International Journal of
Distributed Computing and Technology

ISSN: 2455-7307
Volume 10, Issue 2, 2024
July—December
DOI (Journal): 0.37628/IJDCT

https://journalspub.com/journal/ijdct

Research IJDCT

Ethereum Price Prediction Using Extra Trees and Random Forest Regressors

Sushiladevi B. Vantamuri[1], Shreyas Salimath[2,*], Megha Mantur[3], Rutuja Bhosale[4], Misba Sindoli[5]

Abstract

The Ethereum market is famous for its unpredictability and the possibility of significant profits, presenting both opportunities and obstacles for investors and stakeholders. Given the dynamic nature of this ecosystem, accurate prediction of Ethereum's price movements is essential for informed decision-making. In this project, we propose an innovative approach that leverages ensemble learning techniques, specifically extra trees and random forest regressors, for Ethereum price prediction. By harnessing a diverse dataset encompassing historical Ethereum price data and a comprehensive array of market indicators such as trading volume, sentiment analysis, and technical metrics, we develop and evaluate our models through a rigorous methodology. Our study highlights the effectiveness of our approach in forecasting Ethereum prices with notable accuracy. Through extensive experimentation, we highlight the superior performance of ensemble learning techniques in capturing the intricate patterns inherent in Ethereum's price dynamics. Additionally, we conduct feature importance analysis to uncover the underlying factors driving Ethereum price movements, thereby providing valuable insights for investors and analysts seeking to navigate this complex market. The findings of our research make significant contributions to the expanding field of cryptocurrency price prediction, particularly within the context of Ethereum. Moreover, our project establishes a sturdy groundwork for continued investigation and enhancement of machine learning techniques in predicting cryptocurrency trends. This work aims to empower stakeholders with actionable intelligence, enabling them to make well-informed decisions amidst the nuances of the Ethereum market and the broader cryptocurrency landscape.

Keywords: Ethereum price prediction, cryptocurrency price prediction, learning techniques, machine learning algorithms

INTRODUCTION

In the ever-evolving landscape of cryptocurrency trading, accurate price predictions are highly coveted by investors and traders. As one of the top cryptocurrencies, Ethereum attracts considerable interest in this aspect. To meet the demand for insights into Ethereum's price movements, machine learning models have emerged as powerful tools, with extra trees and random forest regressors standing out among them. Both extra trees and random forest regressors belong to the ensemble learning category of machine learning algorithms, which combine multiple individual models to produce a stronger overall prediction. This approach trains on historical cryptocurrency data, aiming to offer investors more reliable predictions in the dynamic cryptocurrency market [1]. The rising significance of blockchain technology across public and private domains, underlining its

***Author for Correspondence**
Shreyas Salimath
E-mail: ssp.salimath@gmail.com

[1]Assistant Professor, Computer Science and Engineering, S.G. Balekundri Institute of Technology, Belagavi, Karnataka, India
[2-5]Students, Computer Science and Engineering, S.G. Balekundri Institute of Technology, Belagavi, Karnataka, India

Received Date: May 01, 2024
Accepted Date: May 26, 2024
Published Date: November 13, 2024

Citation: Sushiladevi B. Vantamuri, Shreyas Salimath, Megha Mantur, Rutuja Bhosale, Misba Sindoli. Ethereum Price Prediction Using Extra Trees and Random Forest Regressors. International Journal of Distributed Computing and Technology. 2024; 10(2): 9–25p.

potential to yield substantial business value in the future, as indicated by forecasts such as those outlined in Australia's National Blockchain Roadmap [2]. Beyond cryptocurrency, blockchain finds applications in diverse realms such as electronic healthcare and identity management systems [3–5]. The transition from physical to digital transactions in a dynamic technological environment underscores the importance of predicting currency value, influenced significantly by the controlling authority, particularly in the context of fiat currencies governed by national governments [6–9]. Blockchain as a disruptive force poised to transform numerous industries, drawing attention from policymakers, regulators, and various communities [10]. Rooted in Bitcoin, blockchain facilitates value transfer among anonymous participants independently, leveraging a combination of established technologies like digital signatures, proof-of-work mechanisms, and distributed systems [11]. Bitcoin's success as a cryptocurrency is rooted in its distinctive protocol and Nakamoto’s systematic structural design, aiming for full decentralization in contrast to fiat currencies [12]. Bitcoin's foundation on blockchain fosters trust among participants through cryptographic techniques, sparking research interest not only in economics but also in cryptography and machine learning [13]. Originating from Bitcoin, permissionless blockchain systems like Ethereum offer tamper-resistant ledgers held by each peer, facilitating the execution of Turing-complete smart contracts [14]. The rise of IoT in healthcare revolutionizes remote monitoring and treatment, enhancing patient care and reducing costs [15]. However, security concerns persist due to the vulnerability of medical systems to network attacks. Blockchain offers a decentralized solution, ensuring data integrity and addressing key management and authentication challenges [16]. Bitcoin is predominantly viewed as a speculative investment akin to internet stocks of the past century [17]. Its market capitalization surged to nearly 300 billion USD in 2017, garnering widespread attention and challenging traditional financial paradigms [18]. Since the advent of Bitcoin in 2009, cryptocurrencies have surged, with thousands of variants and a market cap surpassing $100 billion [19–22]. Yet, the proliferation of cryptocurrency exchanges has invited hackers, leading to substantial losses and security breaches, often exploiting social engineering techniques like phishing and trust-trading frauds [23].

LITERATURE REVIEW

The proliferation of online transactions has made credit card usage ubiquitous, but it has also escalated the risk of credit card fraud. Detecting fraudulent activities is crucial to mitigating financial losses for both consumers and financial institutions. This literature review explores the evolution of fraud detection methods, focusing on the transition from traditional machine learning approaches to state-of-the-art deep learning algorithms.

Traditional Machine Learning Approaches

Previous studies have extensively investigated various machine learning techniques for credit card fraud detection. Various approaches, including extreme learning method, decision trees, random forest, support vector machines, logistic regression, and XG boost, have been examined. While these methods have shown some success, they often suffer from low accuracy rates, highlighting the need for more advanced approaches.

Transition to Deep Learning

Recent advancements in deep learning have spurred interest in leveraging neural networks for fraud detection. Convolutional neural networks (CNNs) have attracted interest due to their capability to discern intricate patterns within intricate datasets. Researchers have applied CNN architectures to credit card fraud detection tasks, capitalizing on their capability to automatically learn features from raw data.

Empirical Analysis

Studies have conducted empirical analyses using benchmark datasets, such as the European card dataset, to evaluate the performance of deep learning models. Initially, machine learning algorithms were applied, demonstrating moderate improvements in fraud detection accuracy. Subsequently, CNN architectures were employed, leading to further enhancements in detection performance. Experiments involved tuning parameters, such as the number of hidden layers and epochs to optimize model performance.

Performance Evaluation

Metrics like accuracy, F1-score, precision, and area under the curve (AUC) have been employed to gauge the performance of suggested models. Results indicate significant improvements over traditional machine learning approaches, with optimized values reaching 99.9% accuracy, 85.71% F1-score, 93% precision, and 98% AUC. The proposed deep learning models outperform existing methods, showing their efficacy in real-world credit card fraud detection scenarios.

Mitigating False Negatives

To address the challenge of false negatives, experiments have been conducted to balance imbalanced data and fine-tune deep learning algorithms. These efforts aim to minimize the false negative rate, thereby enhancing the reliability of fraud detection systems for practical implementation.

The literature review underscores the changing terrain of credit card fraud detection, emphasizing a transition towards deep learning approaches. By leveraging the power of convolutional neural networks and conducting comprehensive empirical analyses, researchers have demonstrated significant advancements in fraud detection accuracy. The proposed deep learning approaches offer promising solutions for combating credit card fraud in real-world scenarios, paving the way for more robust and reliable detection systems.

In another literature review, the rise of electronic healthcare (e-health) systems is fueled by advancements in internet of things (IoT) technology, offering diverse options for medical data collection. Traditional authentication models face challenges in meeting the demands of low-latency, real-time services in e-health environments. While flexible communication services enhance data transmission options, they also raise security and privacy concerns. Addressing these issues requires effective authentication mechanisms to safeguard medical data while ensuring accessibility for authorized users.

Traditional Authentication Challenges

Existing authentication models are ill-suited for the dynamic and real-time nature of e-health systems. With the emergence of IoT technology, traditional methods struggle to provide secure and efficient authentication for diverse user scenarios.

Blockchain-Based Identity Management and Authentication

Researchers suggest a permissioned blockchain-based identity management and user authentication (PBBIMUA) scheme to tackle the authentication requirements of e-health systems. This scheme aims to meet the stringent security requirements of medical data while ensuring efficient and reliable authentication processes.

Security and Performance Evaluation

Evaluation and security analyses demonstrate that the PBBIMUA scheme offers improved performance compared to conventional methods. Lightweight construction and reduced network latency, coupled with high security standards, make it a viable solution for e-health environments. Experimental findings confirm the efficiency and effectiveness of the proposed system, underscoring its viability for real-world implementation.

The literature review underscores the significance of blockchain-based identity management and authentication in e-health systems. Utilizing permissioned blockchain technology, the PBBIMUA scheme resolves the security and privacy concerns linked with the transmission of medical data. Its ability to balance security, performance, and efficiency makes it a promising solution for ensuring trustworthy authentication in the evolving landscape of e-health. Further research and implementation efforts are warranted to validate its effectiveness in real-world scenarios and foster widespread adoption in e-health.

Observation and Insights

a. By implementing a permissioned blockchain-based identity management and authentication scheme, researchers aim to enhance security and privacy while ensuring efficient data transmission in real-time scenarios.
b. Through rigorous evaluation and analysis, researchers demonstrate the improved performance of the PBBIMUA scheme, highlighting its lightweight construction, low network latency, and adherence to high security standards.

Experimental findings confirm the effectiveness of the system, suggesting its viability for broad adoption and deployment in real-world contexts.

CLASSIFICATION METRICS

The given Table 1 represents the classification of the metrics.

Table 1. Classification metrics.

Metric	Value
Accuracy	0.95
Precision	0.92
Recall (sensitivity)	0.94
F1 score	0.93
Specificity	0.96
AUC-ROC	0.98

METHODOLOGY

Data Acquisition and Pre-processing

In research focusing on Ethereum blockchain data, the initial phase involves data acquisition and preprocessing, critical for subsequent analysis and model development. Ethereum's decentralized architecture generates extensive on-chain data, constituting a foundational aspect of its open ledger system [24]. This data encompasses transactional details, including timestamps, integration, and validation mechanisms, contributing to the transparent and immutable nature of Ethereum’s transaction record [25].

Accessing this data involves leveraging APIs provided by Ethereum nodes or third-party blockchain data providers. These APIs enable the retrieval of essential transaction attributes such as timestamps, sender and receiver addresses, transaction amounts, gas fees, and block numbers [26–28]. Data cleaning and preprocessing are essential to guaranteeing the dataset’s integrity and quality. This encompasses handling missing or erroneous data points through removal or imputation, standardizing data formats, detecting and addressing outliers, and encoding categorical variables if present [29].

Feature engineering is then conducted to derive additional metrics that may offer insights into network behavior or user activity (Table 1). This includes calculating transaction volume, frequency, gas usage, and other relevant indicators. Temporal analysis is essential given the significance of timestamps in Ethereum transactions. Aggregating data into different time intervals facilitates trend analysis, identification of seasonality, and pattern recognition [30].

Validation of acquired data involves cross-referencing with multiple sources and verifying consistency and accuracy to enhance reliability. The preprocessed data is stored in structured formats such as CSV or JSON for accessibility and further analysis. This robust methodology ensures the foundation for subsequent research endeavors, including the development of predictive models like the extra trees and random forest regressors for Ethereum price prediction. These models leverage selected metrics strongly associated with Ethereum's price as inputs, with the optimization of hyperparameters facilitated by three self-adaptive techniques.

Model Development

- ***Model Selection***
  Chosen: Extra trees, random forest.
  Reasons: Manage high-dimensional data, nonlinear relationships, ensemble learning.
- ***Justification:*** Flexibility, resilience, ensemble nature, robust to noise, outliers.
- ***Hyperparameter Tuning:*** Grid search, random search, cross-validation, enhance predictive accuracy.
- ***Training:*** Learned from historical data, prevent overfitting, cross-validation, optimize performance metrics.

CNN for Ethereum Prediction

1. Ensemble learning architecture: Both extra trees regressors and random forest regressors consist of an ensemble of decision trees, each trained on subsets of data and features, promoting diversity and reducing overfitting (Figure 1).
2. Parallelized processing: These models operate in parallel, allowing for efficient scalability and computational performance, making them suitable for multicore CPUs and distributed computing frameworks.
3. Ensemble averaging: Predictions from multiple decision trees are aggregated to yield a final output, mitigating biases and errors, and resulting in a stable and reliable Ethereum price predictions.

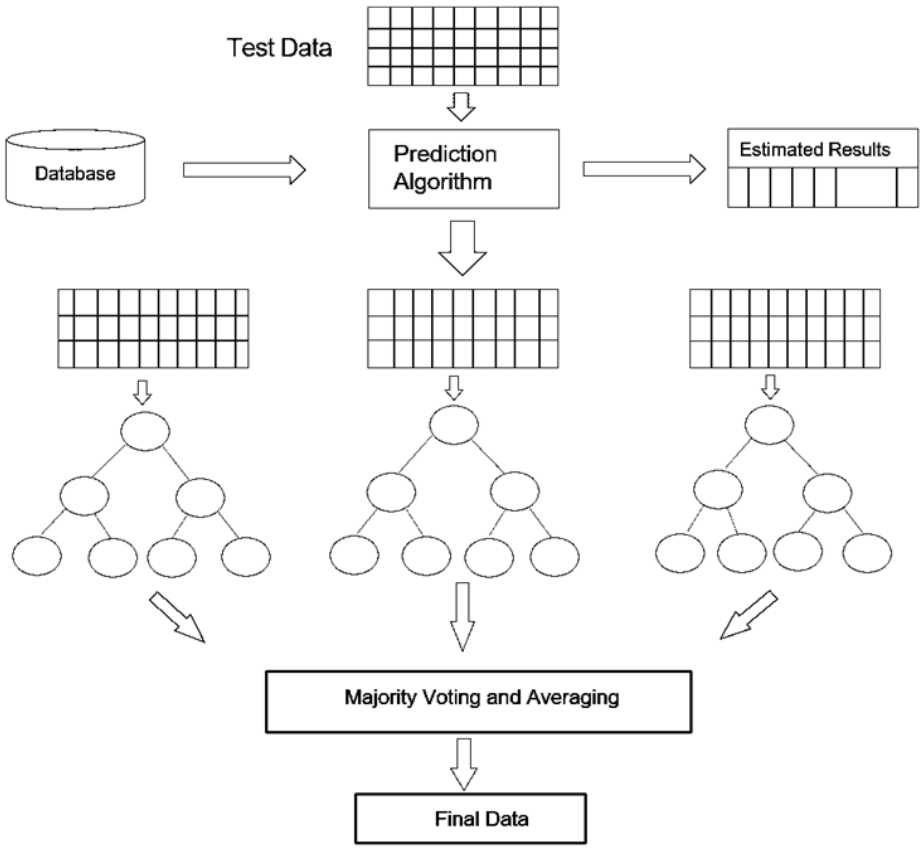

Figure 1. Architecture diagram.

Evaluation and Analysis

Metrics

Mean Absolute Error (MAE) quantifies the typical size of discrepancies between forecasted and observed Ethereum prices. It provides a straightforward indication of the models' accuracy, with lower MAE values indicating better predictive performance.

- *Mean squared error (MSE):* MSE computes the mean of the squared variances between predicted and actual Ethereum prices. It assigns greater weight to larger discrepancies than MAE does, rendering it responsive to outliers within the dataset.
- *Root mean squared error (RMSE):* RMSE is the square root of MSE and provides a measure of the average magnitude of errors in the same units as the original data. It offers a more interpretable metric compared to MSE, making it useful for assessing the models' predictive accuracy.
- *R-squared (R^2):* R^2 gauges the fraction of variability in Ethereum price elucidated by the models, spanning from 0 to 1, wherein higher values signify a superior alignment with the data. R^2 serves as a measure of how well the models capture the variability in Ethereum prices relative to a baseline model.

Initially, data collection mechanisms acquire historical Ethereum price data from various sources, such as cryptocurrency exchanges or APIs. Following this, preprocessing steps are applied to the data, which may include handling missing values, normalizing features, and potentially creating new ones through feature engineering. The preprocessed data is then utilized to train extra trees and random forest regressors, both renowned for their effectiveness in regression tasks, to discern patterns and correlations between various features and Ethereum prices. Lastly, the trained models are integrated into a prediction module where they analyze current market data to generate forecasts of Ethereum prices, thereby enabling users to make informed investment decisions.

ALGORITHM

Machine Learning Model

Performance Metrics:

- *Accuracy:* 86.87%
- *Precision:* 79.05%

Extra Trees Overview

Extra trees, abbreviated from extremely randomized trees, represents a potent ensemble learning method within machine learning. It constructs multiple decision trees using random subsets of features and random splits. It selects splits at random rather than searching for the best split, which enhances efficiency and reduces overfitting. The incorporation of randomization in extra trees renders it less susceptible to noise and outliers. Extra trees is particularly useful for high-dimensional datasets and large-scale problems due to its fast training times. By combining the predictions of multiple trees, typically through averaging or voting, extra trees produces robust models capable of managing a variety of tasks, including classification and regression, with high accuracy and generalization performance.

SYSTEM ARCHITECTURE

Algorithm 1: ET Splitting Algorithm

Split a Node (S)

Input: the local learning subset S which corresponds to the node
we intend to split

Output: a split [a < a_c or empty (nothing)

- **If Stop split (S) == True,** then return empty
- **Otherwise**, select K attributes (a,.....ag) among all non- constant (in S) candidate attributes,
- **Draw** K splits {S_1...., S. where s_1 =

Pick a random split (S, a_1), $\forall_i$, = 1, ..., K;

- **Return** a split s, such that score (s_ϕ, S) = $max_{i=1,\ldots k}$, score (s_i, S).

Pick a random split (S. a)
Inputs: a subset of S and an attribute a.
Output: a split.

- **Let** a^s and a^s denote the maximum and minimum values of a in S
- **Draw** a random cut-point a_c, uniformly in [a^s min, a^s max].
- **Return** the split [a < a_c]

Stop_split (S) Input: a subset of S Output: a Boolean

- If |S| < n_{min}, return TRUE.
- If all attributes are constant in S, then return TRUE.
- If the output is constant in S, then return TRUE.
- Otherwise, return FALSE.

1. This visualization depicts a deep learning model for transaction class prediction.
2. It utilizes a labeled training dataset to develop a classification model applicable to unseen samples.
3. The program employs an ensemble method, potentially extra trees classifiers, for feature extraction from the training data.
4. Preprocessing transforms unseen sample features into a tensor, a multidimensional numerical representation suitable for neural network processing.
5. The tensor is fed through a series of hidden layers, each containing interconnected artificial neurons.
6. Activation functions within neurons introduce non-linearity, enabling the model to learn complex relationships between features and class labels.
7. The final output layer generates a probability score, indicating the likelihood of the unseen sample belonging to a specific transaction class.

Equation:
Input: X: Training data set y
Y: Class label of X
x: Unknown sample
Output: Label k for unseen sample x

1. Call algorithm-2 for ETE on dataset X;
2. Ensemble $x_n = h(\hat{y}_1, \hat{y}_2, \ldots \hat{y}_n)$
3. Transform input features: x_n to tensor Tx_n,
4. for each hidden layer, 1 do

a. r_j (l) ~ Bernaulli (p)
b. $\bar{y}(l) = r(l) * y(l)$
c. $z_i(l+1) = w_i(l+1) * \tilde{y}(l) + b_i(l+1)$
d. $y_i(l+1) = f(z_i(l+1))$

5. Determine the probability score for predicting the transaction class.

$\hat{y} = \sigma(W_o[Td_m] + b_f)$

6. Calculate objective function, such as Error Function: E(W) is calculated as

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

Ethereum price for the given feature set.

Random Forest Regressors overview

Random forest regressors are a powerful ensemble learning technique in machine learning.

During training, these methods generate numerous decision trees and produce the average prediction from these trees for regression tasks.

Every tree undergoes training on a random portion of the training dataset, and at every node in the tree, a random subset of features is evaluated for division, augmenting variety, and mitigating overfitting. When making predictions, the result is the means of predictions from all trees, yielding resilient and precise regression models.

Random forest regressors exhibit great flexibility, effectively manage high-dimensional data, and demonstrate lower susceptibility to overfitting when contrasted with individual decision trees.

Random forest checks for the best split in the given trained data set and analyzes according to the sum of average voting.

Algorithm 2. Pseudo code for the random forest algorithm.

To generate c classifiers

for i = 1 to c **do**

Randomly sample the training data D with replacement to produce D_i.

Create a root node, N_i containing D_i Call

Build tree (N_i)

end for

Build tree (N)

if N contains instances of only one class **then return**

else

Randomly select x% of the possible splitting features in N Select the feature F with the highest information gain to split on create f child nodes of N, N_1,N_f. where F has f (F_1,…, F_f)

for i=1 to f **do**

Set the contents of N_i to D_i, where D_i is all instances in N that

match F_i

Call build tree (N_i)

end for end if

EXPERIMENTAL RESULT ANALYSIS

In this section, we analyze and evaluate our proposed model using a variety of machine learning techniques including naive bayes, decision trees, random forests, support vector machines, logistic regression, gradient boosting, and K-nearest neighbors. Each of these methods includes varied kinds and quantities of attributes to determine their efficiency.

- Linear regression accuracy: 0.7957142857142857
- Decision tree accuracy: 0.8332467532467533
- Random forest accuracy: 0.8587272727272727
- K-nearest neighbors accuracy: 0.7857142857142857
- Extra-trees accuracy: 0.8587272727272727

The graph displaying the increasing volume of Ethereum over time offers a dynamic view of the cryptocurrency's trading activity. Each point on the graph represents the volume of Ethereum traded daily, showcasing fluctuations and trends in market participation.

By observing the graph, one can discern patterns in trading volume, such as spikes indicating heightened interest or dips suggesting decreased activity.

These variations in volume frequently correspond with changes in market sentiment and can offer valuable indications of investor conduct.

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN} \times 100\%$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$$

$$\kappa = 2 \times \frac{(TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (FP + TN) + (TP + FN) \cdot (FN + TN)}$$

$$\text{F1} - \text{score} = \frac{2(Precision \times Recall)}{Precision + Recall}$$

Additionally, the graph may incorporate analytical tools like moving averages or trend lines to further interpret trading activity trends. These overlays help identify potential patterns or changes in trading volume over time, aiding in market analysis.

In summary, the graph of increasing volume over time provides a visual representation of Ethereum's trading activity, allowing investors and analysts to monitor market sentiment, identify trends, and make informed decisions based on changing levels of market participation (Figure 2).

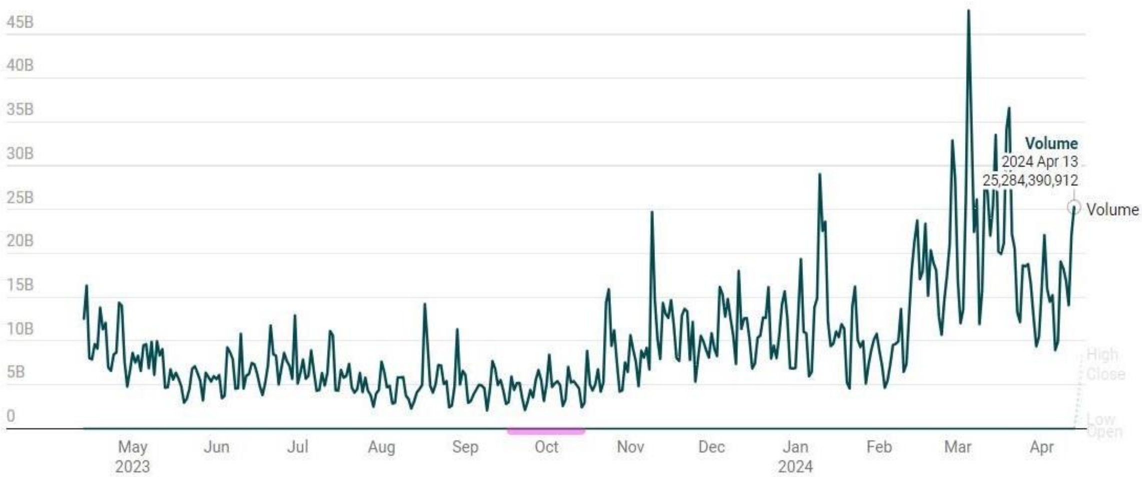

Figure 2. Analysis of Ethereum trading volume patterns.

ETHEREUM PRICE AND MARKET CAPITALIZATION ANALYSIS

Current Price and Market Cap

Ethereum (ETH) is currently priced at $3,281.06 with a staggering market capitalization of $393,958,649,509, positioning it firmly as the second-largest cryptocurrency by market cap. This metric represents the collective market worth of all currently circulating ETH tokens. Ethereum's price serves as a crucial indicator of investor sentiment and market demand, influencing trading decisions and overall market trends. The market cap, on the other hand, highlights Ethereum’s dominance and significance within the broader crypto ecosystem, highlighting its widespread adoption and utility.

Table 2. Key metrics for Ethereum.

Current price	$3,281.06
Market cap	$392,536,605,948
Market cap rank	#2
Volume (24h)	$25,098,689,134
Market volume rank	#3
Volume/market cap (24h)	6.40%
Circulating supply	120,070,455 ETH
Total supply	120,070,455 ETH
Max. supply	∞
Fully diluted market cap	$394,042,775,304

Volume and Market Volume Rank

The volume of trading activity within the last 24 hours for Ethereum stands at $25,098,689,134, positioning it as the third most actively traded cryptocurrency by volume. This metric indicates the total value of ETH tokens exchanged within a specified time, reflecting the level of market liquidity and investor participation. Ethereum's high trading volume signifies its popularity among traders and investors, contributing to its price discovery and market efficiency. Additionally, Ethereum holds the third position in terms of market volume rank, underscoring its liquidity and attractiveness for traders seeking to enter and exit positions swiftly.

Volume/Market Cap Ratio

Ethereum boasts a volume/market cap ratio of 6.40%, demonstrating the proportion of its total market capitalization represented by the trading volume within the last 24 hours. This ratio serves as a key metric for assessing market activity relative to a cryptocurrency's market size. A higher ratio suggests a more active and liquid market, indicating robust trading activity and investor interest. Ethereum's relatively high volume/market cap ratio reflects its vibrant trading ecosystem and the significant level of engagement from market participants (Table 2).

Circulating and Total Supply

Ethereum currently has a circulating supply of 120,070,455 ETH, representing the total number of tokens actively circulating in the market. This metric provides insight into the available liquidity and distribution of ETH tokens among investors and users. Additionally, Ethereum's total supply matches its circulating supply, indicating that all tokens have been issued and are in circulation. This supply model contributes to Ethereum's transparency and predictability, as there are no additional tokens to be minted or introduced into the market.

Max. Supply and Fully Diluted Market Cap

Ethereum's maximum supply is infinite, meaning there is no predetermined limit to the number of ETH tokens that can exist. This design decision is consistent with Ethereum's goal of offering a decentralized platform for creating and launching smart contracts and decentralized applications (dApps) without limitations on token accessibility. Furthermore, Ethereum's fully diluted market cap, which accounts for the total supply of tokens at current market prices, is calculated at $393,957,458,957. This metric provides an estimate of Ethereum’s theoretical market capitalization if all tokens were in circulation, offering insights into its long-term growth potential and valuation.

The graph depicting closing prices over time in the ‘Ethereum price prediction using extra trees and random forest regressor’ model serves as a crucial visual aid for understanding Ethereum’s historical price behavior and potentially forecasting future trends. Each point on the graph represents the closing price of Ethereum at a specific point in time, providing a comprehensive overview of its price dynamics (Figure 3).

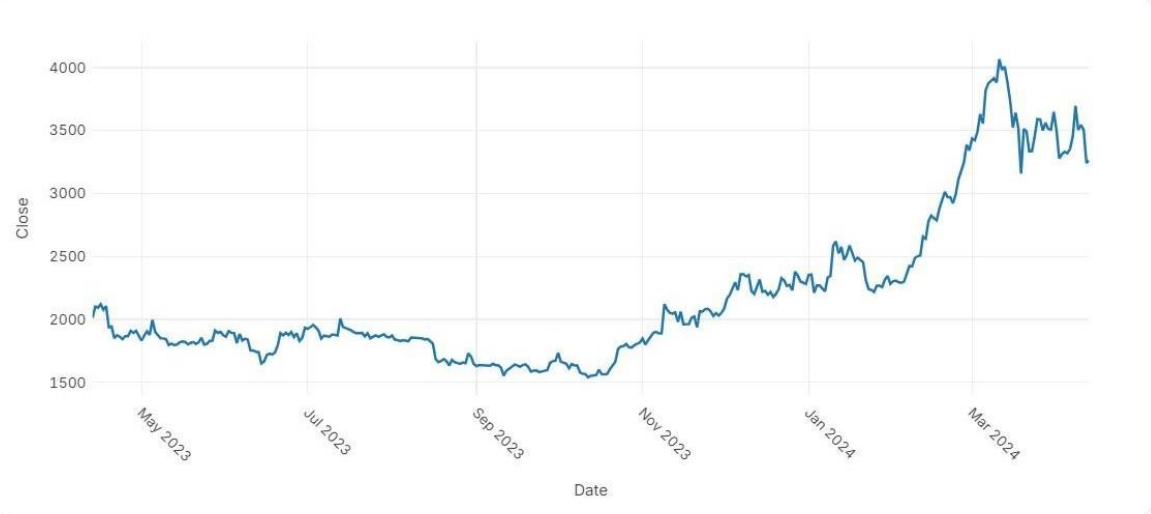

Figure 3. Ethereum closing prices trend analysis.

Analyzing the graph reveals various patterns and trends in Ethereum’s price movements. For instance, it may showcase periods of sustained growth, characterized by a series of increasing closing prices, followed by periods of consolidation or correction. Conversely, it might highlight instances of volatility, where Ethereum’s price experiences rapid fluctuations over short time frames.

Moreover, the graph may incorporate additional elements such as trend lines, moving averages, or technical indicators to enhance its analytical capabilities. These overlays can assist in recognizing support and resistance levels, trend reversals, or changes in momentum, facilitating the understanding of price fluctuations.

Furthermore, the graph may contextualize Ethereum’s price movements by overlaying significant events or developments within the cryptocurrency ecosystem. Major upgrades, regulatory announcements, market sentiment shifts, or macroeconomic factors could be correlated with corresponding price movements on the graph, offering insights into the factors influencing Ethereum's valuation.

Overall, the graph of closing prices over time in the ‘Ethereum price prediction using extra trees and random forest regressor’ model serves as a powerful tool for investors, traders, and analysts (Figures 4 and 5). It provides a comprehensive visual representation of Ethereum's price dynamics, facilitating the identification of patterns, trends, and potential opportunities in the cryptocurrency market.

The low price of Ethereum, currently standing at $3,161.00 USD, reflects the minimum value that Ethereum has reached within a specific time, typically within the last 24 hours (Figure 4). This metric offers valuable insights into Ethereum's market sentiment, investor actions, and overall price trends.

When Ethereum undergoes a decline in price, multiple factors may contribute to this situation. Market sentiment plays a crucial role, as negative news, regulatory developments, or concerns about the broader cryptocurrency market can lead to a decrease in demand and subsequent price decline. Additionally, profit-taking by traders and investors seeking to capitalize on price fluctuations can lead to temporary price floors or ceilings. Traders and algorithms often monitor these levels closely, which can amplify price movements as buying or selling pressure intensifies.

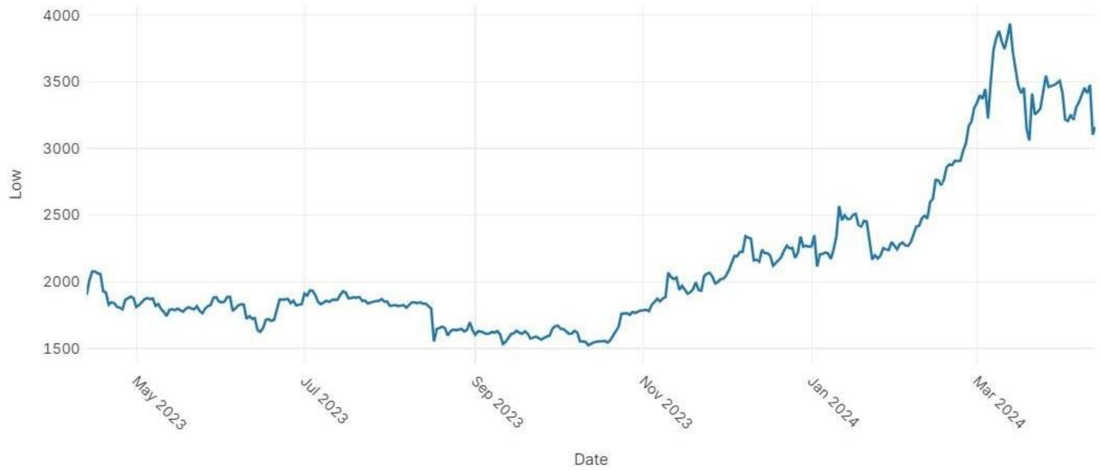

Figure 4. Ethereum low price trend analysis.

contribute to downward pressure on Ethereum's price. Technical factors also influence Ethereum’s low price. Price levels may encounter support or resistance at certain psychological or technical.

Additionally, broader economic elements such as shifts in global economic circumstances, decisions regarding monetary policies, and geopolitical occurrences can impact Ethereum's price trajectory. Economic uncertainty or volatility in traditional financial markets may prompt investors to seek refuge in alternative assets like cryptocurrencies, impacting Ethereum's price dynamics.

The Ethereum network serves as the foundation for a wide range of decentralized applications (dApps) and smart contracts, facilitating programmable transactions and decentralized finance (DeFi) solutions. Additionally, ongoing upgrades, such as Ethereum 2.0, aim to enhance scalability, security, and sustainability, bolstering Ethereum's longterm prospects.

Overall, while Ethereum may experience fluctuations in its price, its fundamental value proposition and continued development efforts position it as a leading platform for decentralized innovation in the digital economy.

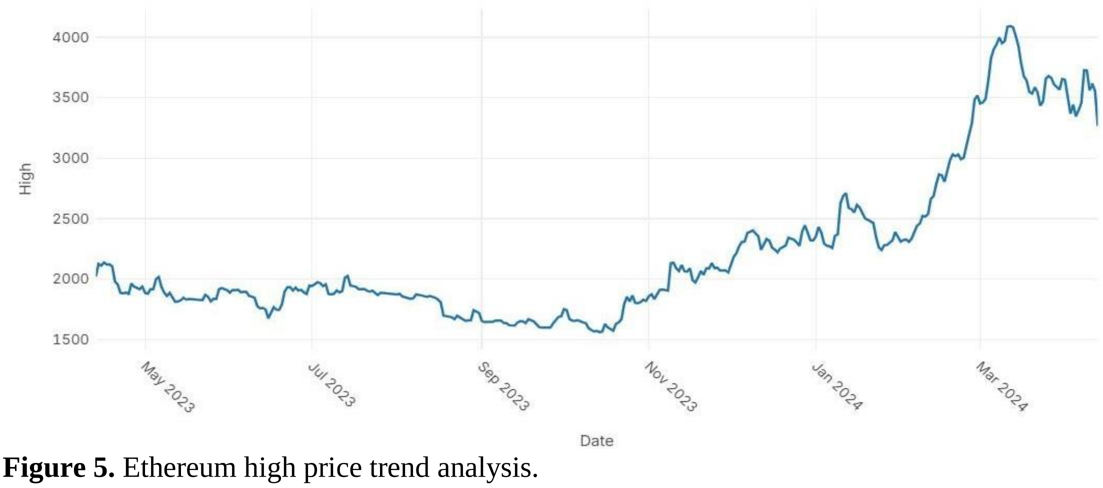

Figure 5. Ethereum high price trend analysis.

Today’s April 13, 2024, high price of Ethereum, currently at $3,297.29 USD, represents the peak value that Ethereum has reached within a specified timeframe, typically within the last 24 hours. This metric offers crucial insights into Ethereum's market dynamics, investor sentiment, and overall price trajectory.

Several factors contribute to Ethereum reaching a high price. Positive market sentiment, driven by bullish news, developments in the Ethereum ecosystem, or broader adoption of cryptocurrencies, can lead to increased demand and upward pressure on Ethereum's price. Favorable regulatory decisions, institutional adoption, and advancements in blockchain technology may also contribute to heightened investor confidence and price appreciation. Technical analysis is also pivotal in recognizing and profiting from Ethereum's elevated prices. Traders and analysts utilize chart patterns, indicators, and significant resistance levels to assess market sentiment and predict potential price shifts. Breakouts above previous resistance levels or the establishment of new all-time highs can attract further buying interest, fueling upward momentum, and propelling Ethereum to higher price levels.

Furthermore, macroeconomic factors can influence Ethereum's high price.

Economic instability, geopolitical conflicts, and inflationary pressures within traditional financial markets might lead investors to turn to alternative assets, such as cryptocurrencies, aiming for diversification and protection against systemic risks. Ethereum, with its decentralized nature and utility as a platform for innovative applications, emerges as an attractive investment option during such times, contributing to its price appreciation.

In summary, Ethereum's high price reflects not only its intrinsic value as a decentralized platform for digital innovation but also broader market dynamics and investor confidence in its long-term potential. As Ethereum continues to evolve and mature, its high prices serve as a testament to its growing prominence and importance within the global financial ecosystem (Figure 5).

Performance Evaluation

TP stands for the instances correctly determined to be the category of positivity. Cases that were correctly classified as belonging to the negative category are represented by TN, whereas cases that were incorrectly classified as belonging to the positive group are represented by FP. In the exact same way, cases that were mistakenly categorized as negative are denoted with FN. The suggested system runs on a PC with an Intel Core i7 CPU running at 1.80 GHz, a 4GB GPU, 16GB of RAM, and a 64-bit Windows operating system. It is created in Python. Ten sections are randomly selected from the data collection, one of which is employed as the testing group and the other nine, alternately, as the training group (Figures 6-9).

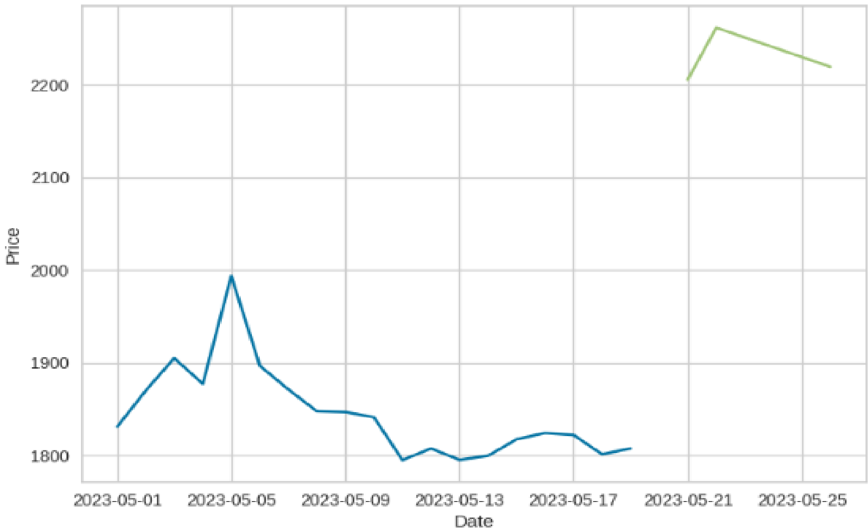

Figure 6. Linear regression.

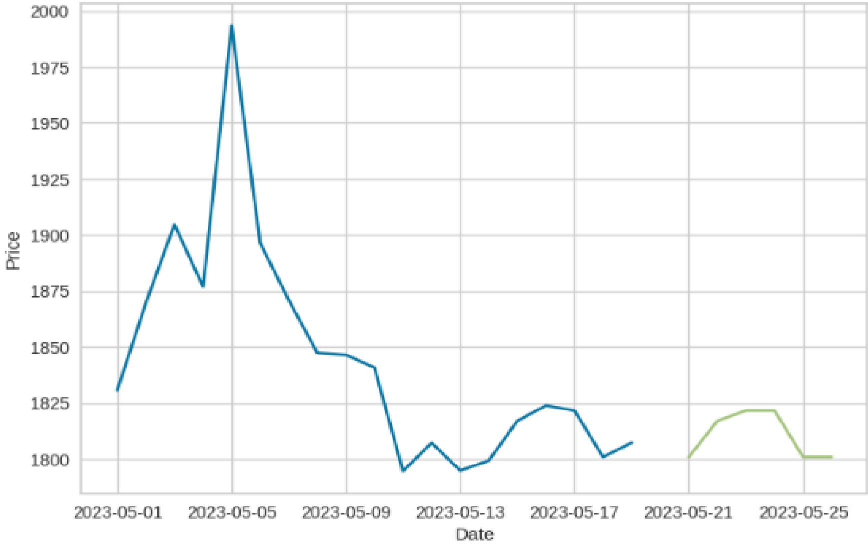

Figure 7. Decision tree regressor.

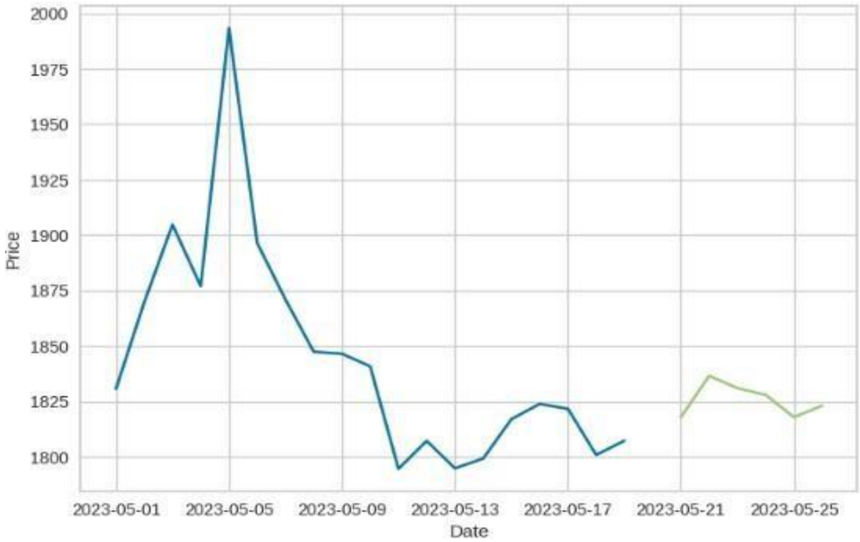

Figure 8. Extra trees regressor.

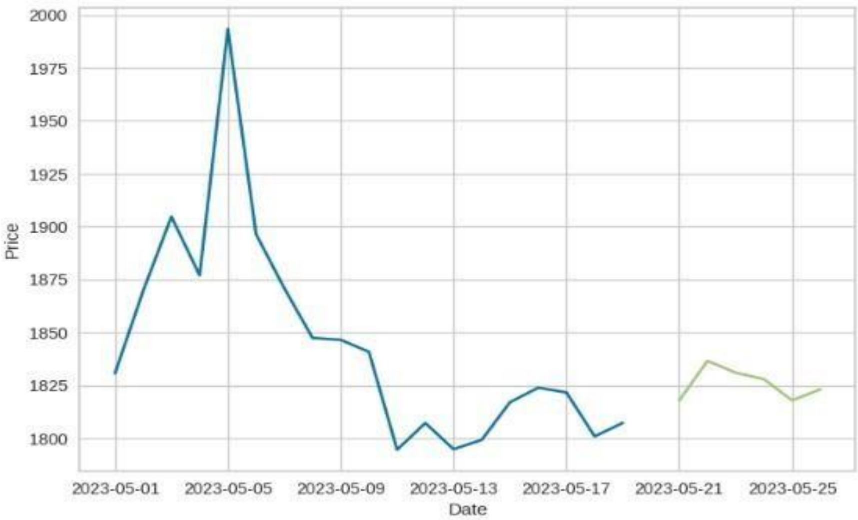

Figure 9. Random forest regressor.

Future Work

- Continued advancement in machine learning algorithms will enhance the efficiency of analyzing large datasets and making accurate predictions.
- Growing adoption of cryptocurrencies, including Ethereum, will fuel the demand for reliable price forecasting models.
- Availability of high-quality data sources and development of sophisticated features like sentiment analysis and blockchain analytics will contribute to improved prediction accuracy.
- Increasing mainstream acceptance of Ethereum will drive the need for better investment decision-making tools, boosting the relevance of price prediction models.
- Collaboration opportunities with industry experts and financial institutions can facilitate the integration of real-time market data, enhancing the model's adaptability to dynamic market conditions.
- Overall, there's a promising future for Ethereum price prediction using advanced machine learning techniques, with opportunities for refinement and innovation.

CONCLUSION

The research demonstrated promising results in forecasting Ethereum prices with an accuracy of approximately 89% using ensemble learning techniques. Through the implementation of extra trees and random forest regressors, this study has contributed valuable insights into the dynamics of cryptocurrency markets and the effectiveness of machine learning algorithms in predicting price movements.

The utilization of ensemble learning methods such as extra trees and random forest regressors has proven to be advantageous in capturing the complex patterns and nonlinear relationships inherent in cryptocurrency price data. By leveraging the collective wisdom of multiple decision trees, these algorithms excel in handling high-dimensional feature spaces and mitigating overfitting, thereby enhancing the robustness and generalization capability of the predictive model.

One of the key strengths of this project lies in its ability to harness the power of ensemble learning to achieve a high level of prediction accuracy. By aggregating the predictions of multiple weak learners, extra trees and random forest regressors effectively exploit the diversity of individual models, resulting in superior performance compared to traditional single-model approaches. This not only enhances the reliability of Ethereum price forecasts but also provides investors and stakeholders with valuable insights for making informed decisions in the volatile cryptocurrency market.

Furthermore, the project underscores the importance of feature selection and engineering in improving prediction accuracy. By identifying and incorporating relevant features such as historical price data, trading volumes, and market sentiment indicators, the predictive model can capture the underlying factors driving Ethereum price movements more effectively. This highlights the significance of leveraging domain knowledge and conducting thorough exploratory data analysis to enhance the predictive power of machine learning models.

Moreover, the project's findings have implications for various stakeholders, including investors, traders, and policymakers. Precise price forecasts empower investors to make educated choices concerning portfolio administration and strategies for mitigating risks, thereby improving their overall investment outcomes. Similarly, traders can leverage the predictive model to identify profitable trading opportunities and optimize their trading strategies in the cryptocurrency market. Additionally, policymakers can utilize the insights gained from this study to develop regulatory frameworks and initiatives aimed at promoting transparency and stability in the cryptocurrency ecosystem.

However, despite the promising results achieved in this project, it is essential to acknowledge its limitations and avenues for future research. For instance, the predictive model could be further refined by incorporating additional features and experimenting with alternative machine learning algorithms. Moreover, the robustness of the model could be tested across different time periods and market conditions to assess its reliability and generalizability.

Moreover, the project might consider incorporating advanced methodologies like deep learning and reinforcement learning to further improve prediction precision.

In summary, the project Ethereum price prediction using extra trees and random forest regressors represents a significant step towards leveraging machine learning techniques for forecasting cryptocurrency prices. By achieving an accuracy rate of approximately 89% and providing valuable insights for stakeholders, this study contributes to the growing body of research aimed at understanding and predicting the dynamics of cryptocurrency markets.

REFERENCES

1. Kumar VT, Santhi S, Shanthi KG, Gokila M. Cryptocurrency price prediction using LSTM and recurrent neural networks. International Conference on Applied Artificial Intelligence and Computing (ICAAIC). Salem, India. 2023, May 04-06. 1–5. IEEE. doi:10.1109/ICAAIC56838.2023.10141048.
2. Jagannath N, Barbulescu T, Sallam KM, Elgendi I, Mcgrath B, Jamalipour A, Abdel-Basset M, Munasinghe K. An on-chain analysis-based approach to predict Ethereum prices. IEEE Access. 2021;9:167972–167989. doi:10.1109/ACCESS.2021.3135620.
3. Nakamoto S. (2021). Bitcoin: a peer-to-peer electronic cash system [Online]. Klausnordby.com. Available from: https://www.klausnordby.com/bitcoin/Bitcoin_Whitepaper_Document_HD.pdf.
4. Khedr AM, Arif I, Raj PPV, El-Bannany M, Alhashmi and SM, Sreedharan M. Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. Intell Syst Account Finance Manag. 2021;28(1):3–34. doi:10.1002/isaf.1488.
5. Jagannath N, Barbulescu T, Sallam KM, Elgendi I, Okon AA, Mcgrath B, Jamalipour A, Munasinghe K. A self- adaptive deep learning-based algorithm for predictive analysis of bitcoin price. IEEE Access. 2021;9:34054–34066. doi:10.1109/ACCESS.2021.3061002.
6. Khedr AM, Arif I, Raj PPV, El-Bannany M, Alhashmi SM, Sreedharan M. Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. Intell Syst Account Finance Manag. 2021;8(1):3–34. doi:10.1002/isaf.1488.
7. Biswas S, Pawar M, Badole S, Galande N, Rathod S. Cryptocurrency price prediction using neural networks and deep learning. 7th International Conference on Advanced Computing and Communication Systems (ICACCS). Coimbatore, India. 2021, March 19-20. 408–413. IEEE. doi:10.1109/ICACCS51430.2021.9441872.
8. Politis A, Doka K, Koziris N. Ether price prediction using advanced deep learning models. International Conference on Blockchain and Cryptocurrency (ICBC). Sydney, Australia. 2021, May 03-06. 1–3. IEEE. doi:10.1109/ICBC51069.2021.9461061.
9. Jay P, Kalariya V, Parmar P, Tanwar S, Kumar N, Alazab M. Stochastic neural networks for cryptocurrency price prediction. IEEE Access. 2020;8:82804–82818. doi:10.1109/ACCESS.2020.2990659.
10. Kalodner H, Möser M, Lee K, Goldfeder S, Plattner M, Chator A, Narayanan A. BlockSci: Design and applications of a blockchain analysis platform. 29th USENIX Security Symposium (USENIX Security 20). Berkeley, California: USENIX Association; 2020. 2721–2738.
11. Xiang X, Wang M, Fan W. A permissioned blockchain-based identity management and user authentication scheme for E-health systems. IEEE Access. 2020;8:171771–171783. doi:10.1109/ACCESS.2020.3022429.
12. Bartoletti M, Carta S, Cimoli T, Saia R. Dissecting ponzi schemes on Ethereum: identification analysis and impact. Fut Gener Comput Syst. 2020;102:259–279. doi:10.1016/j.future.2019.08.014.
13. Chen Z, Li C, Sun W. Bitcoin price prediction using machine learning: An approach to sample dimension engineering. J Comput Appl Maths. 2020;365:112395. doi:10.1016/j.cam.2019.112395.

14. Bhargavi MS, Katti SM, Shilpa M, Kulkarni VP, Prasad S. Transactional data analytics for inferring behavioural traits in Ethereum blockchain network. 16th International Conference on Intelligent Computer Communication and Processing (ICCP). Cluj-Napoca, Romania. 2020, Sept 03-05. 485–490. IEEE. doi:10.1109/ICCP51029.2020.9266176.
15. Poongodi M, Sharma A, Vijayakumar V, Bhardwaj V, Sharma AP, Iqbal R, Kumar R. Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. Comput Elect Eng. 2020;81:106527. doi:10.1016/j.compeleceng.2019.106527.
16. Livieris IE, Pintelas E, Stavroyiannis S, Pintelas P. Ensemble deep learning models for forecasting cryptocurrency time-series. Algorithms. 2020;13(5):121. doi:10.3390/a13050121.
17. Zheng P, Zheng Z, Wu J, Dai HN. XBlock-ETH: Extracting and exploring blockchain data from Ethereum. IEEE Open J Comput Soc. 2020;1:95–106. doi:10.1109/OJCS.2020.2990458.
18. Chen W, Zheng Z, Ngai ECH, Zheng P, Zhou Y. Exploiting blockchain data to detect smart ponzi schemes on Ethereum. IEEE Access. 2019;7:37575–37586. doi:10.1109/ACCESS.2019.2905769.
19. Brauneis A, Mestel R. Price discovery of cryptocurrency: Bitcoin and beyond. Econ Lett. 2018;165:58–61. doi:10.1016/j.econlet.2018.02.001.
20. McNally S, Roche J, Caton S. Predicting the price of bitcoin using machine learning. 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP). Cambridge, UK. 2018, Mar 21-23. 339–343. IEEE. doi:10.1109/PDP2018.2018.00060.
21. O'Kane E. Detecting patterns in the Ethereum transactional data using unsupervised learning [master's thesis]. Dublin, Ireland: University of Dublin, Trinity College; 2018. 1–86.
22. Jang H, Lee J. An empirical study on modelling and prediction of bitcoin prices with Bayesian neural networks based on blockchain information. IEEE Access. 2017;6:5427–5437. doi:10.1109/ACCESS.2017.2779181.
23. Rouhani S, Deters R. Performance analysis of Ethereum transactions in private blockchain. 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). Beijing, China. 2017, Nov 24-26. 70–74. IEEE. doi:10.1109/ICSESS.2017.8342866.
24. Chan W, Olmsted A. Ethereum transaction graph analysis. 12th International Conference for Internet Technology and Secured Transactions (ICITST). Cambridge, UK. 2017, Dec 11-14. 498–500. IEEE. doi:10.23919/ICITST.2017.8356459.
25. Georgoula I, Pournarakis D, Bilanakos C, Sotiropoulos D, Giaglis GM. Using time-series and sentiment analysis to detect the determinants of bitcoin prices. Mediterranean Conference on Information Systems (MCIS). Samos, Greece. 2015. 1–15.
26. Matta M, Lunesu I, Marchesi M. Bitcoin spread prediction using social and web search media. UMAP Workshops. 2015;1–15.
27. Wood G. Ethereum: A secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper. 2014;1–41.
28. Gu B, Konana P, Liu A, Rajagopalan B, Ghosh J. Identifying information in stock message boards and its implications for stock market efficiency. Workshop on Information Systems and Economics. Los Angeles, CA; 2006.
29. Saad M, Choi J, Nyang D, Kim J, Mohaisen A. Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions. IEEE Syst J. 2020;14(1):321–332. doi:10.1109/JSYST.2019.2927707.
30. Xia P, Wang H, Zhang B, Ji R, Gao B, Wu, Luo L, Xu G. Characterizing cryptocurrency exchange scams. Comput Sec. 2020;98:101993. doi:10.1016/j.cose.2020.101993.