Research Article

Predicting Bitcoin's Price Evolution: A Comparative Analysis of Meta - Learning Algorithms

Adeev Mardia

KC High International School, Nungambakkam, Chennai - 600034, Tamil Nadu, India

ABSTRACT

This manuscript presents a comparative study of three machine learning models—Long Short-Term Memory (LSTM), Random Forest Regressor (RFR), and Support Vector Machine (SVM)—for predicting Bitcoin closing prices over the period from 2015 to 2023. The study focuses on evaluating the predictive performance of these models in terms of key metrics such as Mean Squared Error (MSE) and R-squared values, which measure their ability to forecast Bitcoin's price trends. The findings reveal that the Random Forest Regressor (RFR) model outperforms both LSTM and SVM in terms of accuracy and robustness. RFR demonstrated the lowest MSE and the highest R-squared value, effectively capturing both short-term fluctuations and long-term trends in Bitcoin prices. LSTM exhibited moderate performance, struggling to capture extreme volatility, while SVM showed the poorest results, with the highest MSE and lowest R-squared value. In addition, this study explores the relationship between Bitcoin's closing prices and trading volume, identifying a significant correlation that provides insights into market sentiment and price volatility. Challenges such as data preprocessing for SVM and hyperparameter tuning for LSTM are also discussed. The results underscore the potential of machine learning, particularly Random Forest, in enhancing cryptocurrency trading strategies and risk management. Future research will focus on integrating additional features and optimising models for real-time applications.

Keywords: Bitcoin; Machine Learning; Random Forest; Crypto Trading; Price Prediction; LSTM; Risk Management

INTRODUCTION

Bitcoin, the first decentralised (system that operates without a central authority or governing body) cryptocurrency, has grown from a niche digital asset to a globally recognized financial phenomenon since its inception in 2009. Unlike traditional currencies, Bitcoin operates without a central authority, relying on a peer-to-peer network and blockchain technology to ensure transparency and security in transactions. This innovative structure has disrupted traditional financial systems, introducing a digital alternative to government-backed currencies. However, Bitcoin's most striking feature is its extreme price volatility, influenced by a complex interplay of market sentiment, regulatory changes, macroeconomic factors, and technological advancements. This volatility

Corresponding author: Adeev Mardia, E-mail: adeevmardia@gmail.com.

Received October 18, 2024**; Accepted** November 18, 2024
https://doi.org/10.70251/HYJR2348.248296

presents both opportunities and risks for investors, making price prediction an essential yet challenging endeavour.

The cryptocurrency market operates 24/7, in contrast to traditional financial markets, which are characterised by regulated exchanges and limited trading hours. Unlike stocks and bonds, Bitcoin lacks intrinsic value, dividends, or interest payments, further complicating its valuation. Instead, its price dynamics are heavily influenced by speculative behaviour, news cycles, and social media trends. These unique characteristics of cryptocurrencies render traditional financial models inadequate for predicting Bitcoin prices, paving the way for machine learning techniques that can analyse vast datasets and identify complex patterns beyond the reach of conventional approaches.

Machine learning (ML), a subset of artificial intelligence (AI), has emerged as a powerful tool for financial forecasting due to its ability to process large datasets and capture non-linear relationships among variables. For Bitcoin price prediction, ML models like Random Forest Regressor (RFR), Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks are particularly promising. Each of these models brings unique strengths to the table. RFR, an ensemble learning method (machine learning technique where multiple models are combined to improve predictive accuracy), is adept at handling noisy, high-dimensional financial datasets and provides robust predictions. LSTM, a specialised recurrent neural network, excels at capturing temporal dependencies in sequential data, making it well-suited for analysing time series like Bitcoin prices. Meanwhile, SVM is known for its ability to handle non-linear relationships, especially when paired with the kernel trick, though its performance depends significantly on data preprocessing and scaling.

Previous research has demonstrated the potential of these ML models in cryptocurrency trading, but a comprehensive comparison of their performance remains limited. Studies have highlighted the effectiveness of ensemble methods like Random Forest in mitigating overfitting and generalising well across diverse datasets. Similarly, LSTM networks have been praised for their ability to model long-term dependencies, capturing trends and shifts in market dynamics. SVMs, while less commonly applied to financial time series, have shown promise in certain scenarios involving linear and non-linear classification. Despite these advances, significant challenges persist, including the dynamic nature of cryptocurrency markets, the risk of overfitting, and the need for continuous model updates to incorporate new data.

The relationship between Bitcoin's price and trading volume has also attracted considerable attention in the literature, with studies suggesting that trading volume can serve as a proxy for market sentiment and activity levels. A high trading volume often correlates with significant price movements, offering traders valuable insights into potential trends and reversals. By incorporating trading volume and other macroeconomic indicators, ML models can enhance their predictive power and provide a deeper understanding of market behaviour.

Despite its promise, the application of ML to cryptocurrency trading is not without limitations. The quality and availability of historical data can significantly impact model performance. Additionally, the interpretability of complex models like LSTM remains a concern, as traders and investors often prefer transparent decision-making tools. Overfitting (a situation where a machine learning model is too closely fitted to the training data, leading to poor performance on new, unseen data) is another critical issue, particularly when models attempt to capture noise rather than meaningful patterns in the data. Addressing these challenges requires careful feature engineering (process of selecting, modifying, or creating new features from raw data to improve the performance of a machine learning model), hyperparameter tuning (process of adjusting the pre-set parameters of a machine learning model to optimise its performance), and a thorough understanding of the limitations and assumptions underlying each model.

This study seeks to address these gaps by providing a comprehensive evaluation of three machine learning models—Random Forest Regressor, Long Short-Term Memory networks, and Support Vector Machines—for predicting Bitcoin's closing prices from 2015 to 2023. The research aims to answer the following questions:

- Which ML model demonstrates the highest predictive accuracy and robustness in forecasting Bitcoin prices?
- How do these models handle the unique challenges of cryptocurrency markets, such as high volatility and the absence of intrinsic value?
- What insights can be derived from the relationship between Bitcoin's price and trading volume?

To achieve these objectives, the study analyses historical Bitcoin data, including features like opening price, high price, low price, closing price, adjusted closing price, and trading volume. These features are pre-processed to enhance their suitability for time series forecasting, including the calculation of daily returns to capture price fluctuations. Each model undergoes rigorous hyperparameter tuning to optimise its performance, and the results are evaluated using metrics such as Mean Squared Error (MSE) and R-squared values.

The findings reveal that Random Forest outperforms both LSTM and SVM in terms of predictive accuracy and robustness, achieving the lowest MSE and highest R-squared value. LSTM provides moderate accuracy, capturing broader trends but struggling with extreme volatility. SVM, despite its theoretical strengths, underperforms significantly in this context, highlighting the importance of model selection based on the specific characteristics of the dataset and task.

This study contributes to the growing body of research on the application of AI in financial markets, offering valuable insights for traders and investors navigating the volatile world of cryptocurrencies. By comparing the strengths and limitations of different ML models, the research underscores the potential of ensemble methods like Random Forest in capturing the dynamic and complex nature of Bitcoin prices. The findings also highlight the need for continuous model refinement and the integration of additional features, such as macroeconomic indicators and sentiment analysis (the process of using natural language processing techniques to determine and extract the emotional tone or sentiment from a piece of text), to enhance predictive accuracy further.

In summary, this research demonstrates the transformative potential of AI and machine learning in cryptocurrency trading, offering tools and strategies that can help market participants make more informed decisions. By bridging the gap between theoretical advancements and practical applications, the study lays the groundwork for future research aimed at developing more sophisticated and reliable predictive models for digital assets. The hypothesis of this research is that advanced machine learning models, particularly Random Forest and LSTM, can more accurately predict Bitcoin price movements compared to traditional methods like SVM, by effectively capturing the complex relationships between historical price data and market dynamics.

LITERATURE REVIEW

The application of AI and machine learning in financial markets has grown significantly over the past decade. Researchers and practitioners have focused on developing models to predict stock prices, cryptocurrency values, and market trends using various algorithms, including neural networks, support vector machines (SVMs), and ensemble methods like random forests. These models aim to identify patterns and forecast price movements, thereby assisting investors in making informed decisions.

Time Series Analysis in Financial Markets

Time series analysis has been a fundamental approach for modelling and predicting market behaviours. Techniques such as ARIMA (AutoRegressive Integrated Moving Average, a time series forecasting model that combines autoregression, differencing, and moving averages to predict future values) have been traditionally used for forecasting in stock markets, leveraging past data to predict future price movements (4). However, with the advent of more sophisticated AI techniques, these traditional models have been complemented or replaced by machine learning models that can capture more complex patterns and dependencies in the data. Research has shown that models like LSTM (Long Short-Term Memory) networks are particularly effective in capturing temporal dependencies due to their ability to maintain long-term memory, making them suitable for financial time series data (5).

Machine Learning in Cryptocurrency Trading

Cryptocurrencies like Bitcoin exhibit higher volatility than traditional financial assets, posing unique challenges and opportunities for predictive modelling. Studies such as those by Mallqui and Fernandes (2019) have demonstrated the effectiveness of machine learning models in predicting cryptocurrency prices. Their research employed various models, including decision trees, random forests, and SVMs, highlighting that ensemble models like random forests often outperform single models in terms of accuracy due to their ability to reduce overfitting and improve generalisation (6).

Support Vector Machines and Predictive Modeling

Support Vector Machines (SVMs) have been a popular choice for financial forecasting due to their robustness in handling high-dimensional data and their ability to model non-linear relationships (7). Studies by Tripathy N and others have applied SVMs to stock price prediction, showing promising results in capturing the complex patterns of financial time series (8). However, the limitations of SVMs in terms of scalability and their sensitivity to the choice of kernel parameters have led to a growing interest in alternative methods like deep learning, which offer greater flexibility in modelling complex data structures (9).

Neural Networks and Deep Learning

Neural networks, particularly deep learning models such as LSTMs, have gained traction in financial forecasting due to their ability to model complex, non-linear relationships (6). A research by Fischer and Krauss demonstrated the superiority of LSTM models over traditional methods in predicting stock prices by effectively capturing temporal dependencies and non-linear patterns in the data (10). Recent studies have extended these findings to cryptocurrency markets, where LSTM models have shown promise in predicting price trends despite the high volatility and noise characteristic of these markets (11).

Ensemble Methods: Random Forests

Random Forests, an ensemble learning method, have been widely applied in various domains for classification and regression tasks due to their ability to handle large datasets with higher dimensionality. In financial forecasting, they have been used to model the non-linear relationships between market indicators and asset prices (12). Research by Patel et al. showed that random forests could effectively predict stock prices by leveraging a combination of multiple decision trees to improve predictive accuracy and robustness against overfitting (13).

Correlation Analysis in Financial Markets

Correlation analysis plays a critical role in understanding the relationships between different financial indicators. In the context of cryptocurrency trading, analysing the correlation between variables like volume and closing price can provide insights into market behaviour and trader sentiment (14). Studies have shown mixed results regarding the strength and direction of these correlations, suggesting that while some relationships may appear weak on a broader scale, they could be significant in specific market conditions or timeframes (15).

Data Description

The dataset used in this study consists of historical data for Bitcoin trading. It captures various financial indicators such as the opening price, high price, low price, closing price, adjusted closing price and trading volume. This dataset spans from 2015-2023 and provides a comprehensive view of Bitcoin's price movements and market behaviour over an extended period. The dataset's chronological structure is essential for conducting time series analysis which is critical in understanding the trends and patterns within the cryptocurrency market.

The Date column serves as the index for the dataset, providing a sequential timeline that allows for the analysis of Bitcoin's performance over time. Each entry corresponds to a specific day, making it possible to observe daily price movements and volume changes. This temporal organisation is vital for time series forecasting models, such as the LSTM and Random Forest Regressor models used in this study, which rely on past data to predict future trends.

The Open price represents the price at which Bitcoin was first traded when the market opened on a given day (Table 1). This metric is crucial for understanding the market's starting point and how sentiment might evolve throughout the day. In contrast, the High and Low prices provide insights into the volatility of Bitcoin on a daily basis. The high price reflects the maximum value Bitcoin reached during the day, while the low price indicates the minimum value. These metrics are essential for analysing the price range and volatility, which are significant factors in trading strategies and risk management.

The Close price is one of the most critical indicators in the dataset, representing the final price at which Bitcoin traded at the end of the day. This value is widely used by traders and analysts to gauge the overall performance of Bitcoin on a given day and is often compared with the opening price to determine the market's daily direction. Additionally, the Adjusted Close price accounts for any

Table 1. Screenshot of a section of the dataset

Date	Open	High	Low	Close	AdjClose	Volume
03/01/15	314.846008	315.149994	281.082001	281.082001	281.082001	33054400
04/01/15	281.145996	287.230011	257.612	264.195007	264.195007	55629100
05/01/15	265.084015	278.341003	265.084015	274.473999	274.473999	43962800
06/01/15	274.610992	287.553009	272.696014	286.188995	286.188995	23245700
07/01/15	286.076996	298.753998	283.07901	294.337006	294.337006	24866800

dividends, stock splits, or other adjustments, providing a more accurate representation of Bitcoin's true market value over time.

The Volume metric indicates the total number of Bitcoin units traded during a specific day. This data point is crucial for understanding the market's liquidity and the level of activity among traders. High trading volumes often correspond to increased market interest and can signal potential price movements. By analysing the relationship between volume and price, traders can gain insights into market sentiment and potential future trends.

Overall, this dataset offers a rich source of information for analysing Bitcoin's market behaviour over the years. By leveraging these features, this study aims to build robust predictive models that can forecast future price movements, providing valuable insights for traders and investors in the cryptocurrency market.

METHODS AND MATERIALS

The methodology section of this study outlines the comprehensive approach taken to analyse and predict Bitcoin price movements using various machine learning models. The primary focus of this study is on three models: Random Forest, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) neural networks. Each model brings a unique set of strengths and methodologies to the table, providing a diverse range of analytical perspectives on the time series data of Bitcoin prices.

The initial step in the methodology involves preparing the data for analysis. The dataset, spanning from 2015 to 2023, contains daily Bitcoin prices and trading volumes. It was first organised chronologically and cleaned to remove any missing or anomalous values that could skew the results. Feature engineering was then performed to create new variables, such as daily returns, which capture the day-to-day percentage change in Bitcoin's closing price. These features are essential for training the models, as they provide a comprehensive view of the factors that might influence price movements.

The choice of these three models — Random Forest, SVM, and LSTM — was guided by their complementary strengths. Random Forest offers a robust baseline with good generalisation capabilities; SVM provides flexibility in handling non-linear relationships, and LSTM leverages the sequential nature of time series data to capture long-term dependencies. Together, these models provide a comprehensive toolkit for predicting Bitcoin price movements, allowing for a nuanced analysis that accounts for various aspects of the data.

Random Forest Regression

Random Forest is an ensemble learning method primarily used for classification (a machine learning task where the goal is to assign data points to predefined categories or classes based on their features) and regression tasks. It builds multiple decision trees during training and outputs the mean prediction of the individual trees for regression or the majority vote for classification. The main advantage of Random Forest lies in its ability to handle large datasets with higher dimensionality and its robustness to overfitting, especially compared to a single decision tree (12).

Decision Trees

The core component of Random Forest is the decision tree. A decision tree is a flowchart-like structure where an internal node represents a feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome.

Mathematical Representation of Decision Tree: At each node, a split is made based on a certain feature that minimises the impurity (e.g., variance in regression, Gini impurity or entropy in classification).

$$\text{Split on feature } f^* = \arg\min_f \sum_{j=1}^{m} (y_j - \bar{y})^2$$

(Equation 1: Impurity in Decision Trees)

where *f* is a feature, y_j are the target values, and *y bar* is the mean target value after the split.

The decision tree tries to find the best feature and the best threshold value that minimises the prediction error.

Random Feature Selection. One of the unique aspects of Random Forest compared to a single decision tree is the random selection of features at each split (Figure 1). This randomness increases the diversity among the trees and helps in reducing correlation among them, thus improving the overall model performance.

Mathematical Representation of Random Feature Selection: If there are F total features, a subset f (where f<F) is randomly selected at each node. The split is made based on this random subset:

$$\text{Best split } = \arg\min_{f'} \sum_{j=1}^{m} (y_j - \bar{y}_{f'})^2$$

(Equation 2: Random Feature Selection)

Finally, the average of all decision trees is calculated. The final prediction in Random Forest is made by aggregating the predictions from all individual trees.

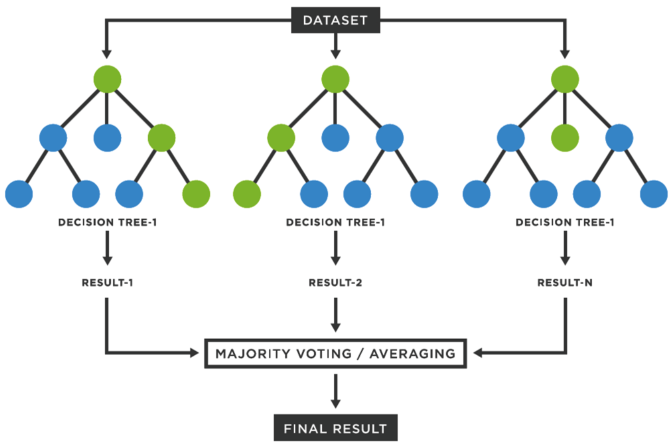

Figure 1. Decision Trees Visualisation (21).

Why was Random Forest used in this Research? Random Forest was chosen for this research because of its ability to handle large datasets and its robustness against overfitting. In predicting Bitcoin closing prices, the Random Forest model can capture the complex interactions between features like open, high, low, adjusted close prices, volume, and daily returns. The model's ensemble approach, where multiple decision trees vote on the final prediction, allows it to generalise well to new data, making it particularly suitable for financial time series forecasting. Moreover, Random Forest provides feature importance metrics, enabling a deeper understanding of which factors most influence Bitcoin prices.

The mathematical foundations discussed here highlight the rigorous approach Random Forest uses to ensure accurate, reliable, and interpretable predictions, making it a powerful tool in financial data analysis.

Long - Short Term Memory

Description of LSTM. LSTM is a type of recurrent neural network (RNN) that is particularly well-suited for sequence prediction problems, such as time series forecasting. LSTMs are designed to handle long-term dependencies in data, overcoming the vanishing gradient problem commonly encountered in traditional RNNs (a type of neural network designed for processing sequential data, where outputs from previous steps are used as inputs for the current step).

Why did I use LSTM? LSTM is ideal for this project because it can capture temporal dependencies in the Bitcoin closing prices over time. The model is robust against long-term trends and can maintain information over several time steps, which is crucial in financial time series data where past trends significantly impact future prices. While LSTM models can be computationally intensive, their ability to accurately model sequences makes them invaluable for predicting financial data like cryptocurrency prices.

Support Vector Machines

Description of SVMs. Support Vector Machines (SVM) are a set of supervised learning methods used for classification, regression, and outlier detection. In your research, SVM is used for regression tasks, known as Support Vector Regression (SVR).

The goal of SVM is to find the optimal hyperplane that best separates the data points in a high-dimensional space. For regression, this hyperplane is positioned to minimise the error within a margin, which is controlled by a parameter C.

Why did I use SVM in my research? In your project, SVM was chosen for its ability to handle high-dimensional data and its robustness in avoiding overfitting. Despite its computational complexity, SVM is particularly effective in scenarios where there is a clear margin of separation. For time series data, while SVM can be powerful, it often struggles with capturing long-term dependencies and temporal patterns, which might explain its lower performance compared to models like LSTM. However, the kernel trick allows SVM to perform non-linear regression, which can still be valuable for capturing some patterns in the data

Mathematics Involved

Mathematics involved in RFR. Bootstrap Aggregation (or bagging) is a technique to reduce variance in machine learning models, particularly decision trees. The idea is to generate multiple datasets from the original dataset using bootstrapping and then train a model on each dataset.

Bootstrapping: Suppose we have a dataset. $D = \{(x_1, y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$ with n samples. A bootstrap sample D_i is generated by randomly selecting n samples from D with replacement. This means some samples may be repeated in D_i, while others may not appear at all (Figure 2).

Bagging involves training multiple models (e.g., decision trees) on different bootstrap samples and then aggregating their predictions. For regression, the aggregation is typically the average of the predictions.

$$\hat{y} = \frac{1}{T} \Sigma_{i=1}^{T} \hat{y}_i$$

(Equation 3: Bagging)

Where $\hat{y}_i$ is the prediction from the i-th model, and T is

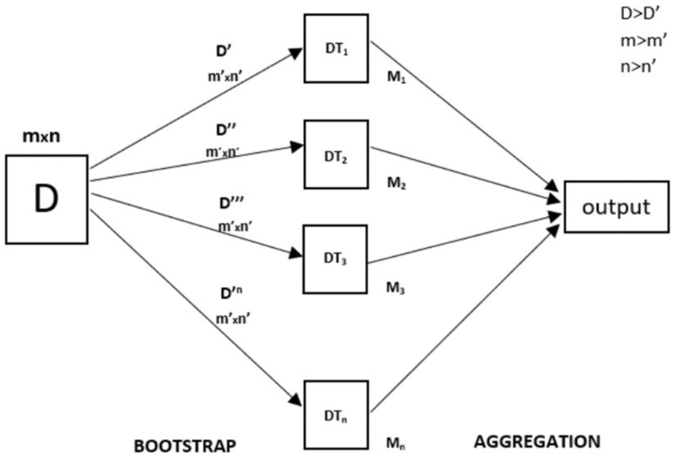

Figure 2. Bootstrap Aggregation (20).

the total number of models. Bagging helps in reducing the variance of the model by averaging out the errors from each individual model.

Mathematics involved in LSTM.

Forget Gate: This gate decides what information from the previous cell state should be discarded.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

(Equation 4: Forget Gate - LSTM)

Input Gate: This gate determines what new information should be added to the cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

(Equation 5: Input Gate - LSTM)

Cell State Update: This represents the updated cell state after taking into account the forget and input gates.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

(Equation 6: Cell State Update - LSTM)

Output Gate: This gate controls what part of the cell state should be output as the hidden state.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

(Equation 7: Output Gate LSTM)

Hidden State: This represents the memory of the network at a given time step and is passed to the next time step.

$$h_t = o_t * \tanh(C_t)$$

(Equation 8: Hidden State - LSTM)

The main mathematical operations involved in an LSTM cell include matrix multiplications, additions, and element-wise operations like *sigmoid* (σ) and *hyperbolic tangent (tanh)*. These operations ensure that the network can remember information over long sequences, making LSTM a powerful tool for time series prediction (23) (Figure 3).

Mathematics involved in SVMs.

Objective Function: SVM minimises a regularised loss function that balances the trade-off between the model complexity (represented by the norm of the weight vector ||w||) and the error margin. The parameter C controls this trade-off. A higher C value allows the model to penalise errors more, potentially leading to overfitting, while a lower C value might underfit the data.

Kernel Trick: In cases where the data is not linearly separable, SVM employs the kernel trick to map the data into a higher-dimensional space where a linear separation is possible. Common kernels include the Radial Basis Function (RBF), polynomial, and sigmoid kernels.

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right)$$

(Equation 9: Kernel Trick - SVM)

Support Vectors: The model's predictions are based on a subset of the training data known as support vectors. These are the critical data points that lie on the margin and directly influence the position of the decision boundary.

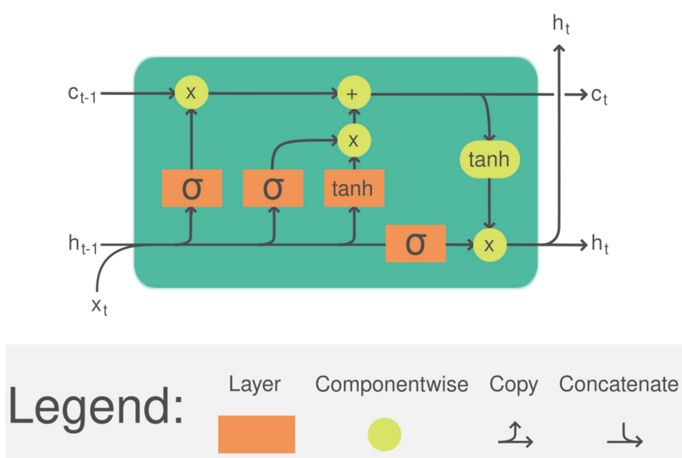

Figure 3. LSTM Architecture.

$$y_i(w \cdot x_i + b) = 1$$

(Equation 10: Support Vectors - SVM)

Decision Function: SVM aims to find the hyperplane that best separates the data into different classes or, in the case of regression, fits the data points with the maximum margin. The decision function for SVM can be written as:

$$f(x) = \text{sign}(w \cdot x + b$$

(Equation 11: Decision Function - SVM)

Objective Function: The SVM model minimises the following objective function:

$$\min_{w,b} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \max(0, 1 - y_i(w \cdot x_i + b))$$

(Equation 12: Decision Function - SVM)

RESULTS

General Results - Representation & Interpretation

Figure 4 presents a time series plot of Bitcoin's closing prices from 2015 to 2023. The plot vividly captures the inherent volatility of the cryptocurrency market, with prices exhibiting significant fluctuations over this period. The graph reveals several distinct phases, including sharp upward trends, steep declines, and periods of relative stability. These fluctuations are indicative of the speculative nature of Bitcoin, reflecting both market sentiment and external factors that influence its value. The overall trend shows a dramatic increase in price from 2015 through 2021, followed by more pronounced variability in subsequent years. This visual representation underscores the challenges and opportunities in predicting Bitcoin prices over time.

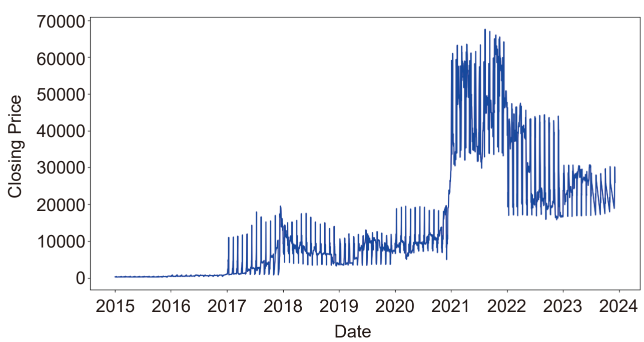

Figure 4. Bitcoin Closing Prices (USD) from 2015 to 2023.

Figure 5 depicts a smoothed version of the Bitcoin closing prices from 2015 to 2023, providing a clearer view of the underlying trends by reducing the noise from daily fluctuations. This smoothing process highlights the broader movements in the market, making it easier to identify long-term trends and patterns. The graph reveals the general trajectory of Bitcoin's price over the years, with prominent peaks and troughs more readily apparent. By filtering out short-term volatility, this smoothed representation facilitates more accurate analysis and interpretation, offering insights into the overall direction of the market and aiding in better forecasting and strategic decision-making.

Long - Short Term Memory Results

Figure 6 illustrates the Actual vs. Predicted Closing Prices for Bitcoin using the Long Short-Term Memory (LSTM) model from 2022 to 2023, with the blue line representing actual prices and the red line representing predicted prices. The LSTM model captures general trends and broader movements, but struggles with predicting

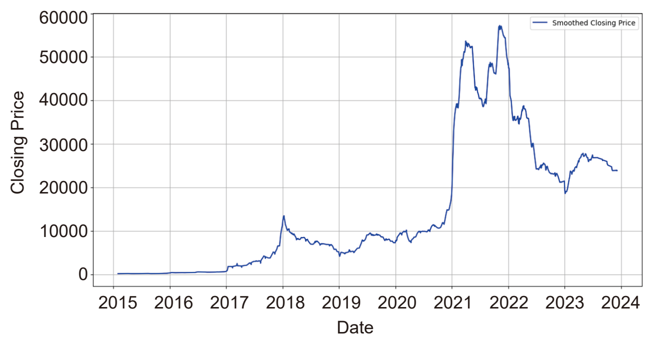

Figure 5. Smoothed Graph of Bitcoin Closing Prices from 2015-2023.

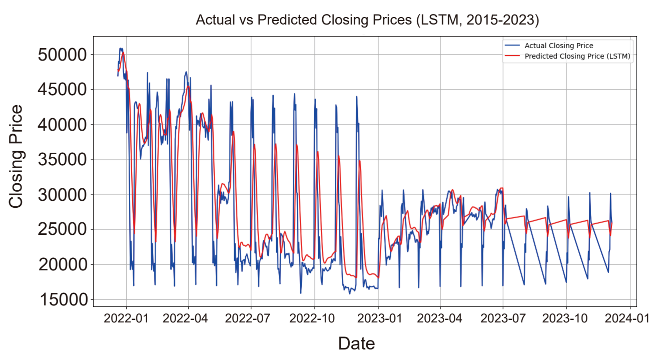

Figure 6. Actual vs Predicted Closing Prices for Bitcoin using LSTM.

sharp price fluctuations, especially during periods of high volatility, as seen by the smoother red line compared to the sharp drops in the blue line. The model's MSE of 84,173,172.78 indicates significant prediction errors, and the R-squared value of 0.461 shows that only 46.1% of the variance in Bitcoin prices is explained, indicating moderate predictive power but room for improvement.

Random Forest Regressor Results

Figure 7 shows Bitcoin Price Prediction using the Random Forest model, where the blue line represents actual closing prices and the red line represents predicted prices. The model generally follows the actual price trend but shows sharp jumps, indicating sensitivity to data variations. With a relatively low MSE of 214,071.05 and an impressive R-squared value of 0.998, the Random Forest Regressor demonstrates a strong fit to the data, explaining nearly all of the variance in Bitcoin prices and suggesting high predictive accuracy.

Support Vector Machines Results

Figure 8 comparing Actual vs Predicted Bitcoin Closing Prices using the Support Vector Machine (SVM) model from 2015 to 2023 reveals that the SVM model struggles with accurate predictions. The red line, representing the predicted values, consistently underestimates the actual prices (blue line), with an MSE of 121,142,959.61 and an R-squared value of 0.142, indicating low predictive power and poor accuracy. The model's failure to capture the full volatility and sharp fluctuations of Bitcoin prices suggests that SVM, while capable of modelling general trends, is not well-suited for forecasting extreme price movements in highly volatile markets like cryptocurrencies.

Brief Comparison of Models (Table 2)

Exploring Other Metrics

Figure 9 highlights a minor positive correlation between Bitcoin's closing price and trading volume, with higher volumes concentrated in the $30,000 to $60,000 price range, suggesting significant market activity driven by investor enthusiasm or specific market events. At lower

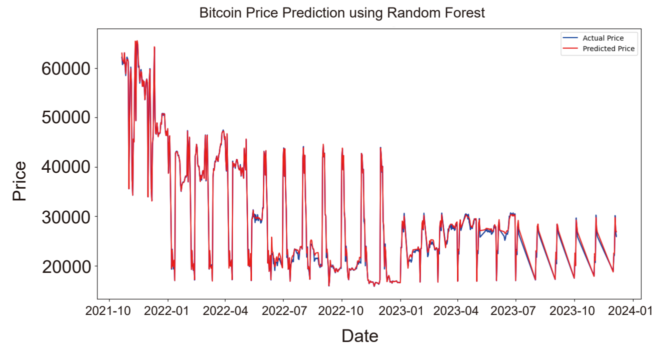

Figure 7. Actual vs Predicted Closing Prices for Bitcoin using RFR.

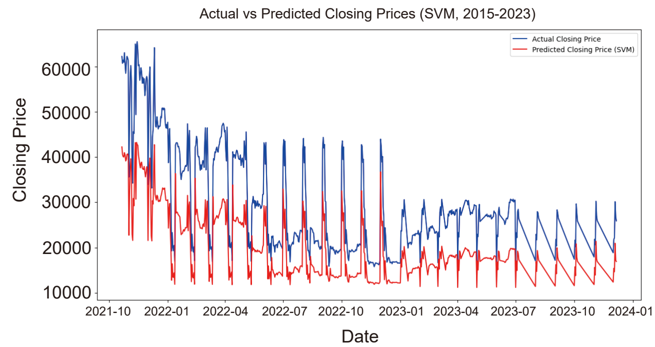

Figure 8. Actual vs Predicted Closing Prices for Bitcoin using SVM.

Table 2. Model Comparison

Model	MSE Value	Advantages	Disadvantages
LSTM	84,173,172.78	- Effective at capturing long-term dependencies - Suitable for time series prediction	- Computationally expensive - Requires large amounts of data for training
Random Forest Regressor	214,071.05	- Handles non-linear data well - Robust to overfitting with ensemble learning	- Slower for real-time predictions - Harder to interpret results
Support Vector Machine	121,142,959.61	- Works well with high-dimensional data - Effective in cases with clear margin of separation	- Sensitive to feature scaling - Difficult to tune hyperparameters

price levels (below $30,000), trading volume is more dispersed, indicating fewer high-volume days, potentially linked to market corrections or sell-offs. Additionally, extreme trading volumes (3.5*10^11) show varying prices, pointing to instances of heightened volatility or reactions to news. This underscores the importance of analysing both price and volume together to better understand market sentiment, especially in speculative markets like cryptocurrencies.

Analysis of Results

The performance of three machine learning models—Support Vector Machine (SVM), Random Forest Regressor (RFR), and Long Short-Term Memory (LSTM) networks—was evaluated for predicting Bitcoin's closing prices over the period from 2015 to 2023. The results highlight significant differences in predictive accuracy, with Random Forest performing the best, followed by LSTM, while SVM struggled to capture key price dynamics.

Table 3 summarises the key performance metrics of each model:

Key Findings.

- **1.** Random Forest's Superiority: Random Forest demonstrated the highest accuracy, with the lowest MSE and highest R-squared value of 0.998. This indicates that the model explains nearly all of the variance in Bitcoin's price movements. Its ensemble learning approach allows it to handle complex, non-linear relationships in the data, making it well-suited for Bitcoin price prediction in dynamic market conditions.
- 2. Limitations of LSTM: The LSTM model, while capturing broader trends, showed significant deviation during high-volatility periods. With an MSE of 84,173,172.78 and an R-squared value of 0.461, LSTM performed moderately, indicating its strength in stable market conditions but its vulnerability during price surges and dips.
- 3. SVM's Underperformance: SVM, with an MSE of 121,142,959.61 and an R-squared of 0.142, underperformed significantly. It was unable to capture Bitcoin's sharp price fluctuations, which are crucial in predicting extreme market conditions. This suggests that SVM is less suited for highly volatile assets like Bitcoin.

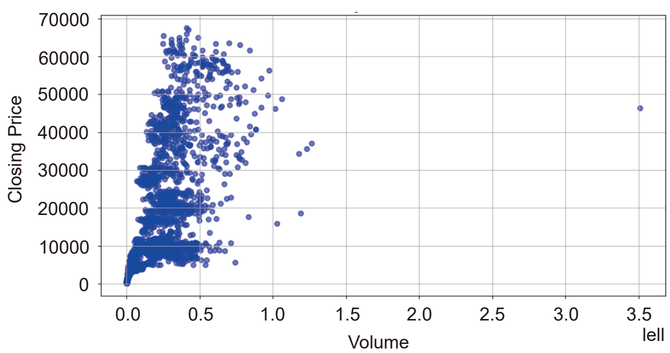

Figure 9. Closing Price vs Volume for Bitcoin.

Connection to Real-World Crypto Market Challenges. The performance differences between these models underscore the challenges of predicting cryptocurrency prices, which are highly volatile and affected by factors like market sentiment, regulatory changes, and investor behaviour. Random Forest's high R-squared value indicates its potential for more accurate predictions, offering critical insights for traders who need to manage risk in real-time. By leveraging models like Random Forest, investors can better forecast price trends, enabling them to make more informed decisions about when to buy, sell, or hold Bitcoin.

Adaptive Risk Management. The findings are particularly relevant to adaptive risk management in cryptocurrency trading. In volatile markets, accurate price prediction is crucial for mitigating risks. Random Forest, with its robustness and predictive power, can be used to identify potential price movements and adjust investment strategies accordingly. For example, the model can help investors understand when the market is likely to experience sharp corrections or significant price surges, allowing them to hedge their investments and reduce

Table 3.

Model	MSE	R-squared	Key Trends Observed
SVM	121,142,959.61	0.142	Underpredict Bitcoin prices, struggles with volatility
Random Forest	214,071.05	0.998	Excellent fit to data, captures trends well
LSTM	84,173,172.78	0.461	Good at capturing long-term trends, but struggles with volatility

exposure to extreme price changes.

LSTM, while valuable for capturing long-term trends, could be combined with Random Forest for a hybrid model, leveraging both models' strengths. This approach could enhance risk management by providing both short-term volatility insights and long-term trend forecasting.

DISCUSSION

Problems Faced

I faced a plethora of challenges with regards to coding, plotting graphs, etc. during the entire research project. One of the primary difficulties was dealing with the scaling of data, particularly when using Support Vector Machines (SVM). SVMs are sensitive to feature scaling, and the initial attempts to train the model without properly scaling the input data led to inaccurate predictions. The closing prices of Bitcoin, spanning several years, varied widely in range, while other features like trading volume had different scales. Without normalisation, the SVM model performed poorly, as it was unable to balance the influence of each feature properly. After identifying this issue, I had to ensure that MinMaxScaler or StandardScaler was applied to the data, which significantly improved the model's predictions. However, even after scaling, SVM still underperformed compared to other models, which led to exploring other algorithms (26).

Another challenge I faced was handling the gaps and irregularities in the dataset. The Bitcoin market is open 24/7, and my dataset included several days of missing data, particularly weekends or holidays when trading volume was low. This led to abrupt jumps in the graph, which confused the models during the training process. I had to clean the dataset, ensuring the time series data was continuous and reliable, either by filling in missing data points using interpolation techniques or by removing inconsistencies. Managing this missing data and ensuring proper preprocessing were critical steps that consumed a significant amount of time. In retrospect, handling the data more thoroughly in the initial stages would have saved time and effort during model training.

Another issue involved tuning hyperparameters for each model. While Random Forests and LSTM were more forgiving in their performance, the SVM and even the Random Forest model required careful tuning of parameters such as the number of estimators, the kernel function for SVM, and learning rates. Hyperparameter optimization was time-consuming as I had to rely on GridSearchCV and cross-validation, which took significant computational resources. This process was essential in extracting optimal performance from the models but added a layer of complexity that required constant troubleshooting and experimentation.

Outcomes

The results of this research focus on the performance of machine learning models, specifically Random Forest Regressor, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks, in predicting the closing prices of Bitcoin from 2015 to 2023. The dataset used consisted of various features, including Date, Open, High, Low, Close, AdjClose, and Volume, with a focus on predicting the 'Close' price. Each model's performance was evaluated by comparing the predicted prices with the actual closing prices of Bitcoin during this period.

The Random Forest model yielded the most accurate predictions overall, as evident from the graph comparing actual versus predicted closing prices. It was able to capture the trends of price movements, particularly during periods of high volatility, such as the Bitcoin price spikes in late 2021. The Random Forest's capability to handle non-linear relationships and interactions between features contributed to its superior performance. It closely followed the pattern of the actual closing prices and had minimal deviations, making it the most reliable model among those tested. Additionally, the feature importance analysis showed that the 'Volume' and 'Open' price significantly influenced the model's predictions.

On the other hand, the SVM struggled to accurately predict the closing prices, especially during periods of sharp fluctuations. The SVM model was sensitive to the scaling of data, and even after applying appropriate scaling techniques, its predictions remained significantly off. The graph of SVM's predicted prices showed more erratic behaviour, particularly during price spikes, where it failed to capture the magnitude of price changes. LSTM, while generally more accurate than SVM, also had periods of underperformance but showed promise during steady price periods due to its ability to handle time-series data effectively. The overall findings indicate that Random Forest was the best model for predicting Bitcoin closing prices, with LSTM being a viable alternative for long-term trend analysis and SVM being the least effective.

Does the data support the hypothesis?

The results strongly support the initial hypothesis that machine learning models, particularly ensemble models like Random Forest, can offer accurate predictions for Bitcoin price movements. This hypothesis was grounded in the assumption that sophisticated algorithms could analyse large volumes of historical data to forecast future prices more effectively than traditional models. Random Forest's ability to consistently produce reliable results across various price ranges confirms this assumption. Additionally, LSTM, which was initially expected to handle time-series data exceptionally well, performed adequately, supporting the hypothesis, though not outperforming Random Forest. However, the underperformance of the SVM model challenges part of the hypothesis. The expectation that SVM, with appropriate scaling, would provide a good baseline model for prediction was not met. This discrepancy suggests that the complexity of Bitcoin's price movements and the inherent challenges in scaling features for SVM limited its potential. Thus, while the overall hypothesis stands validated, there were unexpected challenges with specific models like SVM.

Limitations of the study

One of the primary limitations of this study is the reliance on historical price data, which, while useful for training machine learning models, cannot account for sudden market shocks, regulatory changes, or unexpected economic events that drastically affect Bitcoin prices. The models' predictions are based purely on past trends and do not factor in external variables such as geopolitical events, market sentiment, or investor behaviour, which play a significant role in cryptocurrency markets. Another limitation was encountered during the SVM model's training, where scaling posed a challenge. While efforts were made to standardise the data appropriately, the erratic nature of Bitcoin price movements made it difficult for SVM to generalise well (29). Additionally, the study only focused on Bitcoin, limiting the generalizability of the findings to other cryptocurrencies. Other cryptocurrencies have different volatility patterns and market dynamics, which may require different approaches. Lastly, the time period studied, while spanning several years, may still be insufficient to fully capture all the long-term trends and anomalies within the cryptocurrency market.

Given the insights gained and the limitations identified, several next steps can be taken to enhance the study. First, future research should consider incorporating external variables into the models, such as news sentiment analysis, social media trends, and macroeconomic indicators, to capture a more holistic view of the factors influencing Bitcoin prices. Additionally, more advanced feature engineering techniques could be employed to better capture the relationship between volume and price, as well as other underutilised features. In terms of model development, it would be valuable to explore hybrid models that combine the strengths of Random Forest and LSTM, allowing for better handling of both short-term volatility and long-term trends. Furthermore, expanding the analysis to include other cryptocurrencies, such as Ethereum or Litecoin, would provide a broader understanding of whether these models perform consistently across different markets. Lastly, running the models on a live data stream to perform real-time predictions could be an exciting step forward in demonstrating the practical application of these techniques in the fast-moving world of crypto trading.

CONCLUSION

This research highlights the potential of machine learning models, particularly Random Forest and LSTM, in predicting Bitcoin price movements within the highly volatile cryptocurrency market. By evaluating and comparing the performance of Random Forest Regressor (RFR), Long Short-Term Memory (LSTM), and Support Vector Machine (SVM) models, we identified key strengths and limitations that offer insights for financial forecasting in this unique asset class. Random Forest emerged as the most reliable model in terms of predictive accuracy, achieving a low Mean Squared Error (MSE) and high R-squared value. Its ensemble learning approach proved adept at managing the complex, non-linear relationships present in Bitcoin's historical data. Conversely, while LSTM showed promise in capturing temporal dependencies, it struggled with extreme volatility, and SVM underperformed due to challenges in scaling and kernel selection.

The findings underscore the critical importance of feature selection, such as incorporating daily returns and trading volume, to enhance the models' accuracy. Moreover, the study revealed significant correlations between trading volume and price movements, providing valuable insights into market sentiment and activity. These results demonstrate that machine learning can provide nuanced predictions that account for the multifaceted factors influencing cryptocurrency prices, offering an edge over traditional financial models.

The implications of this research extend beyond Bitcoin price prediction to broader applications in financial markets. Traders, investors, and financial institutions can leverage these machine learning tools to navigate the unpredictability of cryptocurrencies more effectively, enhancing risk management and decision-making. For individual traders, understanding the predictive capabilities of AI models can support more informed strategies, potentially mitigating losses in volatile markets. Furthermore, these models can be adapted for other financial assets, including stocks, commodities, and forex, suggesting a wide range of applicability for financial forecasting across diverse markets.

While the study provides a solid foundation for AI-driven price prediction, it also sheds light on challenges and areas for future research. One significant limitation is the underperformance of the SVM model, primarily due to its sensitivity to data scaling and inability to capture Bitcoin's high volatility effectively. Enhancing SVM performance through advanced kernel techniques, feature scaling strategies, and hyperparameter optimization could be a focus for future work. Similarly, LSTM models, though effective at capturing sequential patterns, require further refinement to handle extreme price fluctuations better. Techniques such as hybrid modelling, combining LSTM with ensemble methods like Random Forest, could help address these limitations and improve prediction accuracy.

Additionally, this research highlights the need for continuous model retraining and adaptation to ensure relevance in the rapidly evolving cryptocurrency market. Incorporating external factors, such as macroeconomic indicators, geopolitical events, and social media sentiment, could further enrich the models and provide a more comprehensive understanding of price drivers. Exploring these directions would not only enhance model robustness but also address the dynamic nature of cryptocurrency markets.

From an academic perspective, the study contributes to the growing field of AI-driven financial analysis by demonstrating the utility of machine learning models in tackling the challenges posed by volatile assets. It aligns with prior research affirming the effectiveness of Random Forest and LSTM in financial forecasting while offering new insights into their relative strengths and weaknesses when applied to cryptocurrencies. Furthermore, the findings validate the hypothesis that advanced machine learning models, particularly ensemble methods like Random Forest, outperform traditional techniques in capturing complex relationships within financial data.

Practically, this research lays the groundwork for developing AI-powered tools tailored to cryptocurrency trading, which could democratise access to sophisticated analytics for retail investors. By making these tools user-friendly and widely available, the financial industry can bridge the gap between professional and individual traders, fostering greater market inclusivity and efficiency.

In conclusion, this study reinforces the transformative potential of AI and machine learning in financial forecasting. By demonstrating the applicability of models like Random Forest and LSTM in predicting Bitcoin prices, it sets the stage for future advancements in this domain. However, the limitations identified—such as SVM's underperformance and the need for broader feature incorporation—highlight areas where further research is necessary. Addressing these challenges will not only refine predictive accuracy but also expand the applicability of machine learning models to other financial contexts.

As the cryptocurrency market continues to evolve, the role of AI in navigating its complexities will undoubtedly grow. This research represents a step forward in leveraging machine learning for financial innovation, paving the way for more robust, adaptive, and accessible tools that can empower both individual traders and institutional investors in managing the risks and opportunities of this dynamic market.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my mentor, Kira Miller of Veritas AI at Lumiere Education, for her invaluable guidance and support throughout this research.

REFERENCES

1. Chen Y. Understanding Bitcoin: A Case Study Method to Understand Market Dynamics, Strategies, and Risks of Cryptocurrency. Advances in economics, management and political sciences. 2023 Dec 28; 62 (1): 61-68. https://doi.org/10.54254/2754-1169/62/20231315
2. Singh M. Exploring the Impact: AI Technology in the Cryptocurrency Market [Internet]. AIM Research | Artificial Intelligence Market Insights. 2024. Available from: https://aimresearch.co/market-industry/ai-in-cryptocurrency-the-crucial-role-of-ai-in-shaping-the-future-of-cryptocurrency (accessed on 2024-10-16)
3. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science [Internet]. 2021; 2 (3): 1-21. Available from: https://link.springer.com/article/10.1007/S42979-021-00592-X (accessed on 2024-9-26). https://doi.org/10.1007/s42979-021-00592-x
4. Hayes A. Understanding Time Series [Internet]. Investopedia. 2022. Available from: https://www.investopedia.com/terms/t/timeseries.asp (accessed on 2024-9-17)
5. Long Short-Term Memory Network - an overview | Scien-

ceDirect Topics [Internet]. www.sciencedirect.com. Available from: https://www.sciencedirect.com/topics/computer-science/long-short-term-memory-network (accessed on 2024-9-4)
6. R. Siva, Bharathi Subrahmanian, Chaturya. K. Analyzing the machine Learning methods to predict Bitcoin Pricing. World Journal of Advanced Research and Reviews [Internet]. 2024 Jan 30; 21 (1): 1288-1294. Available from: https://wjarr.com/sites/default/files/WJARR-2024-0074.pdf (accessed on 2024-9-10). https://doi.org/10.30574/wjarr.2024.21.1.0074
7. Erfanian S, Zhou Y, Razzaq A, Abbas A, Safeer AA, Li T. Predicting Bitcoin (BTC) Price in the Context of Economic Theories: A Machine Learning Approach. Entropy. 2022 Oct 18; 24 (10): 1487. https://doi.org/10.3390/e24101487
8. Tripathy N. Stock Price Prediction Using Support Vector Machine Approach [Internet]. Available from: https://www.dpublication.com/wp-content/uploads/2019/11/24-ME.pdf (accessed on 2024-9-10). https://doi.org/10.33422/conferenceme.2019.11.641
9. Eitrich T, Lang B. Efficient optimization of support vector machine learning parameters for unbalanced datasets. Journal of Computational and Applied Mathematics. 2006 Nov; 196 (2): 425-436. https://doi.org/10.1016/j.cam.2005.09.009
10. Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research. 2018 Oct; 270 (2): 654-669. https://doi.org/10.1016/j.ejor.2017.11.054
11. V Akila, MVS Nitin, I Prasanth, S Reddy, A Kumar. A Cryptocurrency Price Prediction Model using Deep Learning. Tummala SK, Kosaraju S, Bobba PB, Singh SK, editors. E3S Web of Conferences [Internet]. 2023 [cited 2023 July 7]; 391: 01112. Available from: https://www.e3s-conferences.org/articles/e3sconf/pdf/2023/28/e3sconf_icmed-icmpc2023_01112.pdf (accessed on 2024-8-27) https://doi.org/10.1051/e3sconf/202339101112
12. Vikram K, Sivaraman N, Balamurugan DrP. Crypto Currency Market Price Prediction Using Data Science Process. International Journal for Research in Applied Science and Engineering Technology. 2022 Feb 28; 10 (2): 1451-1454. https://doi.org/10.22214/ijraset.2022.40521
13. Mishra O, Joshi U, Patil H, Dongardive J, Student. Comparing LSTM and Random Forests for Stock Price Movement Forecasting. International Journal for Research Trends and Innovation [Internet]. 9:2456-3315. Available from: https://www.ijrti.org/papers/IJRTI2401047.pdf (accessed on 2024-8-30)
14. Bowden J, Gemayel R. Sentiment and trading decisions in an ambiguous environment: A study on cryptocurrency traders. Journal of International Financial Markets, Institutions and Money. 2022 Sep; 80: 101622. https://doi.org/10.1016/j.intfin.2022.101622
15. Hayes A. Understanding Trend Analysis and Trend Trading Strategies [Internet]. Investopedia. 2023. Available from: https://www.investopedia.com/terms/t/trendanalysis.asp (accessed on 2024-9-10)
16. Random Forest Algorithm in Machine Learning [Internet]. AlmaBetter. Available from: https://www.almabetter.com/bytes/tutorials/data-science/random-forest (accessed on 2024-9-10)
17. IBM. What is support vector machine? | IBM [Internet]. www.ibm.com. 2023. Available from: https://www.ibm.com/topics/support-vector-machine (accessed on 2024-8-25)
18. Rasifaghihi N. Support Vector Regressor - Theory and Coding Exercise in Python [Internet]. Medium. 2023. Available from: https://medium.com/@niousha.rf/support-vector-regressor-theory-and-coding-exercise-in-python-ca6a7dfda927 (accessed on 2024-9-10)
19. Alhamid M. LSTM and Bidirectional LSTM for Regression [Internet]. Medium. 2021. Available from: https://towardsdatascience.com/lstm-and-bidirectional-lstm-for-regression-4fddf910c655 (accessed on 2024-9-10)
20. Dutta A. Random Forest Regression in Python - GeeksforGeeks [Internet]. GeeksforGeeks. 2019. Available from: https://www.geeksforgeeks.org/random-forest-regression-in-python/ (accessed on 2024-9-10)
21. Spotfire | Demystifying the Random Forest Algorithm for Accurate Predictions [Internet]. Spotfire. Available from: https://www.spotfire.com/glossary/what-is-a-random-forest (accessed on 2024-9-10)
22. scikit-learn. 3.2.4.3.2. sklearn.ensemble.RandomForestRegressor - scikit-learn 0.20.3 documentation [Internet]. Scikit-learn.org. 2018. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html (accessed on 2024-9-4)
23. Shenoy K. LSTM Back-Propagation Derivation | Kartik Shenoy | Medium [Internet]. Medium. Medium; 2020 [cited 2024 Sep 10]. Available from: https://kartik2112.medium.com/lstm-back-propagation-behind-the-scenes-andrew-ng-style-notations-7207b8606cb2 (accessed on 2024-9-10)
24. Haugh M. Machine Learning for OR & FE Support Vector Machines (and the Kernel Trick) [Internet]. [cited 2024 Sep 10]. Available from: https://www.columbia.edu/~mh2078/MachineLearningORFE/SVMs_MasterSlides.pdf (accessed on 2024-9-2)
25. Random Forest vs Support Vector Machines vs Neural Network [Internet]. iunera. 2020. Available from: https://www.iunera.com/kraken/fabric/random-forest-vs-support-vector-machine-vs-neural-network/ (accessed on 2024-9-9)
26. Why is SVM sensitive to scaling of features? [Internet]. Cross Validated. 2017 [cited 2024 Sep 10]. Available from: https://stats.stackexchange.com/questions/279911/why-is-svm-sensitive-to-scaling-of-features (accessed on 2024-9-12)
27. Sebastião H, Godinho P. Forecasting and trading crypto-

currencies with machine learning under changing market conditions. Financial Innovation. 2021 Jan 6; 7 (1). https://doi.org/10.1186/s40854-020-00217-x

28. McNally S, Roche J, Caton S. Predicting the Price of Bitcoin Using Machine Learning. 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP). 2018 Mar; https://doi.org/10.1109/PDP2018.2018.00060
29. machine learning - When to use Random Forest over SVM and vice versa? [Internet]. Data Science Stack Exchange. Available from: https://datascience.stackexchange.com/questions/6838/when-to-use-random-forest-over-svm-and-vice-versa (accessed on 2024-9-15)
30. Hayes A. Financial Markets: The Basics [Internet]. Investopedia. 2024. Available from: https://www.investopedia.com/terms/f/financial-market.asp (accessed on 2024-10-2)
31. Cryptocurrency vs Stock Exchange: The Difference Between Crypto and Stock Exchanges [Internet]. www.cointree.com. Available from: https://www.cointree.com/learn/cryptocurrency-exchange-vs-stock-exchange/ (accessed on 2024-9-17)