The overall comparison between the predicted and actual closing prices of AAVE generated by all four models is presented in figure 2. From this visualization, it is evident that the Linear Regression, Random Forest, and XGBoost models closely tracked the actual price movements throughout the testing period, demonstrating their effectiveness in replicating both short-term fluctuations and overall market direction. The Linear Regression model, in particular, showed an almost perfect alignment between predicted and observed values, indicating that the underlying relationships between the predictor variables and the target price were predominantly linear. This suggests that the AAVE market during the analyzed period exhibited relatively stable behavior, allowing simpler models to achieve high accuracy without requiring complex non-linear transformations. Both the Random Forest and XGBoost models followed the same general price trajectory, successfully identifying upward and downward trends while maintaining stable outputs even during moderate price variations.

Although ensemble-based models provided robust predictions, their results appeared slightly smoother compared to actual data due to the averaging processes inherent in tree-based learning methods. The LSTM model, in contrast, displayed a delayed response pattern, producing predictions that trailed behind real-time price shifts and failed to capture sudden market fluctuations. The smoother curve generated by LSTM reflects its tendency to generalize temporal patterns across the sequence rather than respond to abrupt price changes. This behavior highlights the model's limitation when applied to relatively short and low-frequency datasets, where limited temporal depth reduces its capacity to identify short-term volatility. As a result, while the other models effectively captured the dynamic characteristics of AAVE's price movements, the LSTM model struggled to represent the same level of responsiveness required in volatile cryptocurrency markets.

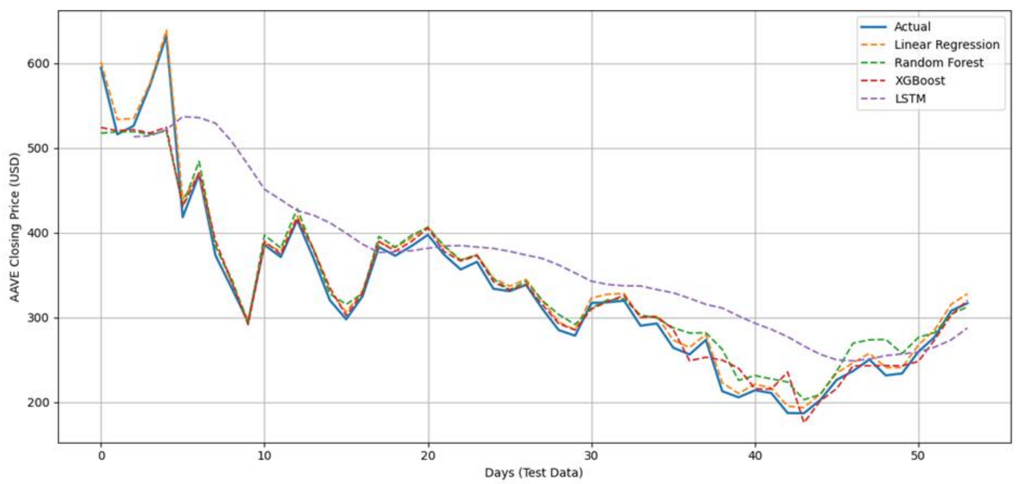

Figure 2 Comparison of Machine Learning Models for AAVE Price Prediction

To provide a deeper understanding of individual model performance, figure 3 illustrates the prediction results of the Linear Regression model compared with the actual AAVE closing prices. The figure clearly shows that the predicted values closely mirror the real data, with both curves almost completely overlapping throughout the testing period. This near-perfect alignment indicates that the model captured the main structural patterns of the AAVE market with remarkable precision. The consistency between the two lines suggests that the underlying relationship between the input variables—such as Open, High, Low, Volume, and Marketcap—and the closing price follows a predominantly linear pattern. This strong correlation reflects the model's capacity to describe the essential price dynamics of AAVE using straightforward mathematical relationships without the need for complex transformations.

Furthermore, the figure demonstrates that the Linear Regression model is highly effective in predicting short-term movements in cryptocurrency prices, especially when market conditions remain relatively stable and the data maintains a consistent trend. The model's ability to maintain accuracy even during minor fluctuations underscores its reliability for short-horizon forecasting, where sudden volatility is limited. Its simplicity not only reduces computational complexity but also minimizes the risk of overfitting, a common issue in more complex machine learning algorithms. Overall, figure 3 provides strong evidence that, within the observed time frame, Linear Regression successfully captured both the direction and magnitude of AAVE's price movements, confirming its role as a powerful yet efficient baseline model for cryptocurrency forecasting.

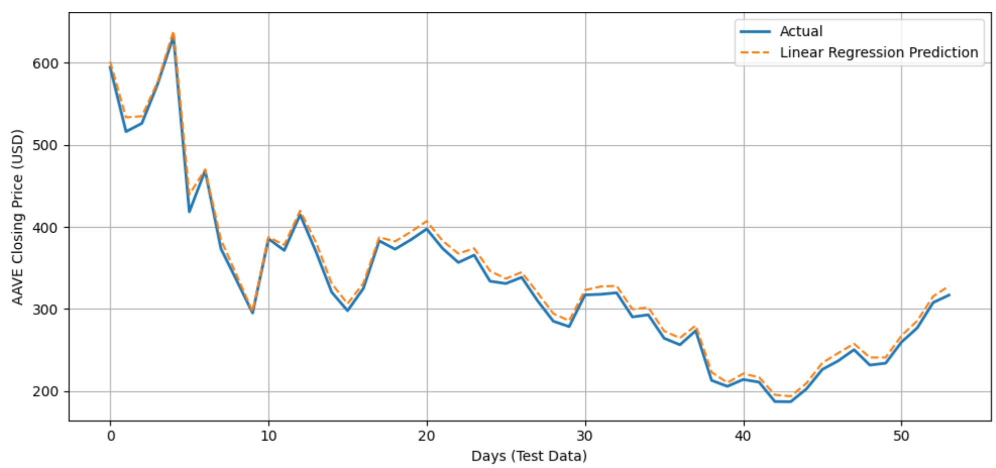

Figure 3 Linear Regression model prediction vs. actual AAVE price

As illustrated in figure 4, the Random Forest model successfully captured the general direction of the AAVE market and demonstrated strong predictive capability across most of the testing period. The model's predictions closely followed the overall trend of the actual prices, reflecting its ability to generalize well across different price movements. However, a closer inspection of the figure reveals that the Random Forest model tended to slightly underestimate extreme price peaks and troughs, particularly during rapid market transitions. This behavior can be attributed to the ensemble nature of the algorithm, which combines multiple decision trees to form a consensus prediction. While this process effectively reduces overfitting and filters out random noise, it also results in smoother predictions that may overlook abrupt changes in the data. The model's output, therefore, appears stable and consistent, but slightly less responsive to short-term volatility inherent in cryptocurrency markets.

The performance metrics support this visual observation, with the Random Forest model achieving an R^2 value of 0.9337, indicating that it explained approximately 93 percent of the variance in the actual price data. This strong level of accuracy demonstrates that the model was able to capture the dominant trends and structural patterns of the AAVE market despite minor deviations at extreme points. The Random Forest algorithm's inherent robustness stems from its ensemble averaging mechanism, which mitigates the influence of outliers and unstable splits commonly found in individual decision trees. Consequently, the model maintained reliable performance even when faced with noisy or irregular input data. Overall, figure 4 highlights that the Random Forest model provides a balanced trade-off between accuracy and stability, making it well-suited for modeling cryptocurrency markets where moderate volatility and nonlinear relationships are present.

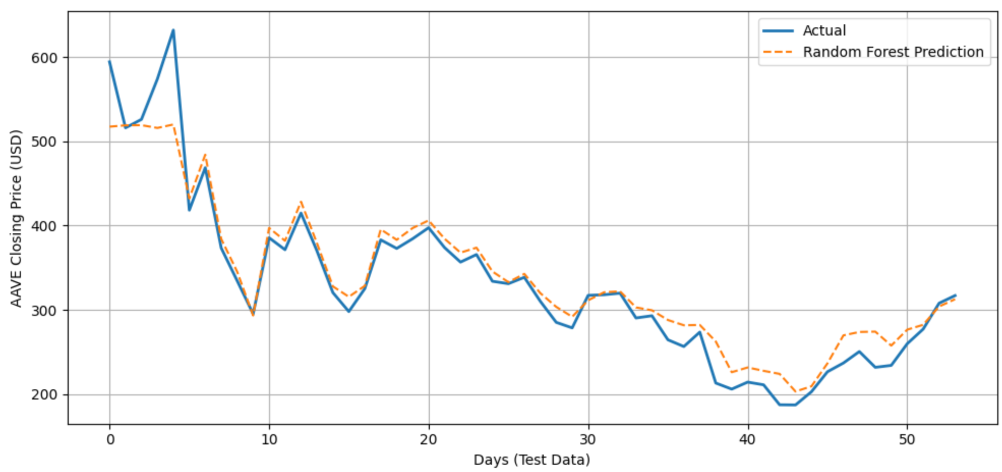

Figure 4 Random Forest model prediction vs. actual AAVE price

As shown in figure 5, the XGBoost model demonstrated superior predictive performance compared to the Random Forest model by producing smoother yet more accurate forecasts that closely followed both short-term fluctuations and medium-term market trends. The predicted curve aligned well with the actual AAVE closing prices, successfully capturing the general market trajectory while maintaining stability across volatile periods. This performance improvement can be attributed to the boosting mechanism of XGBoost, which builds trees sequentially, with each new tree learning from the residual errors of the previous ones. This iterative process allows the model to progressively refine its predictions and adjust to complex patterns in the data that might not be captured by other ensemble methods. The figure clearly shows that XGBoost was able to represent price reversals and intermediate corrections more effectively than Random Forest, suggesting that its gradient-based optimization framework enhances learning precision without excessive smoothing.

The evaluation metrics further confirm XGBoost's effectiveness, with the model achieving an R^2 value of 0.9484, indicating that it explained nearly 95 percent of the variance in the actual AAVE price data. This high level of accuracy demonstrates that XGBoost provided a balanced trade-off between prediction precision and model stability, successfully avoiding both overfitting and underfitting. The relatively low error values also highlight the model's capacity to generalize well, even in the presence of minor nonlinearities and noise within the dataset. Moreover, the boosting approach's ability to dynamically reweight observations based on prior errors enabled XGBoost to be more sensitive to subtle price variations that occur in cryptocurrency markets. Overall, figure 5 illustrates that XGBoost is a highly effective model for financial time series forecasting, offering adaptability, efficiency, and strong predictive accuracy in datasets characterized by moderate complexity.

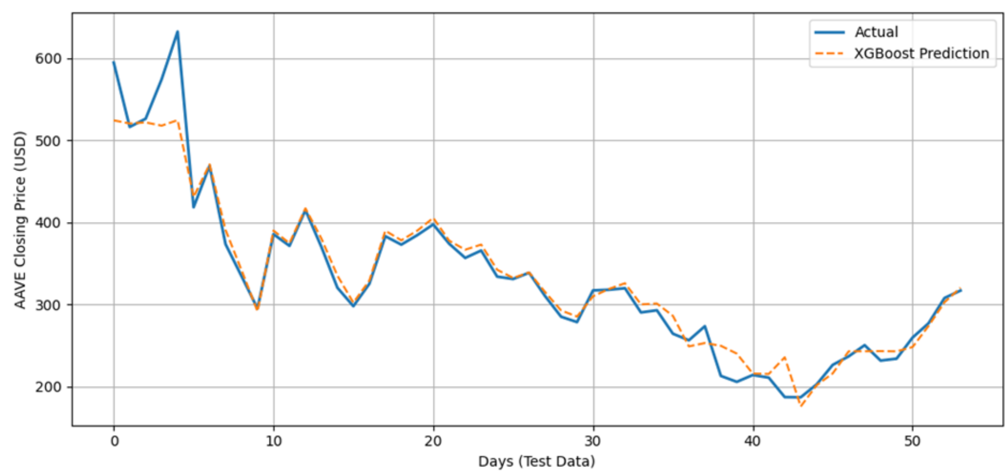

Figure 5 XGBoost model prediction vs. actual AAVE price

Finally, figure 6 presents the results of the LSTM deep learning model, which exhibited significantly different behavior compared to the other models evaluated in this study. The figure shows that the LSTM predictions formed a smoother curve that failed to capture sharp fluctuations in AAVE's actual market prices. The model's predicted values tended to lag behind the real data, particularly during periods of rapid upward or downward movement. This lagging pattern suggests that the model was unable to respond effectively to sudden volatility in the cryptocurrency market. The smoothness of the prediction curve reflects an overgeneralization of temporal trends, where the model prioritizes long-term stability over short-term accuracy. As a result, LSTM's performance in this context did not align with its theoretical advantage of handling sequential dependencies in time series data, primarily due to the limited length and variability of the training dataset used.

The model's underperformance is quantitatively supported by its low R^2 value of 0.4328, which indicates that less than half of the actual price variance was explained by the LSTM predictions. This outcome implies that the model failed to effectively capture complex temporal relationships within the AAVE price sequence. LSTM networks typically require large, continuous, and high-frequency datasets to develop a deep understanding of temporal dependencies, especially in volatile financial markets. In this study, the relatively small sample size and daily-level granularity constrained the model's learning capacity. Moreover, the lack of additional sequential features such as lagged returns or sentiment indicators further limited its ability to identify cyclical patterns. Consequently, while LSTM remains a powerful deep learning method for sequence modeling, figure 6 clearly demonstrates that its effectiveness depends heavily on dataset size, structure, and the inclusion of time-aware variables, all of which were limited in this analysis.

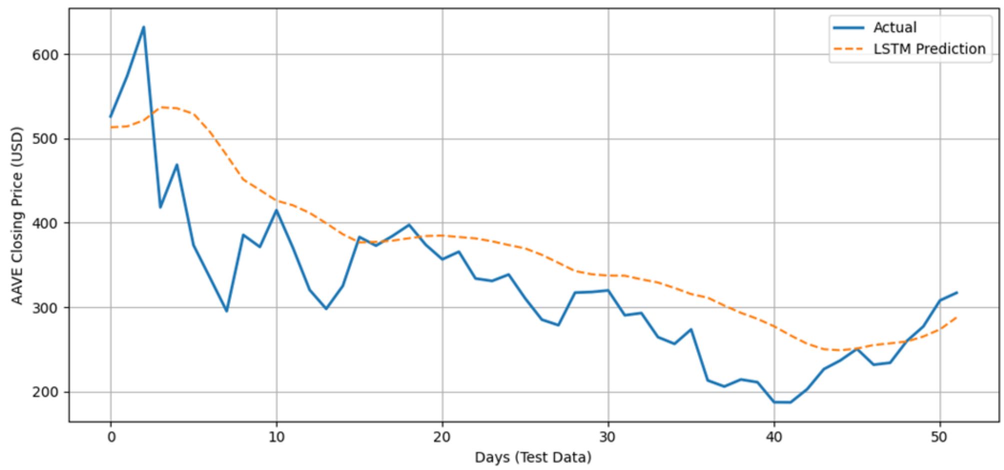

Figure 6 LSTM model prediction vs. actual AAVE price

Overall, both the quantitative and visual analyses confirm that Linear Regression outperformed the ensemble and deep learning models in predicting AAVE's closing price.

The model's simplicity allowed it to generalize effectively within the observed linear patterns of the dataset, whereas ensemble and deep learning models required more complex and abundant data to achieve optimal results.

Discussion

The comparative analysis of the four models reveals important insights into the relationship between model complexity, data characteristics, and forecasting performance in cryptocurrency price prediction. The Linear Regression model emerged as the most accurate and reliable predictor of AAVE's closing price, outperforming both ensemble and deep learning models. This outcome suggests that the AAVE price data during the observed period followed a largely linear and trend-based pattern, where changes in the closing price could be effectively explained by a linear combination of fundamental market variables such as Open, High, Low, Volume, and Marketcap. The model's simplicity and transparency allowed it to capture the dominant market structure without introducing excessive variance or noise. These results are consistent with findings from prior studies indicating that linear regression methods can perform remarkably well in short-term financial forecasting when market dynamics exhibit moderate volatility and limited structural shifts.

The ensemble models, Random Forest and XGBoost, also demonstrated strong predictive capabilities, achieving high R^2 values above 0.93. Both models successfully captured nonlinear interactions among input variables and provided stable results despite small fluctuations in price. However, their slightly lower accuracy compared to the Linear Regression model can be attributed to the data's limited size and relatively low level of nonlinear complexity. Ensemble methods are designed to perform best with large, heterogeneous datasets containing multiple nonlinear relationships and feature interactions. In the present study, where the predictors were few and largely correlated, ensemble models added unnecessary model complexity that led to over-smoothing and a small increase in prediction error. Nonetheless, the XGBoost model performed marginally better than Random Forest, owing to its gradient boosting mechanism, which iteratively minimizes residual errors and adapts more efficiently to underlying data patterns. This behavior aligns with prior research suggesting that boosting-based algorithms tend to outperform bagging approaches when handling moderately nonlinear datasets with limited noise.

In contrast, the LSTM model produced the weakest results, with an R^2 of only 0.4328. Despite being theoretically well-suited for sequential data modeling, the LSTM network underperformed due to the dataset's restricted length and low temporal granularity. Deep learning architectures like LSTM require extensive historical data to identify long-term dependencies and seasonal trends. In this study, the model was trained on daily closing prices covering a limited time span, which provided insufficient temporal context for the network to learn meaningful patterns. Moreover, the absence of additional sequential features such as lagged returns, moving averages with different window sizes, or sentiment indicators from social media further constrained the model's ability to represent complex temporal dynamics. Previous studies have similarly shown that LSTM models tend to underperform in small-sample financial datasets, particularly when compared to simpler machine learning approaches.

The overall findings highlight a crucial trade-off between model complexity and data adequacy in financial forecasting. While advanced models such as LSTM and XGBoost are capable of capturing intricate nonlinearities and long-term dependencies, their advantages diminish when applied to limited datasets with weak temporal structures. In contrast, simpler models like Linear Regression can achieve superior performance under such conditions, as they efficiently capture dominant linear relationships without overfitting. This suggests that model selection in cryptocurrency forecasting should not rely solely on algorithmic sophistication but rather on the interplay between data quality, quantity, and the nature of market behavior. For highly volatile markets or larger datasets with strong sequential signals, deep learning architectures may offer significant improvements. However, for short-term forecasting with limited data, classical machine learning models remain highly competitive and, in some cases, preferable.

Conclusion

This study aimed to evaluate and compare the predictive performance of four machine learning algorithms, namely Linear Regression, Random Forest, XGBoost, and LSTM, in forecasting the closing price of the AAVE cryptocurrency. The findings revealed clear variations in performance across the models, highlighting how algorithmic complexity and data characteristics jointly influence forecasting accuracy. Among the tested models, Linear Regression achieved the highest predictive accuracy, recording the lowest MAE and RMSE values and the highest R^2 score of 0.9924. These results indicate that the AAVE price dynamics during the study period were largely governed by linear relationships among core market indicators such as Open, High, Low, Volume, and Marketcap. The model's strong performance demonstrates that when cryptocurrency data displays stable and trend-oriented patterns, simple regression-based approaches can produce reliable forecasts with minimal error. The ensemble models, Random Forest and XGBoost, also showed strong predictive capabilities, with R^2 values of 0.9337 and 0.9484 respectively, suggesting that they effectively captured moderate nonlinearities in the dataset. However, their slightly higher error values indicate that while ensemble methods enhance robustness, they may introduce over-smoothing, reducing responsiveness to rapid market fluctuations.

In contrast, the LSTM deep learning model produced the weakest performance, with a notably lower R^2 value of 0.4328 and significantly higher prediction errors. This underperformance suggests that the model failed to capture meaningful temporal dependencies due to the limited dataset size and low data granularity. Deep learning models such as LSTM are inherently designed to learn from long- and detailed-time sequences, and their effectiveness depends on the availability of extensive historical and high-frequency data. The small sample size and absence of additional sequential features, such as lagged returns, sentiment data, or macroeconomic indicators, constrained the model's ability to generalize complex patterns in AAVE price movements. These findings underscore the importance of aligning model complexity with data characteristics. Simpler models like Linear Regression and ensemble methods are more effective when data availability is limited and relationships are relatively straightforward, while deep learning models are better suited for larger datasets with richer temporal information. Future research should expand the dataset by incorporating longer time periods, high-frequency trading data, and external variables such as investor sentiment or blockchain activity, as well as explore hybrid approaches that combine ensemble learning and deep learning to enhance prediction accuracy and model stability.

Declarations

Author Contributions

Conceptualization: M.Ł.; Methodology: M.Ł.; Software: M.Ł.; Validation: M.Ł.; Formal Analysis: M.Ł.; Investigation: M.Ł.; Resources: M.Ł.; Data Curation: M.Ł.; Writing Original Draft Preparation: M.Ł.; Writing Review and Editing: M.Ł.; Visualization: M.Ł.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Bouteska, M. Z. Abedin, P. Hajek, and K. Yuan, “Cryptocurrency price forecasting: A comparative analysis of ensemble learning and deep learning methods,” *International Review of Financial Analysis*, vol. 92, art. 103055, Mar. 2024, doi: 10.1016/j.irfa.2023.103055.

[2] F. Rodrigues and M. Machado, “High-frequency cryptocurrency price forecasting using machine learning models: A comparative study,” *Information*, vol. 16, no. 4, art. 300, Apr. 2025, doi: 10.3390/info16040300.

[3] P. L. Seabe, M. Kaya, and R. Budka, “Forecasting cryptocurrency prices using LSTM, GRU, and Bi-Directional LSTM,” *Fractal and Fractional*, vol. 7, no. 2, art. 203, 2023, doi: 10.3390/fractalfract7020203.

[4] H. Kırkgöz and O. Kurt, "Modeling bitcoin network energy demand: Price-adjusted hybrid deep learning approach to complex time series forecasting," *Chaos, Solitons & Fractals*, vol. 200, pt. 2, p. 117075, 2025, doi: 10.1016/j.chaos.2025.117075.

[5] P. Boozary, S. Sheykhan, and H. GhorbanTanhaei, "Forecasting the Bitcoin price using the various Machine Learning: A systematic review in data-driven marketing," *Systems and Soft Computing*, vol. 7, p. 200209, 2025, doi: 10.1016/j.sasc.2025.200209.

[6] R. Guo, T. Y. Kim, and J. Liu, "Estimating and forecasting Bitcoin daily prices using ARIMA-GARCH models," *Business Analytics Journal*, vol. 45, no. 1, pp. 11–23, 2024, doi: 10.1108/BAJ-05-2024-0027.

[7] Y. Wang, G. Andreeva, and B. Martin-Barragan, "Machine learning approaches to forecasting cryptocurrency volatility: Considering internal and external determinants," *International Review of Financial Analysis*, vol. 90, p. 102914, 2023, doi: 10.1016/j.irfa.2023.102914.

[8] H. Alnami, M. Mohzary, B. Assiri, and H. Zangoti, "An integrated framework for cryptocurrency price forecasting and anomaly detection using machine learning," *Applied Sciences*, vol. 15, no. 4, art. 1864, Feb. 2025, doi: 10.3390/app15041864.

[9] D. H. Kim, F. J. Vanheusden, and A. Kim, "Forecasting cryptocurrency markets using recurrence and time-frequency analysis-based machine learning algorithms," *Finance Research Letters*, vol. 85, no. November, art. 108268, 2025, doi: 10.1016/j.frl.2025.108268.

[10] S. Lee, D.-C. Vu, H.-H. Pham, and D. H. Tran, "Enhancing cryptocurrency forecasting using temporal features on high-frequency data," in *Intelligent Systems and Data Science (ISDS)*, *Communications in Computer and Information Science*, vol. 2714, no. October, pp. 343–356, 2025, doi: 10.1007/978-981-95-3358-9_25.

[11] "A hybrid ensemble approach for cryptocurrency price forecasting combining ARIMA, LSTM, and LightGBM," in *Smart Computing Paradigms: AI and Networking Applications*, LNNS, vol. 1749, no. November, pp. 423–434, Nov. 2024, doi: 10.1007/978-981-97-7880-5_36.

[12] S. Abossedgh, A. Yeganeh, and A. Johannssen, "Machine learning-driven feature selection and anomaly detection for Bitcoin price analysis," *Applied Soft Computing*, vol. 188, p. 114382, 2026, doi: 10.1016/j.asoc.2025.114382.

[13] R. Hasanli and M. Dursun, "Integrating high-dimensional technical indicators into machine learning models for predicting cryptocurrency price movements," *FinTech*, vol. 4, no. 4, art. 77, 2025, doi: 10.3390/fintech4040077.

[14] A. Karta Wijaya et al., "Optimizing Bitcoin price forecasting using multi-model approaches including LSTM, GRU, Prophet, VaR, and ES," *Jurnal Teknik Informatika*, vol. 6, no. 3, pp. 1095–1112, Jun. 2025, doi: 10.52436/1.jutif.2025.6.3.4078.

[15] V. Vahidpour et al., "A multi-layer machine learning approach for cryptocurrency price prediction," *International Journal of Intelligent Computing and Cybernetics*, vol. 18, no. 4, pp. 706–730, 2025, doi: 10.1108/IJICC-03-2025-0128.

[16] P. Jaquart, S. Köpke, and C. Weinhardt, "Machine learning for cryptocurrency market prediction and trading," *Journal of Finance and Data Science*, vol. 8, no. November, pp. 331–352, 2022, doi: 10.1016/j.jfds.2022.12.001.

[17] "Forecasting Bitcoin volatility using a hybrid CNN-LSTM model," *International Review of Financial Markets, Institutions and Money*, vol. 72, no. December, art. 102064, 2024, doi: 10.1016/j.intfin.2024.102064.

[18] P. Hiskiawan, J. William, and L. F. Tio Jansel, "A hybrid data science framework for forecasting Bitcoin prices using traditional and AI models," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2089–2101, 2025, doi: 10.30871/jaic.v9i5.10631.

[19] W. Kochliaridis, A. Papadopoulou, and I. Vlahavas, "A machine learning approach to cryptocurrency trading," *Applied Intelligence*, vol. 54, no. April, pp. 5688–5710, 2024, doi: 10.1007/s10489-024-05407-z.

[20] G. Dudek, M. Kasprzyk, and P. Pełka, "Multivariate forecasting of bitcoin volatility with gradient boosting: Deterministic, probabilistic, and feature importance perspectives," *Expert Systems with Applications*, vol. 301, p. 130404, 2026, doi: 10.1016/j.eswa.2025.130404.