applied sciences

MDPI

Article

An Integrated Framework for Cryptocurrency Price Forecasting and Anomaly Detection Using Machine Learning

Hani Alnami [1,*], Muhammad Mohzary [1,2], Basem Assiri [1] and Hussein Zangoti [3]

[1] Department of Computer Science, Jazan University, Jazan 82817, Saudi Arabia; mmohzary@jazanu.edu.sa (M.M.); babumussmar@jazanu.edu.sa (B.A.)
[2] Engineering and Technology Research Center, Jazan University, P.O. Box 114, Jazan 82817, Saudi Arabia
[3] Department of Electrical and Electronic Engineering, Jazan University, Jazan 82817, Saudi Arabia; hmzangoti@jazanu.edu.sa
* Correspondence: hmalnami@jazanu.edu.sa

Abstract: The accurate prediction of cryptocurrency prices is crucial due to the volatility and complexity of digital asset markets, which pose significant challenges to traders, investors, and researchers. This research addresses these challenges by leveraging machine learning and deep learning techniques to forecast closing prices for cryptocurrencies, focusing on Bitcoin, Ethereum, Binance Coin, and Litecoin cryptocurrency datasets. A Random Forest ensemble learning algorithm, a Gradient Boosting model, and a feedforward neural network were implemented to handle the complexities in cryptocurrency data. A Z-Score-based anomaly detection framework was integrated to classify closing prices as normal or abnormal, aiding in identifying significant market events. Evaluation metrics, such as the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2), demonstrate the superior precision and reliability of the Random Forest and Gradient Boosting models. The deep learning model indicates strong generalization capabilities, suggesting potential advantages on more complex datasets. These findings highlight the importance of combining advanced machine learning techniques and cryptocurrencies to develop a robust framework for cryptocurrency forecasting and anomaly detection.

Keywords: cryptocurrency prediction; abnormal close price prediction; machine learning

check for updates

Academic Editor: Gianluca Lax

Received: 18 December 2024
Revised: 31 January 2025
Accepted: 31 January 2025
Published: 11 February 2025

Citation: Alnami, H.; Mohzary, M.; Assiri, B.; Zangoti, H. An Integrated Framework for Cryptocurrency Price Forecasting and Anomaly Detection Using Machine Learning. *Appl. Sci.* **2025**, *15*, 1864. https://doi.org/10.3390/app15041864

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cryptocurrency markets have gained significant attention worldwide because they could change how we handle money. These digital assets are also known for their ups and downs in price, making them attractive but risky [1–4]. The ups and downs of cryptocurrency attract many investors, but they come with both big chances and serious risks. Because prices can change quickly, it is essential to create good models that predict these price changes. These models help reduce risks, support better decision making, and spot new trends in the market [5–7].

Traditional financial forecasting techniques frequently struggle to effectively represent cryptocurrencies' complex price fluctuations. These fluctuations are affected by multiple factors, such as non-linear dynamics, changes in external sentiment, and various elements inherent to the blockchain [8–12]. As a result, the drawbacks of traditional methods have prompted the development of state-of-the-art machine learning and deep learning techniques. These approaches have demonstrated effectiveness in addressing and understanding the complex challenges of predicting the cryptocurrency market, providing a more advanced analytical tool for investors and financial analysts.

This study contributes to the existing knowledge base by integrating machine learning and deep learning to forecast cryptocurrencies' closing prices and identify irregular market trends. This study used historical price data and essential market features such as trading volume and capitalization to develop robust predictive models. By leveraging Random Forest and Gradient Boosting algorithms, along with a feedforward neural network, this research addressed the non-linear patterns and dynamic nature of cryptocurrency data. Furthermore, an anomaly detection framework based on Z-Scores was incorporated to classify closing prices as normal or abnormal, allowing the identification of significant market events that could impact trading strategies.

A fundamental aspect of this research is enhancing cryptocurrencies by enabling early and transparent decisions based on accurate prediction, which helps to prevent or minimize harmful situations. This approach ensures the clarity, immutability, and real-time accessibility of abnormal trading data, thereby enhancing the reliability and utility of the predictive framework. This investigation provides a comprehensive solution for cryptocurrency price forecasting, anomaly detection, and market analysis. By addressing existing limitations in traditional financial forecasting methods, this research contributes to advancing the field of cryptocurrency analytics and offers practical insights for stakeholders in the digital asset ecosystem.

The key contributions of this research are summarized as follows:

- Integration of Advanced Technologies: A robust framework combines machine learning, deep learning, and cryptocurrency to improve price prediction and anomaly detection while ensuring transparency and data immutability.
- Robust Predictive Models: This research utilizes Random Forest, Gradient Boosting, and feedforward neural network models to tackle non-linear behaviors and evolving patterns in cryptocurrency data. It incorporates features such as trading volume and market capitalization to improve the accuracy of the models.
- Anomaly Detection Framework: A Z-score-based mechanism classifies closing prices as normal or abnormal, providing insights into significant market events to support trading strategies and risk management.
- Advancement of Cryptocurrency Analytics: This research offers practical tools for stakeholders to tackle the challenges of the cryptocurrency market by overcoming the limitations of traditional forecasting methods.

2. Literature Review

Several studies have utilized traditional machine learning algorithms to predict cryptocurrency prices. Linear regression (LR) models have been commonly applied to forecast Bitcoin's closing price using critical market features, as demonstrated in [5]. Additionally, other studies have employed LR and support vector machines (SVMs), along with sliding window techniques, to predict daily Bitcoin prices [13]. In [14], Chen et al. used high-dimensional feature classification models to predict daily price movements. Meanwhile, ensemble learning methods, including Random Forests, have been widely applied. For example, ref. [15] used random forest models to predict Bitcoin's closing price, outperforming models like LR and K-nearest neighbors (KNN). Furthermore, ensemble learning methods, such as gradient boosting, have demonstrated superior performance in cryptocurrency forecasting using XGBoost and LightGBM [7]. Moreover, ref. [16] employed several machine learning models, such as recurrent neural networks and tree-based ensembles, to predict the daily movements of more than 100 cryptocurrencies with an accuracy of 59.5% at most. Additionally, a recent study [17] forecasted Ethereum (ETH) price and trends by applying Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) models for price predictions and support vector machines (SVM) for trend classification. The authors achieved a return factor of up to 5.16%.

Time-series forecasting approaches have also been employed in cryptocurrency price prediction. Traditional econometric models like auto-regressive integrated moving average (ARIMA) have been compared with deep learning models. For example, Rebane et al. [18] compared ARIMA with recurrent neural network (RNN)-based seq2seq models and found that neural networks outperformed ARIMA. Additionally, ARIMA models have been applied alongside modern techniques such as FBProphet and XGBoost for predicting fluctuations in Bitcoin prices, as demonstrated in [19]. The results indicated that ARIMA achieved the lowest error rates, making it a strong candidate for cryptocurrency time-series forecasting. Econometric time-series models like Auto-Regressive Conditional Heteroskedasticity (ARCH) [20] and Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) [21] handle heteroskedasticity (non-constant variance) over time to simulate volatility. In addition, Auto-Regressive Integrated Moving Average (ARIMA) [22] has also been used to forecast non-stationary series. Together, these methods form some of the most modern time-series forecasting models [23].

Deep learning algorithms, particularly recurrent neural networks (RNNs), have shown impressive results in cryptocurrency price forecasting. Long Short-Term Memory (LSTM) networks have been a popular choice for handling time-series data due to their ability to capture temporal dependencies. Several studies have applied LSTM networks to predict cryptocurrency prices, outperforming traditional machine learning models. For instance, an LSTM model achieved superior accuracy over linear regression in [5], and other studies have utilized LSTM to predict daily cryptocurrency prices [18]. Moreover, Sezer et al. applied Convolutional Neural Networks (CNNs) to time-series data, transforming price trends into 2D images, which helped capture short-term price dependencies [24]. Additionally, Zhang et al. demonstrated the potential of transformer models like BERT and GPT for price prediction, showing that transformers outperformed RNNs by capturing both short- and long-term dependencies [25].

Hybrid models that combine traditional and deep learning methods have also been proposed for cryptocurrency price prediction. For example, Chong et al. developed a hybrid ARIMA-LSTM model to combine linear and non-linear trends in price data, demonstrating improved accuracy in price forecasting [1]. Similarly, in [6], researchers integrated various cryptocurrency data sources to predict Ethereum prices using advanced machine learning techniques. These models have been shown to benefit from both traditional econometric and modern neural network features. Another research evaluated the use of machine learning in predicting price volatility for cryptocurrency [26]. It compared the performances of linear regression, decision trees, support vector machines, and neural networks, showing promising results in price prediction based on the US financial market. Another study investigated the performance of Gradient Boosting, Random Forest, and Bagging in predicting the prices of cryptocurrencies, such as Bitcoin, Ethereum, Binance, USD, and others. The work showed that Gradient Boosting outperforms the other machine learning models in some currencies, while Random Forest outperforms them in other currencies [27]. More studies confirm the effectiveness of the Random Forest regression algorithm in cryptocurrency price prediction [28–30].

Indeed, several works have focused on a specific cryptocurrency such as Bitcoin to be the base model that can be generalized [30–34]. While others have expanded the research to include more cryptocurrencies [35,36].

3. Methodology and Architecture

This research examined the closing prices of cryptocurrencies—specifically Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), and Litecoin (LTC)—utilizing a Random Forest machine learning algorithm. As demonstrated in Figure 1, the proposed methodology for forecasting closing prices adopts a multi-stage approach that integrates data transformation, machine learning, and cryptocurrencies. Historical price data were initially sourced from various platforms, including stock exchanges and financial APIs. Following data collection, this dataset underwent preprocessing to ensure its quality and relevance, involving cleaning, normalizing, and transforming data into an appropriate format for analysis. Additionally, key features such as trading volume and market capitalization were engineered to enhance the accuracy of the predictive model.

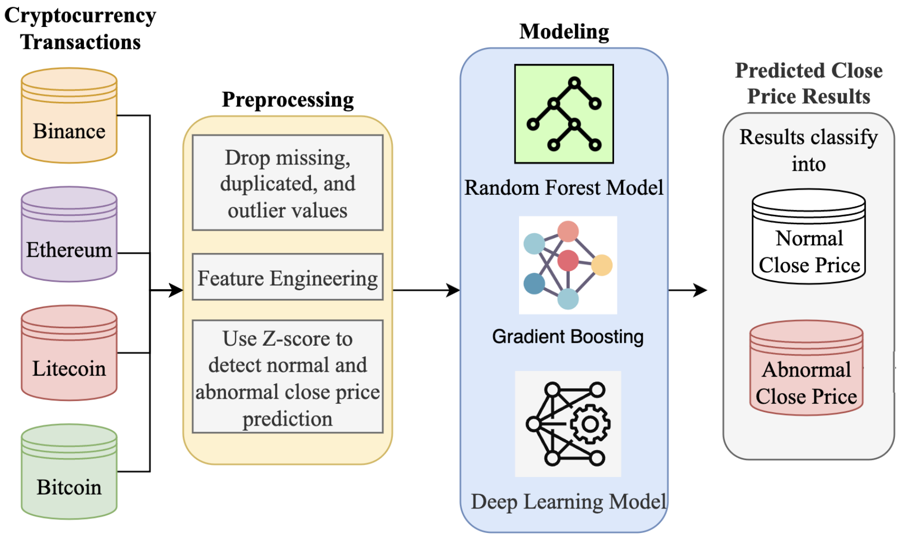

Figure 1. Proposed methodology.

Once the dataset was prepared, the ensemble machine learning algorithm was created to forecast prices. The Random Forest algorithm was chosen for its robustness against overfitting and adeptness at handling non-linear relationships [37]. To determine the model's effectiveness, metrics such as the Mean Absolute Error and Root Mean Squared Error were utilized to evaluate its predictive accuracy.

The methodology employs an anomaly detection framework to identify normal versus abnormal closing price scenarios. Predicted closing prices were consistently compared to actual closing prices to create a baseline of expected behavior. This involved calculating the standard deviation of the predicted closing prices to set a threshold for categorizing them. Specifically, a threshold was defined as the mean plus or minus one standard deviation, allowing predicted closing prices to be classified as either normal or abnormal. Furthermore, statistical tools like Z-scores were used to further classify predictions into normal (within an acceptable range) and abnormal (outliers) categories. Abnormal cases may signal significant market events or anomalies that should be investigated, while typical cases reflect standard market variations. The acceptable Z-score range was from -1 to 1 percent, so if the predicted closing price for the next day lied within this range, it was deemed normal; otherwise, it was classified as abnormal.

4. Dataset Description and Preprocessing

4.1. Description

In the initial phase, data collection involved gathering historical data from reliable cryptocurrency exchanges, such as Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), and Litecoin (LTC). We included these cryptocurrencies to capture a broad spectrum of market conditions, volatility, and different use cases. Bitcoin serves as the market benchmark with the largest market capitalization, and Ethereum serves as that with the second market capitalization that supports smart contracts technology. Binance Coin illustrates an exchange-based utility token, and Litecoin offers fast and lightweight transactions [38]. These varied digital assets provide robust historical data and diverse market dynamics that can enable comprehensive evaluations of price forecasting and anomaly detection. The dataset encompasses critical features, including closing prices, opening prices, high and low prices, trading volume, and market capitalization. The data collection time frame spanned 2015 until 2021 to ensure a comprehensive analysis of price trends [39]. The following are brief descriptions of the dataset's features:

- Date: Represents the observation date and serves as the dataset's temporal index.
- Open: Captures the opening price of the cryptocurrency, reflecting the first transaction at the start of the trading day.
- High: Denotes the highest price achieved within the selected time frame, providing insight into market volatility on the date.
- Low: Represents the lowest price traded during the interval, helping assess the minimum demand and price fluctuation range compared to the high feature.
- Close (Target): Records the final price at the end of the day.
- Volume: Indicates the total of cryptocurrency units traded. Higher volumes suggest greater market interest, while lower volumes may indicate uncertainty.
- Market Cap: Calculated as the closing price times circulating supply, it measures the cryptocurrency's overall market value, used for ranking. A higher market cap suggests stability; a lower one may indicate riskier investments.

4.2. Preprocessing

Following data collection, preprocessing steps were undertaken to enhance the dataset's quality. This process includes data cleaning to address missing values and duplicates. Feature engineering also uses correlation feature selection to determine variables that could improve model performance, such as volume and market capitalization, as presented in Figure 2. Additionally, the standard scaler preprocessing technique was employed to normalize all cryptocurrency datasets, adjusting them to have a mean of 0 and a standard deviation of 1. This step is essential for enhancing the effectiveness of machine learning algorithms [40]. The threshold value for labeling abnormal values is derived from the Z-Score, a statistical measure that quantifies the number of standard deviations a data point is from the mean of a dataset. The Z-Score is calculated using a rolling window approach, where a rolling mean and standard deviation are computed over a specified period (in this case, 30 days). This ensures that the Z-Score dynamically adjusts to local trends and variations in the data, making it suitable for time-series analysis. The decision to use a 30-day rolling window aims to balance the identification of medium-term trends and outliers while reducing the noise from short-term fluctuations. This time frame corresponds with the typical expectation that financial market patterns may display cyclical trends over monthly periods. We acknowledge that this window size might not adequately represent sudden changes. Consequently, additional tests were performed using rolling windows of 15 and 7 days to assess their effects on the results. We recalculated the rolling mean and standard deviation for these new window sizes and analyzed how these changes impacted the Z-Score, anomaly detection, and overall model performance.

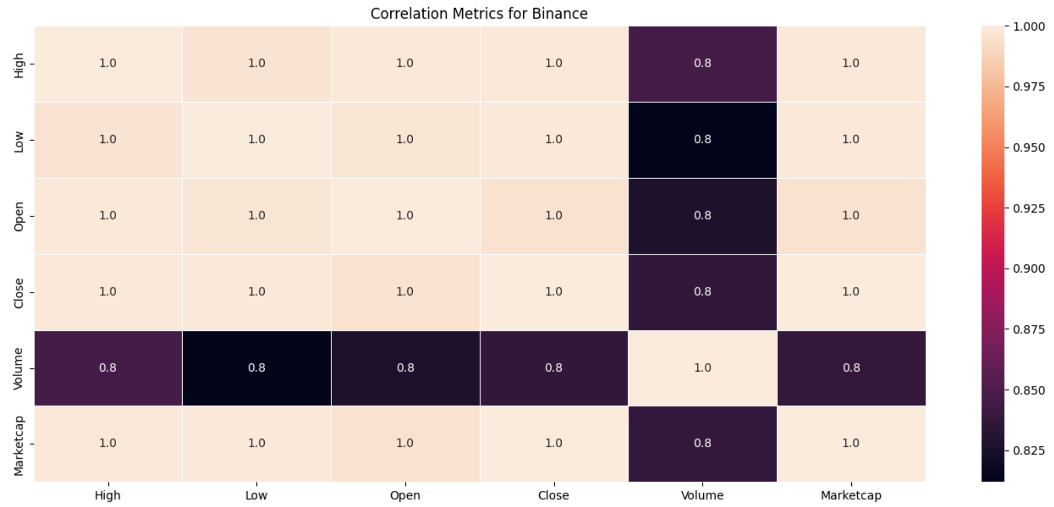

Figure 2. Correlation metrics.

Algorithm 1 identifies abnormal data points based on their deviation from recent trends using a financial instrument's closing prices $C(t)$. The analysis was carried out by computing rolling statistics and normalizing the data using the Z-Score. The algorithm relies on three key components: the rolling mean $\mu(t)$, rolling standard deviation $\sigma(t)$, and Z-Score $Z(t)$. The rolling mean $\mu(t)$ represents the local average of the time-series data over a specified window size w, smoothing short-term fluctuations to highlight longer-term trends. Similarly, the rolling standard deviation $\sigma(t)$ quantifies the dispersion of the data around the rolling mean, capturing the variability within the same window. Using these rolling statistics, the algorithm adapts to local trends and avoids the influence of earlier data, which might no longer be relevant. The Z-Score $Z(t)$ is a normalized measure that quantifies the deviation of a data point from the rolling mean in units of the rolling standard deviation. This normalization enables a consistent comparison of deviations across different points in time, even when the scale or variability of the data changes. The algorithm detects abnormalities by evaluating whether the absolute value of the Z-Score exceeds a predefined threshold $z_{\text{threshold}}$. Data points for which $|Z(t)| > z_{\text{threshold}}$ are flagged as abnormal, and an abnormality indicator $A(t)$ is assigned a value of 1, while normal points are assigned a value of 0. This approach allows for the detection of significant deviations, regardless of whether they are above or below the rolling mean. The rolling window approach ensures that the algorithm is sensitive to recent data, making it well suited for applications where trends and patterns evolve over time. Common use cases include financial data analysis, where sudden price movements can signal market anomalies, or sensor data monitoring, in which deviations may indicate faults or unusual events. However, the method requires sufficient data points within the rolling window to provide reliable statistics, and the presence of outliers within the window may influence the computed rolling mean and standard deviation.

Overall, the algorithm provides a robust framework for anomaly detection. It ensures adaptability across various domains and datasets by leveraging local statistics and normalization.

The dataset was then divided into training and testing subsets, allocating approximately 80% of the data for training and 20% for testing to validate the model's performance.

Algorithm 1 Calculate Z-Score and detect abnormalities in cryptocurrency data

Input:

- $C(t)$: Time-series data of closing prices.
- w: Rolling window size ($w = 30$).
- $z_{\text{threshold}}$: Z-Score threshold ($z_{\text{threshold}} = 1$).

Output:

- $\mu(t)$: Rolling mean of $C(t)$.
- $\sigma(t)$: Rolling standard deviation of $C(t)$.
- $Z(t)$: Z-Score of $C(t)$.
- $A(t)$: Abnormality indicator; $A(t) = 1$ if $|Z(t)| > z_{\text{threshold}}$, otherwise $A(t) = 0$.

Step 1: Calculate Rolling Mean
for each time point t in $C(t)$ **do**
  Compute the rolling mean $\mu(t)$ over a window of size w:

$$\mu(t) = \frac{1}{w} \sum_{i=t-w+1}^{t} C(i)$$

end for
Step 2: Calculate Rolling Standard Deviation
for each time point t in $C(t)$ **do**
  Compute the rolling standard deviation $\sigma(t)$:

$$\sigma(t) = \sqrt{\frac{1}{w} \sum_{i=t-w+1}^{t} (C(i) - \mu(t))^2}$$

end for
Step 3: Calculate Z-Score
for each time point t in $C(t)$ **do**
  Compute the Z-Score $Z(t)$:

$$Z(t) = \frac{C(t) - \mu(t)}{\sigma(t)}$$

end for
Step 4: Identify Abnormalities
for each time point t in $C(t)$ **do**
  if $|Z(t)| > z_{\text{threshold}}$ **then**
    Set $A(t) = 1$ {Mark as abnormal}
  else
    Set $A(t) = 0$ {Normal point}
  end if
end for

5. Modeling and Evaluation

In financial forecasting, accurately predicting cryptocurrency prices is crucial due to the volatile and dynamic characteristics of digital assets. This study investigated the use of machine learning and deep learning models to forecast closing prices of cryptocurrencies, specifically emphasizing Random Forest and feedforward neural network techniques. Both models were assessed for their capacity to manage the complexities and non-linear trends present in financial datasets, utilizing important features like ‘Volume’ and ‘Marketcap’ for prediction. The methodologies along with their implementation details are provided below.

This study employed the Random Forest (RF) model and Gradient Boosting (GB) regressor to predict the closing prices of various cryptocurrencies, including Binance, Ethereum, and Litecoin. The RF model is a robust ensemble learning algorithm frequently utilized for regression and classification, recognized for its ability to handle non-linear relationships and reduce overfitting. It operates by creating 100 decision trees during the training phase and consolidating their predictions to achieve a more accurate and reliable result. Key input features, such as "Volume" and "Marketcap", were used for model training, with the target variable being the "Close" price. Conversely, the GB regressor constructs trees sequentially, with each tree addressing the residual errors of the previous one. This iterative refinement process renders Gradient Boosting especially adept at detecting complex patterns and attaining high accuracy in numerous regression tasks. Both models were trained on 80% of the dataset, with predictions evaluated on the remaining 20% of the test set. To measure accuracy and reliability, we calculated metrics, including the Mean Squared Error, the Root Mean Squared Error, the Mean Absolute Error, and R-squared [41]. The following are the descriptions of each metric:

Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- n: Total number of data points.
- y_i: Actual value of the i-th data point.
- $\hat{y}_i$: Predicted value of the i-th data point.
- $(y_i - \hat{y}_i)^2$: Squared difference between the actual and predicted values.

The MSE measures the average of the squared differences between predicted and actual values, emphasizing larger errors.

Root Mean Squared Error (RMSE)

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

- n: Total number of data points.
- y_i: Actual value of the i-th data point.
- $\hat{y}_i$: Predicted value of the i-th data point.
- $(y_i - \hat{y}_i)^2$: Squared difference between the actual and predicted values.

The RMSE is the square root of the MSE and provides an error measure in the same units as the target variable.

Mean Absolute Error (MAE)

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

- n: Total number of data points.
- y_i: Actual value of the i-th data point.
- $\hat{y}_i$: Predicted value of the i-th data point.
- $|y_i - \hat{y}_i|$: Absolute difference between the actual and predicted values.

MAE measures the average magnitude of errors without considering their direction.

R-squared (R^2)

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

- n: Total number of data points.
- y_i: Actual value of the i-th data point.
- $\hat{y}_i$: Predicted value of the i-th data point.

- $\bar{y}$: Mean of all actual values.
- $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$: Sum of squared residuals (errors).
- $\sum_{i=1}^{n}(y_i - \bar{y})^2$: Total sum of squares (variance in the actual data).

R^2 measures the proportion of the variance in the actual data that the model explains. It ranges from 0 to 1, with higher values indicating better model performance.

This analysis utilizes a deep learning model featuring a feedforward neural network to forecast the closing prices of Binance cryptocurrency, using crucial features like "Volume" and "Marketcap". To enhance the model's performance, input features were standardized with a StandardScaler, normalizing the data to a mean of zero and a standard deviation of one, which aids convergence in training. The neural network was built with the Keras Sequential API and included three hidden layers containing 64, 32, and 16 units, each applying the Rectified Linear Unit (ReLU) activation function to identify non-linear patterns. The output layer is a sole neuron with a linear activation function specifically designed for regression tasks. The model employs the Adam optimizer, which is recognized for its adaptive learning rate. To reduce prediction errors, the loss function was set to Mean Squared Error (MSE). During training, the model was fit for 40 epochs with a batch size of 32, and validation data were used to evaluate its performance.

6. Results and Discussion

The evaluation results for the Random Forest, Gradient Boosting, and Deep Learning models show unique strengths in predicting cryptocurrency prices across various datasets. Figure 3 illustrates the RMSE and MAE outcomes for the three models—Random Forest, Deep Learning, and Gradient Boosting—applied to four cryptocurrency datasets: Binance, Bitcoin, Ethereum, and Litecoin. Since all RMSE and MAE values are below 0.15, this indicates that all models perform with a high degree of accuracy, regardless of the variations in metric values among the models.

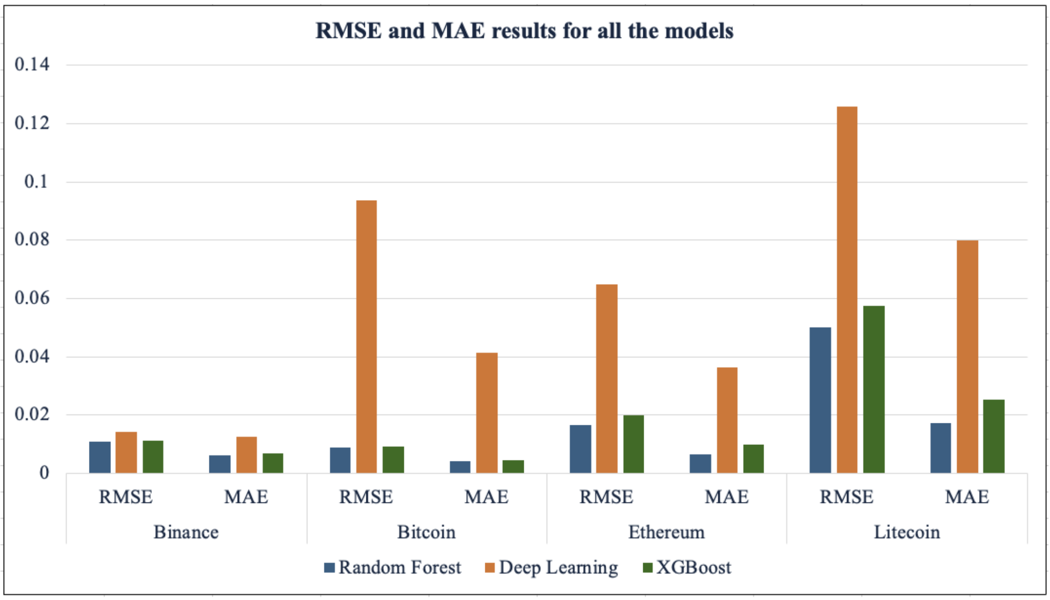

Figure 3. RMSE and MAE results for all the models utilized.

On the Binance dataset, all three models show low RMSE and MAE values, indicating precise performance. RF and GB yield nearly identical results, highlighting their effectiveness in capturing the underlying patterns of Binance's market data. Although Deep Learning presents slightly higher error values than the other models, it still attains metrics very close to zero, affirming its capacity to model Binance's relatively stable market patterns accurately.

On the Bitcoin dataset, all models delivered precise predictions, reflected in their low RMSE and MAE values. While Deep Learning exhibits higher error rates than Random Forest and GB, these errors are still minor, indicating the model's dependable performance. Random Forest and GB present slightly lower errors, highlighting their strength in grasping the complexities of Bitcoin's volatile market. Nevertheless, the substantial fact that all errors are well below 0.15 confirms that each model is capable of accurately predicting Bitcoin prices.

The models for the Ethereum dataset again demonstrate high accuracy. Both Random Forest and GB achieve low, comparable RMSE and MAE values, highlighting their strong predictive power for Ethereum's market. Although Deep Learning shows slightly higher error metrics, these figures remain close to zero, indicating that the model accurately reflects Ethereum's price trends. The overall low metric values across all models signify that Ethereum's market is highly predictable, despite minor variations in error.

All models exhibit consistent accuracy on the Litecoin dataset, reflected in their low RMSE and MAE values. Although Deep Learning records the highest error rates, the values are still comfortably below 0.15, indicating that the model is capable of delivering dependable predictions. Both Random Forest and GB attain low error rates and show marginally better performance, showcasing their ability to adapt to Litecoin's market trends. Despite variations in error rates, all models attain near-zero metrics, underlining their precision and reliability in forecasting Litecoin prices.

Table 1 compares the performances of all the algorithms across the four datasets. Performance was evaluated using the MSE, RMSE, MAE, and (R^2). Below is a detailed analysis of the models for each dataset.

Table 1. Performance metrics for Random Forest (RF), Gradient Boosting (GB), and Deep Learning (DL) models on all the datasets.

Dataset	Algorithm	MSE	RMSE	MAE	R^2
Binance	RF	0.0001	0.0110	0.0062	0.9998
	GB	0.0001	0.0112	0.0070	0.9998
	DL	0.0002	0.0144	0.0125	0.9996
Ethereum	RF	0.0002	0.0167	0.0067	0.9995
	GB	0.0004	0.0201	0.0098	0.9993
	DL	0.0042	0.0648	0.0364	0.9937
Litecoin	RF	0.0025	0.0501	0.0172	0.9972
	GB	0.0032	0.0574	0.0252	0.9963
	DL	0.0158	0.1258	0.0799	0.9825
Bitcoin	RF	0.000084	0.0091	0.0041	0.9998
	GB	0.000097	0.0098	0.0045	0.9998
	DL	0.0087	0.0936	0.0413	0.9879

All three models performed remarkably well on the Binance dataset, achieving MSE values that are nearly zero and R^2 values close to 1. Both RF and GB exhibit an MSE of 0.0001 and an R^2 of 0.9998, showing that they can predict Binance prices with nearly perfect accuracy. Meanwhile, DL has a slightly higher MSE of 0.0002 and an R^2 of 0.9996, indicating a minor decrease in its ability to fit the data compared to RF and GB. Nevertheless, the differences are negligible, and all models exhibit outstanding accuracy for this dataset.

On the Ethereum dataset, RF surpasses the other models with an MSE of 0.0002 and an R^2 of 0.9995, highlighting its effectiveness in capturing the data's underlying patterns.

GB follows closely with a slightly higher MSE of 0.0004 and an R^2 of 0.9993, demonstrating strong performance but falling just short of RF. Conversely, DL shows an increase in MSE to 0.0042 and a lower R^2 of 0.9937; nonetheless, these results remain highly accurate.

For Litecoin, RF delivers the strongest performance, achieving an MSE of 0.0025 and an R^2 of 0.9972. This signifies highly accurate predictions and an excellent fit to the data. Following closely is GB, which records an MSE of 0.0032 and an R^2 of 0.9963, slightly trailing RF yet still performing admirably. DL, on the other hand, exhibits lower performance with a higher MSE of 0.0158 and a reduced R^2 of 0.9825. Nevertheless, the error remains below zero, demonstrating the effective performance of the DL.

On the Bitcoin dataset, RF and GB record the best MSEs at 0.000084 and 0.000097, respectively, surpassing DL, which has an MSE of 0.0087. This signifies that RF and GB offer more accurate predictions than DL based on the MSE. In addition, RF and GB also reach an impressive near-perfect R^2 value of 0.9998, indicating that they closely align with the overall data trends. The DL's R^2 value is 0.9879, slightly lower than those of RF and GB.

The Random Forest model exhibits outstanding efficacy in predicting abnormal closing prices from normal closing prices over different rolling window periods (7, 15, and 30 days), attaining an accuracy of 100% across all four cryptocurrency test datasets, utilizing metrics such as precision, recall, F1-score, and overall accuracy [42]. This remarkable performance signifies that the model accurately classified all instances within the test dataset, where the total number of normal cases in the Binance test data was 126, whereas there were 167 abnormal cases. In the Litecoin dataset, 301 normal cases were recorded, while the abnormal cases amounted to 292. In the Ethereum dataset, there were 194 normal cases compared to 233 abnormal cases. Lastly, in the Bitcoin dataset, 296 abnormal cases were recorded, while the normal cases were counted to be 300.

7. Conclusions

This research introduced a comprehensive framework for predicting cryptocurrency prices and detecting anomalies through the integration of advanced machine learning, deep learning, and cryptocurrencies. Utilizing Random Forest, Gradient Boosting, and feedforward neural network models, this study addressed the inherent complexities and non-linearities in cryptocurrency data. Evaluation metrics such as the MSE, RMSE, MAE, and R^2 demonstrated that the Random Forest model consistently outperformed the other models in terms of precision and robustness for the Binance, Ethereum, and Litecoin datasets. Gradient Boosting performed closely to Random Forest, delivering strong predictions with slightly higher error metrics. Notably, the deep learning model achieved the lowest MSE on the Bitcoin dataset, showcasing its potential to handle certain complex datasets effectively.

This research also made a significant contribution by developing a Z-Score-based anomaly detection framework for cryptocurrencies, effectively categorizing closing prices as normal or abnormal. This detection mechanism provides crucial insights into important market events, equipping traders and investors with an additional layer of decision-making support. This guarantees transparency, immutability, and real-time access for documenting and analyzing anomalies.

Future research could further explore extending this framework to incorporate additional data sources, such as sentiment analysis from social media or news platforms, to enhance the models' predictive accuracy and reliability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app15041864/s1.

Author Contributions: H.A. and B.A.; methodology, H.A. and M.M.; validation, H.A., B.A., H.Z. and M.M.; formal analysis, H.A., M.M., B.A. and H.Z.; investigation, H.A. and M.M.; resources, H.A., M.M., B.A. and H.Z.; data Analysis, H.A. and M.M.; writing—original draft preparation, H.A.; writing—review and editing, H.A., M.M., B.A. and H.Z.; visualization, H.A.; supervision, H.A.; project administration, H.A.; and funding acquisition, H.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the funding of the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia, through project number: (RG24-S0149).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used are cited, and the code is openly available, submitted as Supplementary Materials.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chong, E.; Han, C.; Park, F.C. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Syst. Appl.* **2017**, *83*, 187–205. [CrossRef]
2. Smutny, Z.; Sulc, Z.; Lansky, J. Motivations, barriers and risk-taking when investing in cryptocurrencies. *Mathematics* **2021**, *9*, 1655. [CrossRef]
3. Sun, W.; Dedahanov, A.T.; Shin, H.Y.; Kim, K.S. Switching intention to crypto-currency market: Factors predisposing some individuals to risky investment. *PLoS ONE* **2020**, *15*, e0234155. [CrossRef]
4. Wang, Y. Do cryptocurrency investors in the UK need more protection? *J. Financ. Regul. Compliance* **2024**, *32*, 230–249. [CrossRef]
5. Vaddi, L.; Neelisetty, V.; Vallabhaneni, B.C.; Prakash, K.B. Predicting crypto currency prices using machine learning and deep learning techniques. *Int. J.* **2020**, *9*, 6603–6608. [CrossRef]
6. Kim, H.M.; Bock, G.W.; Lee, G. Predicting Ethereum prices with machine learning based on Blockchain information. *Expert Syst. Appl.* **2021**, *184*, 115480. [CrossRef]
7. Yuan, Z. Gold and Bitcoin Price Prediction based on KNN, XGBoost and LightGBM Model. *Highlights Sci. Eng. Technol.* **2023**, *39*, 720–725. [CrossRef]
8. Assiri, B.; Khan, W.Z. Fair and trustworthy: Lock-free enhanced tendermint blockchain algorithm. *TELKOMNIKA (Telecommun. Comput. Electron. Control)* **2020**, *18*, 2224–2234. [CrossRef]
9. Assiri, B.; Khan, W.Z. Enhanced and lock-free tendermint blockchain protocol. In Proceedings of the 2019 IEEE International Conference on Smart Internet of Things (SmartIoT), Tianjin, China, 9–11 August 2019; pp. 220–226.
10. Assiri, B. A modified and effective blockchain model for e-healthcare systems. *Appl. Sci.* **2023**, *13*, 12630. [CrossRef]
11. Assiri, B. Leader election and blockchain algorithm in cloud environment for e-health. In Proceedings of the 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), Amman, Jordan, 9–11 October 2019; pp. 1–6.
12. Assiri, B.; Alnami, H. Smart Partitioned Blockchain. *Sensors* **2024**, *24*, 4033. [CrossRef]
13. Poongodi, M.; Sharma, A.; Vijayakumar, V.; Bhardwaj, V.; Sharma, A.P.; Iqbal, R.; Kumar, R. Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. *Comput. Electr. Eng.* **2020**, *81*, 106527.
14. Chen, Z.; Li, C.; Sun, W. Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *J. Comput. Appl. Math.* **2020**, *365*, 112395. [CrossRef]
15. Talaat, F.M. An Enhanced Deep Learning Technique to Measure the Impact of Cryptocurrency on the World Payment system using Random Forest. *Am. J. Bus. Oper. Res.* **2022**, *8*, 8–15.
16. Jaquart, P.; Köpke, S.; Weinhardt, C. Machine learning for cryptocurrency market prediction and trading. *J. Financ. Data Sci.* **2022**, *8*, 331–352. [CrossRef]
17. Viéitez, A.; Santos, M.; Naranjo, R. Machine learning Ethereum cryptocurrency prediction and knowledge-based investment strategies. *Knowl.-Based Syst.* **2024**, *299*, 112088. [CrossRef]
18. Rebane, J.; Karlsson, I.; Papapetrou, P.; Denic, S. Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: A comparative study. In Proceedings of the SIGKDD Fintech'18, London, UK, 19–23 August 2018.
19. Iqbal, M.; Iqbal, M.; Jaskani, F.; Iqbal, K.; Hassan, A. Time-series prediction of cryptocurrency market using machine learning techniques. *EAI Endorsed Trans. Creat. Technol.* **2021**, *8*, e4. [CrossRef]
20. Bollerslev, T. Glossary to arch (garch). *CREATES Res. Pap.* **2008**, *49*, 1–46. [CrossRef]
21. Horv, L.; Kokoszka, P. GARCH processes: Structure and estimation. *Bernoulli* **2003**, *9*, 201–227.

22. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A comparison of ARIMA and LSTM in forecasting time series. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1394–1401.
23. Sung, S.H.; Kim, J.M.; Park, B.K.; Kim, S. A Study on Cryptocurrency Log-Return Price Prediction Using Multivariate Time-Series Model. *Axioms* **2022**, *11*, 448. [CrossRef]
24. Sezer, O.B.; Gudelek, M.U.; Ozbayoglu, A.M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Appl. Soft Comput.* **2020**, *90*, 106181. [CrossRef]
25. Li, Y.; Liu, M.; Cao, J.; Wang, X.; Zhang, N. Multi-attribute group decision-making considering opinion dynamics. *Expert Syst. Appl.* **2021**, *184*, 115479. [CrossRef]
26. Islam, M.Z.; Islam, M.S.; Al Montaser, M.A.; Rasel, M.A.B.; Bhowmik, P.K.; Dalim, H.M. Evaluating the effectiveness of machine learning algorithms in predicting cryptocurrency prices under market volatility: A study based on the usa financial market. *Am. J. Manag. Econ. Innov.* **2024**, *6*, 15–38. [CrossRef]
27. Samson, T.K. Comparative analysis of machine learning algorithms for daily cryptocurrency price prediction. *Inf. Dyn. Appl.* **2024**, *3*, 64–76. [CrossRef]
28. Derbentsev, V.; Datsenko, N.; Babenko, V.; Pushko, O.; Pursky, O. Forecasting cryptocurrency prices using ensembles-based machine learning approach. In Proceedings of the 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), Kharkiv, Ukraine, 6–9 October 2020; pp. 707–712.
29. Farouk, M.; Shaker, N.; AbdElminaam, D.S.; Elrashidy, O.; Mandour, L.; Mesbah, M.; Walid, J.; Ahmed, M.; Attia, R.; Ahmed, N.; et al. Bitcoin_ML: An Efficient Framework for Bitcoin Price Prediction Using Machine Learning. *J. Comput. Commun.* **2024**, *3*, 70–87. [CrossRef]
30. Pan, L. Cryptocurrency price prediction based on ARIMA, random forest and LSTM algorithm. *BCP Bus. Manag.* **2023**, *38*, 3396–3404. [CrossRef]
31. Sinha, H. Predicting Bitcoin prices using machine learning techniques with historical data. *Int. J. Creat. Res. Thoughts* **2024**, *12*, c760–c769.
32. Khedr, A.M.; Arif, I.; El-Bannany, M.; Alhashmi, S.M.; Sreedharan, M. Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Intell. Syst. Account. Financ. Manag.* **2021**, *28*, 3–34. [CrossRef]
33. Chen, J. Analysis of bitcoin price prediction using machine learning. *J. Risk Financ. Manag.* **2023**, *16*, 51. [CrossRef]
34. Alamery, F.M.S. Cryptocurrency analysis using machine learning and deep learning approaches. *J. Comput. Electr. Electron. Eng. Sci.* **2023**, *1*, 29–33. [CrossRef]
35. Srivastava, V.; Dwivedi, V.K.; Singh, A.K. Cryptocurrency price prediction using enhanced PSO with extreme gradient boosting algorithm. *Cybern. Inf. Technol.* **2023**, *23*, 170–187. [CrossRef]
36. Fang, F.; Chung, W.; Ventre, C.; Basios, M.; Kanthan, L.; Li, L.; Wu, F. Ascertaining price formation in cryptocurrency markets with machine learning. *Eur. J. Financ.* **2024**, *30*, 78–100. [CrossRef]
37. Parmar, A.; Katariya, R.; Patel, V. A review on random forest: An ensemble classifier. In *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 758–763.
38. Bhosale, J.; Mavale, S. Volatility of select crypto-currencies: A comparison of Bitcoin, Ethereum and Litecoin. *Annu. Res. J. SCMS Pune* **2018**, *6*, 132–141.
39. Cryptocurrency Historical Prices—kaggle.com. Available online: https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory (accessed on 10 September 2024).
40. Scikit-Learn Developers. sklearn.preprocessing.StandardScaler—Scikit-Learn 1.6.0 Documentation. 2025. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (accessed on 15 January 2025).
41. Tatachar, A.V. Comparative assessment of regression models based on model evaluation metrics. *Int. J. Innov. Technol. Explor. Eng.* **2021**, *8*, 853–860.
42. Tharwat, A. Classification assessment methods. *Appl. Comput. Informatics* **2021**, *17*, 168–192. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.