International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Forecasting Cryptocurrency Markets: Predictive Modelling Using Statistical and Machine Learning Approaches

Shafeeq Ur Rahaman[1], Patchipulusu Sudheer[2], Mahe Jabeen Abdul[3]
[1]Monks, CA, USA
[2]Samsung, CA, USA
[3]San Jose State University, CA, USA

ABSTRACT: The rapidly evolving landscape of cryptocurrency markets presents unique challenges and opportunities. The significant daily variations in cryptocurrency exchange rates lead to substantial risks associated with investments in crypto assets. This study aims to forecast the prices of cryptocurrencies using advanced machine learning models. Among seven models that were tested for their prediction and validation efficiency, Neutral Networks performed the best with minimum error. Thus, Long Short-Term Memory (LSTM) neural networks were used for predicting future trends. LSTM model is well-suited for analyzing complex dependencies in financial data. Starting with historical data collection, data preprocessing, feature engineering, normalization and integrative binning, a comprehensive Exploratory Data Analysis (EDA) was conducted on 50 cryptocurrencies. Top performers were identified based on criteria such as trading volume, market capitalization, and price trends. The LSTM model was implemented using Python to predict 90-day price movements data to check intricate patterns and relationships. Model performance was validated by performance metrics such as MAE and RMSE. The findings align with the Adaptive Market Hypothesis (AMH) which suggests that cryptocurrency markets exhibit dynamic efficiency influenced by evolving market conditions and investor behavior. The study shows the potential of machine learning models in financial economics and their role in enhancing risk management strategies and investment decision-making processes.

KEYWORDS: Currency forecasting, Financial Economics, LSTM Neural Network, Machine learning, Model prediction.

INTRODUCTION

Cryptocurrency emerged as a revolutionary financial innovation by an anonymous entity known as Satoshi Nakamoto (January, 2020). Bitcoin was launched in 2009 and marked this beginning of a new era of digital currencies (Malik, 2016). Nakamoto's whitepaper described Bitcoin as a decentralized electronic cash system that facilitates secure direct transactions without relying on intermediaries (Nakamoto & Bitcoin, 2008). Following Bitcoin's success, numerous alternative cryptocurrencies, such as Ethereum, Ripple, and Litecoin, were developed. Today, the cryptocurrency landscape comprises thousands of digital assets with applications in finance, gaming, supply chain, and beyond (Chlioumi, 2022). The rapid evolution of cryptocurrency markets presents both opportunities and challenges for traders and investors. Unlike traditional financial markets, cryptocurrencies operate in a decentralized environment which is characterized by high fluidity and frequent price fluctuations. These unique market dynamics require advanced predictive modeling techniques to forecast future price movements accurately (Lubogo, 2022). Forecasting the behavior of cryptocurrencies can enhance trading strategies and support risk management decisions. Forecasting can inform regulatory frameworks and foster greater market confidence. As the cryptocurrency market continues to evolve, the ability to anticipate price movements and market behavior becomes increasingly vital for predicting sustainable growth and integration into the broader financial landscape (Nabila et al., 2021). However, the limitations of these linear models in capturing the non-linear and complex dependencies within financial time series have led to the exploration of machine learning techniques. Machine learning models have gained domimance in time series forecasting due to their capacity to learn complex patterns from historical data (Elsayed et al., 2021). LSTM models, a type of recurrent neural network (RNN), are widely used to predict long-term variations and sequential relationships. LSTM model is particularly well-suited for modeling the non-linear and temporal nature of cryptocurrency prices (Liu et al., 2020).

Studies have demonstrated the efficacy of LSTM models in predicting stock prices, foreign exchange rates and now cryptocurrency prices. For instance, Ghosh and Neufeld (2022) applied LSTMs to the S&P 500 index and found that LSTMs outperformed

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

traditional models in both accuracy and profitability (Ghosh et al., 2022). Similarly, N Latif et al. (2023) applied LSTMs to Bitcoin price prediction and demonstrated improved performance of LSTMs over ARIMA and other machine learning models (Latif et al., 2023). These findings show the potential of LSTM networks to model the complex, volatile nature of cryptocurrency markets. Recent literature also highlights the application of LSTMs and other deep learning models for cryptocurrency forecasting with mixed results depending on data preprocessing, feature selection, and model configuration (Awotunde et al., 2021; Mudassir et al., 2020). However, the efficacy of these models often depends on the quality of input features and the ability to handle the high variability inherent in cryptocurrency data.

Despite the growing interest in applying LSTM models to financial time series, there remains a gap in their comparative performance against traditional statistical approaches and other machine learning algorithms. This study aims to address this gap by employing seven different model and to shortlist the best model based on maximum efficiency and minimum error. Neural network model performed best showing maximum predictive precision and minimum error. LSTM Neural Network model was selected to predict cryptocurrency prices over a 90-day horizon. By using advanced data preprocessing techniques, feature engineering, normalization, scaling and model validation, this research seeks to provide a robust methodological framework for cryptocurrency market prediction. The results are expected to offer valuable insights into the effectiveness of LSTM models in capturing market dynamics and improving forecast accuracy in the rapidly evolving cryptocurrency space. The findings aim to contribute to the growing body of literature on machine learning applications in financial markets and it will also provide actionable insights for market participants.

METHODOLOGY

A comprehensive and validated methodology was applied for accurate prediction to analyze cryptocurrencies future market trends and patterns. The framework of this research methodology is illustrated in Figure 1.

Data Acquisition

To start with, recent historical daily market price datasets of the top 50 cryptocurrencies were obtained from CoinMarketCap via Yahoo Finance (https://finance.yahoo.com/markets/crypto/all/) using the python based approach. The yfinance library was employed to fetch datasets for each cryptocurrency by specifying their ticker symbols. Yahoo Finance is a widely recognized reliable and consistent financial data provider (Boritz & No, 2020). The time range for the data collection was set from January 1, 2018, to August 30, 2024. Older datasets of corresponding cryptocurrencies from April 28th, 2013 to December 31st, 2017 were retrieved from kaggle (https://www.kaggle.com/). Kaggle offers extensive collection of diverse and high-quality datasets (Quaranta et al., 2021). Among approximately 9000 active and traded cryptocurrencies in the market, 50 were selected based on their high trading volume in the market. The final dataset was exported as a CSV file. Both CSV files were combined to create a unified dataset. The files were first loaded into Python using the pandas library, and columns were aligned by reordering them to a standardized format.

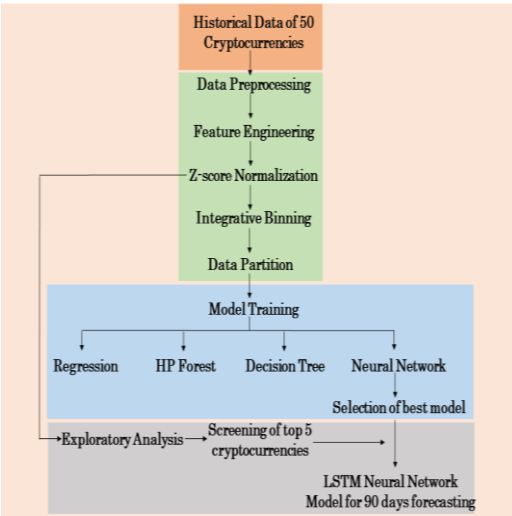

Figure 1. Experimental Flowchart for modelling and forecasting

Data Preprocessing

Data preprocessing was performed to ensure the integrity and utility of the dataset for subsequent analysis. Missing values within the dataset were identified and handled. Each column was examined for null values using the `isnull()` method from the Pandas library. Data integrity was assessed by identifying inconsistencies within the dataset. This involved cross-verifying values with expected ranges and patterns. Discrepancies were flagged for further review or correction. Data types were verified to ensure compatibility with subsequent analyses. Columns were checked for correct data types and converted as necessary to maintain consistency. Type conversion was carried out using the `astype()` method to ensure that all columns conformed to the expected data types. The preprocessing yielded a refined dataset ready for analysis. Imputation was applied to columns with a low percentage of missing values, while rows with extensive missing data were removed to maintain data quality.

Data mismatches were resolved to ensure that dataset was consistent and reliable. Values that did not conform to expected patterns were corrected or excluded. All columns were converted to appropriate data types for accurate analysis and preventing errors during processing.

Feature Engineering

Moving averages, volatility measures, and technical indicators were computed to enrich the dataset. Simple Moving Averages (SMA) and Exponential Moving Averages (EMA) were calculated to identify trends (Rusdiana et al., 2020). Average True Range (ATR) indicator was used to assess market volatility (Cohen, 2023). Various technical indicators were computed using the `pandas_ta` library, including Relative Strength Index (RSI) and Bollinger Bands (Daniswara et al., 2022).

Z-score Normalization

To ensure consistent scaling of the features used in our analysis, z-score normalization was applied to the processed dataset obtained from feature engineering (Urolagin et al., 2021). This step was crucial for standardizing the numerical features. Initially, the dataset was imported using Python's pandas library and non-numeric columns were excluded from the normalization process. The numeric features were then identified and isolated for scaling using the `StandardScaler` from the `scikit-learn` package (Zollanvari, 2023). The normalized features were subsequently reintegrated with the non-numeric data to maintain the original structure of the dataset. The complete normalized dataset was saved as a CSV file for subsequent analysis.

Exploratory Data Analysis (EDA)

Detailed data exploration was conducted to screen the top performers in cryptocurrency marker (Rouf et al., 2021). Trends were visualized by plotting closing prices of cryptocurrencies over time. A bar graph was plotted to compare the average closing values and trading volumes of cryptocurrencies over a specific period. This visualization allowed for a quick assessment of market dominance and overall performance. A correlation matrix was also constructed to evaluate the relationships between the selected features of the cryptocurrencies. Time series analysis of top cryptocurrencies based on the logarithmic prices was done (Fleischer et al., 2022).The log transformation was applied to stabilize variance and reveal long-term trends.

$$\text{Log Price} = \log(\text{Close Price})$$

This transformation helped highlight relative changes rather than absolute price differences. Based on the EDA results, the top 5 cryptocurrencies were selected using criteria such as high trading volume, strong market capitalization, and distinct trends identified in the log time series plots. These top performers were further analyzed for predictive modeling. The EDA was performed by manually data analysis and Python's data visualization libraries, including Matplotlib and Seaborn (Sial et al., 2021), to generate graphs and to identify top performers.

Model Training

SAS Enterprise Miner Client 15.2 was used for machine learning model training (Truong, 2024). The dataset was partitioned into training (60%), validation (20%) and test (20%) subsets. Integrative Binning was applied to categorize continuous variable (Ma et al., 2020). Binning was done to handle non-linear relationships and enhancing model performance. Four models i.e., Neural Networks, Regression Models, High Performance Forest and Decision Tree were trained and evaluated. Regression model assessed linear relationships between features and cryptocurrency prices (Akbulaev et al., 2020). Neural networks captured complex non-linear patterns through multi-layered architectures (Karn et al., 2024). HP forest implemented a high-performance random forest model to improve prediction accuracy through ensemble learning (Xin et al., 2023). Decision Tree utilized tree-based algorithms to interpret and visualize decision paths in predicting cryptocurrency prices (Naghib Moayed & Habibi, 2020). Each model's performance was compared with the normalized data. A final score node was used to aggregate predictions and assess model performance comprehensively. This node provided a detailed evaluation of each model's predictive capabilities.

LSTM Model for forecasting future trends

Since, Neural Network performed best in model training, so the Long Short-Term Memory (LSTM) Neural Network model was used for predicting future prices of top 5 cryptocurrencies (Lindemann et al., 2021). LSTMs mitigate the vanishing gradient problem inherent in traditional RNNs by incorporating a memory cell structure that retains information across long time steps. This is facilitated by three essential components: the forget gate, the input gate, and the output gate (Song et al., 2020).

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

LSTM networks consist of a series of repeating modules (cells) with four interacting layers: the input layer, hidden layers with LSTM units, and an output layer. The core mechanism of an LSTM cell is governed by the 4 gates. The fundamental operation of an LSTM cell is controlled by four gates. The forget gate determines which pieces of information should be removed from the cell state. It utilizes a sigmoid activation function to produce a value ranging from 0 to 1, where 0 indicates 'completely discard' and 1 signifies 'fully retain'.

$$f_t = \sigma (W_f [h_t-1, xt]+bf)$$

Where f_t represents the forget gate output at time step t, W_f is the weight matrix, h_{t-1} is the previous hidden state, x_t is the current input, b_f is the bias term and σ denotes the sigmoid activation function. Input Gate determines which new information should be stored in the cell state. It consists of a sigmoid layer that decides which values to update and a tanh layer that creates a vector of new candidate values.

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

$$C_t=\tanh (W_c\cdot [h_{t-1}, x_t] + b_C)$$

C_t denotes the candidate cell state and W_i,W_c are the weight matrices. The cell state C_t is formed by integrating the previous cell state with the newly acquired information. Finally, the output gate generates the output by processing the updated cell state (Zu & Wei, 2021).

$$C_t = f_t \times C_{t-1} + i_t \times C_t$$

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

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

The LSTM model was implemented in Python using the Keras library, using the TensorFlow backend (Florencio & Moreno, 2021). The dataset was preprocessed by normalizing price-related features, including close prices, log returns, and other technical indicators. Data was split into training (80%) and testing (20%) sets and sequences were created to capture past price movements over a 90-day time window.

The normalized data was reshaped into a three-dimensional format required by LSTM (samples, time steps, features). The model was trained for 50 epochs with a batch size of 32. A 10% validation split was employed to track performance during training (Lechner & Hasani, 2020). The trained model was used to forecast cryptocurrency prices for the next 90 days.

Model Validation

The performance of the model was rigorously assessed using several statistical evaluation metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Final Prediction Error (Khan et al., 2020). These metrics provide quantitative measures of the accuracy and reliability of the model's predictions. The results from these evaluations highlighted the model's proficiency in capturing the intricate temporal dependencies present in cryptocurrency market data.

Results

Data Acquisition and Preprocessing

The data acquisition process successfully gathered comprehensive historical data for 50 major cryptocurrencies, spanning from April 2013 to August 2024. This dataset includes crucial market indicators such as opening price, closing price, highest price, lowest price, and trading volume (in US Dollars) for each day. This comprehensive historical range allows for in-depth analysis of market behavior under different economic scenarios. The combined CSV file, generated from the two input files, contained all the necessary columns in the specified order. Figure 2 is representing the volume of 50 selected cryptocurrencies over the course of 11 years.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

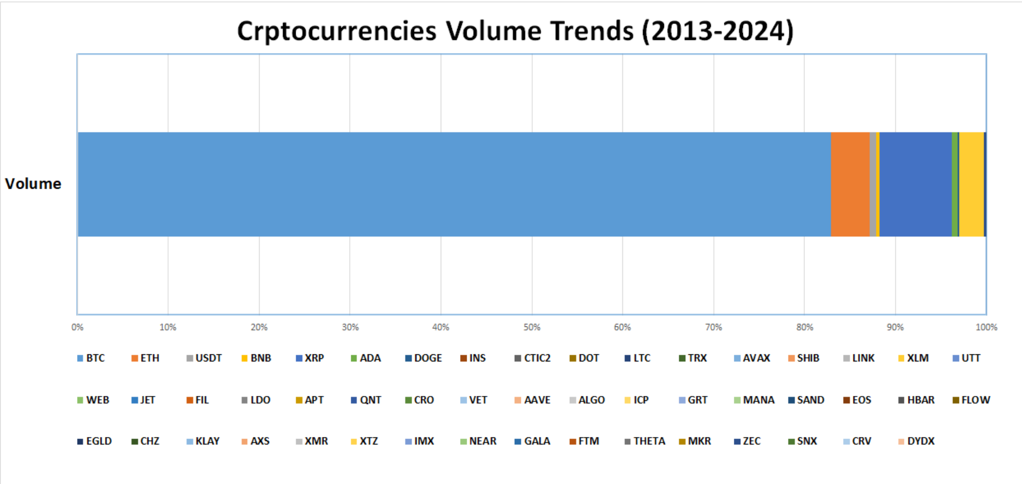

Figure 2 Volume Trends of 50 selected cryptocurrencies over time

Feature Engineering

The calculation of moving averages, volatility measures, and technical indicators provided valuable insights. The inclusion of SMA and EMA helped in trend analysis, while volatility measures such as ATR offered a deeper understanding of market dynamics. Technical indicators enriched the dataset and gave comprehensive features for predictive modeling and analysis. These preprocessing and feature engineering efforts laid a solid foundation for subsequent analysis and modeling tasks. Figure 3 is indicating the SMA 20, EMA 20, and Bollinger Bands indicators. The SMA 20 is the average price of an asset over the last 20 periods. It smooths out price data to identify trends more clearly. When prices are above the SMA 20, it often signals an uptrend (Peng et al., 2021). The EMA 20 is similar to SMA but gives more weight to recent prices. It can be useful for detecting short-term momentum (Market, 2006). Bollinger Bands comprise a SMA with upper and lower bands set at a specified distance, typically ±2 standard deviations (Huyen Chau & Doan, 2024). Price touching the upper band indicate potential overbought condition while price touching the lower band shows potential oversold condition. These indicators help in understanding the market's past behavior. The high peak in late 2015 likely reflects a period of extreme market sentiment driven by the significant market events.

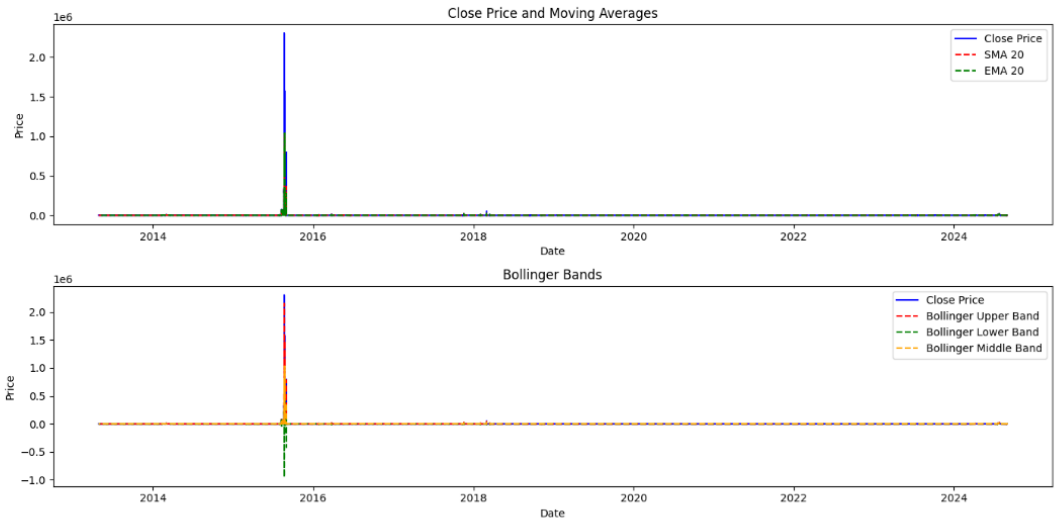

Figure 3 Moving Averages and Bollinger Bands during Feature Engineering

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Z-score Normalization

The z-score normalization successfully standardized the numeric features within the dataset. This step mitigated the effects of different scales and ranges among the features and optimized the data for further statistical analysis and machine learning application. Upon normalization, visual inspection of the data confirmed the effectiveness of the transformation as all numerical variables exhibited a uniform scale. The data was devoid of any outliers that could skew analysis results. The normalized dataset was saved and utilized in subsequent modeling, contributing to improved model convergence and predictive performance.

Exploratory Data Analysis

Among the 50 cryptocurrencies analyzed, Bitcoin, Ethereum, Binance Coin, Litecoin, and Maker emerged as the top performers due to their consistent high trading volumes and significant average price levels. They exhibited clear upward trends and distinctive market behaviors. Figure 4 (a) is representing the closing prices of these cryptocurrencies over time. Bitcoin has the most successful trend since 2020 followed by Maker. Figure 4 (b) is exhibiting the correlation matrix among top performers. The correlation matrix revealed varying degrees of correlation between the closing prices with some exhibiting strong positive correlations. For instance, Bitcoin and Ethereum showed a high correlation which is reflecting their influence on the broader market. Figure 4 (c) is representing the 3d bar graphs of cryptocurrencies. The bar graph is comparing the average closing prices and trading volumes of selected cryptocurrencies. This graphical representation represents the dominance of these cryptocurrencies in the market. This visualization helped in validating the selection of these cryptocurrencies for predictive modeling. These analyses not only facilitated the identification of the top-performing cryptocurrencies but also provided a comprehensive understanding of market dynamics.

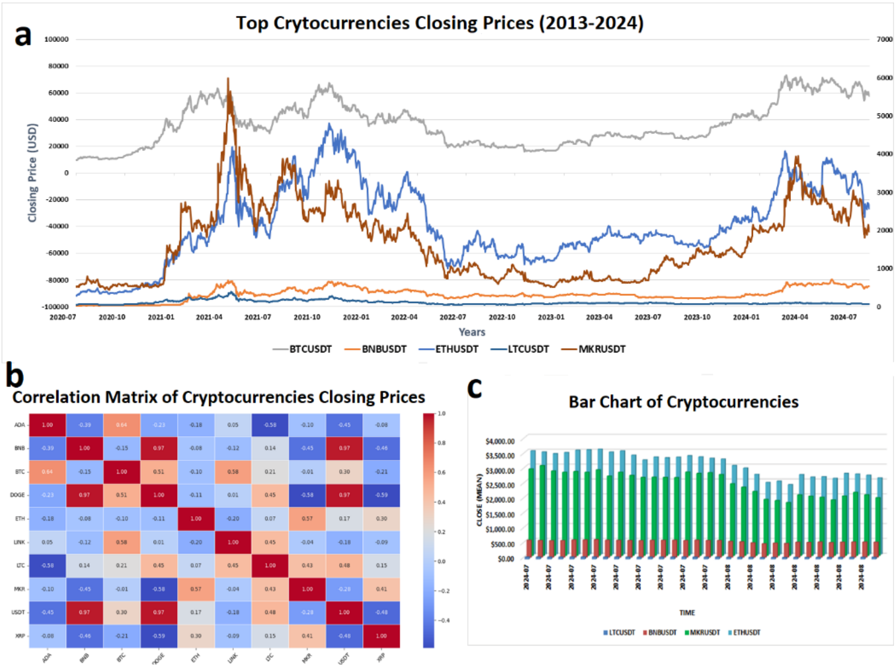

Figure 4 Exploratory data analysis of cryptocurrencies (a) trends of top cryptocurrencies over time (b) Correlation Matrix of Cryptocurrencies (c) 3d Bar Graphs of selected Cryptocurrencies.

The logarithmic-transformed time series graphs were also plotted for each cryptocurrency and revealed significant patterns and growth trends among the cryptocurrencies. The log plots are usually used to demonstrate financial values and effectively highlight periods of exponential growth, corrections, and consolidations (Sung et al., 2022). The logarithmic graphs are illustrated in Figure 5.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

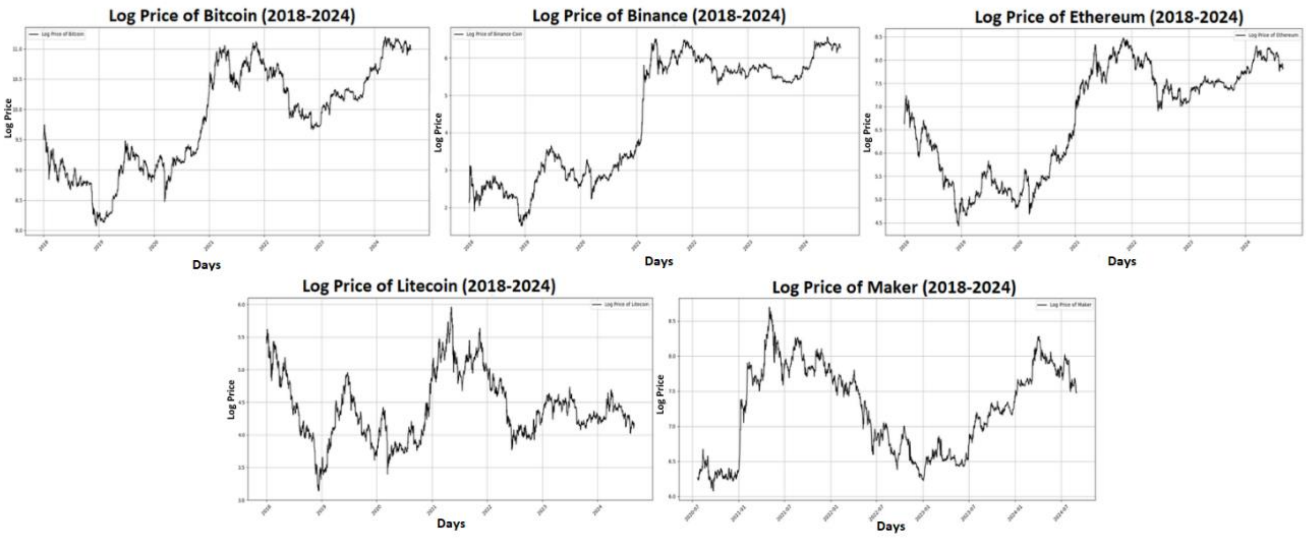

Figure 5 Logarithmic Time series graphs of selected cryptocurrencies.

Model Comparison

Model comparison revealed distinct performance characteristics for each model. Regression Model demonstrated a solid baseline performance. Neural Networks effectively capture complex patterns and non-linear relationships. HP Forest provided robust performance with enhanced predictive accuracy due to its ensemble approach. Decision Tree offered interpretability and clear decision paths. Among these models, the neural network and Decision Tree models were superior in forecasting accuracy and handling data complexity. The score node analysis highlighted these models as the most reliable for predicting cryptocurrency price trends. Neural Networks had the lowest average squared error followed by Decision Tree as shown in the Table 1. Supplementary Figure 6 is representing the model score of Train, Validation and Test data.

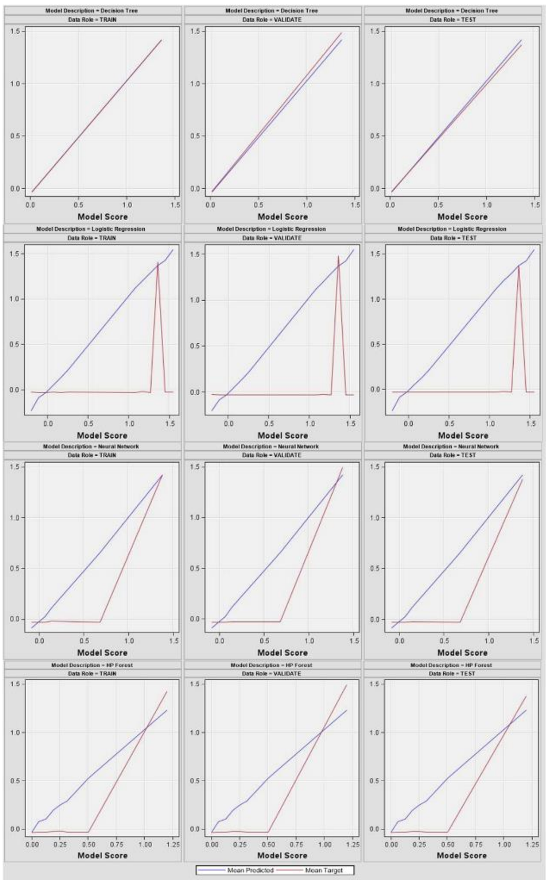

Figure 6 Score of applied Models offered by SAS E-Miner

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Table 1 Valid Average Squared Error of applied Models

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Selection Criteria: Valid Average Squared Error
Y	Neural	Neural	Neural Network	close	1.01450912
	Tree	Tree	Decision Tree	close	1.014666461
	Reg	Reg	Logistic Regression	close	1.016109443
	Reg2	Reg2	Stepwise Regression	close	1.016111252
	Reg3	Reg3	Forward Regression	close	1.016111252
	Reg4	Reg4	Backward Regression	close	1.016111252
	HPDMForest	HPDMForest	HP Forest	close	1.01674716

LSTM Model of 90-days forecasting

The LSTM model effectively captured the sequential patterns and variations within the cryptocurrency price data. The 90-day forecasting results showed that the model could anticipate general trends and turning points in the market with reasonable accuracy. The LSTM model aligned closely with historical price patterns as shown in Figure 6. The results highlight the model's capacity to learn from complex, non-linear market dynamics and provide a valuable tool for forecasting future cryptocurrency prices. Graphical representation indicates the nearly constant trends for Bitcoin. Bitcoin is the oldest and most widely adopted cryptocurrency and exhibits relatively stable price trends compared to others. Moreover, Bitcoin's market is less reactive to news and speculative events. A slightly downward future trends are observed for Ethereum and Binance Coin while constantly fluctuating trends were observed for Maker and Litecoin. Both Ethereum and Binance Coin may be experiencing market corrections following periods of rapid growth. While Maker and Litecoin have smaller market caps and are more susceptible to price fluctuations due to lower liquidity and higher speculative trading.

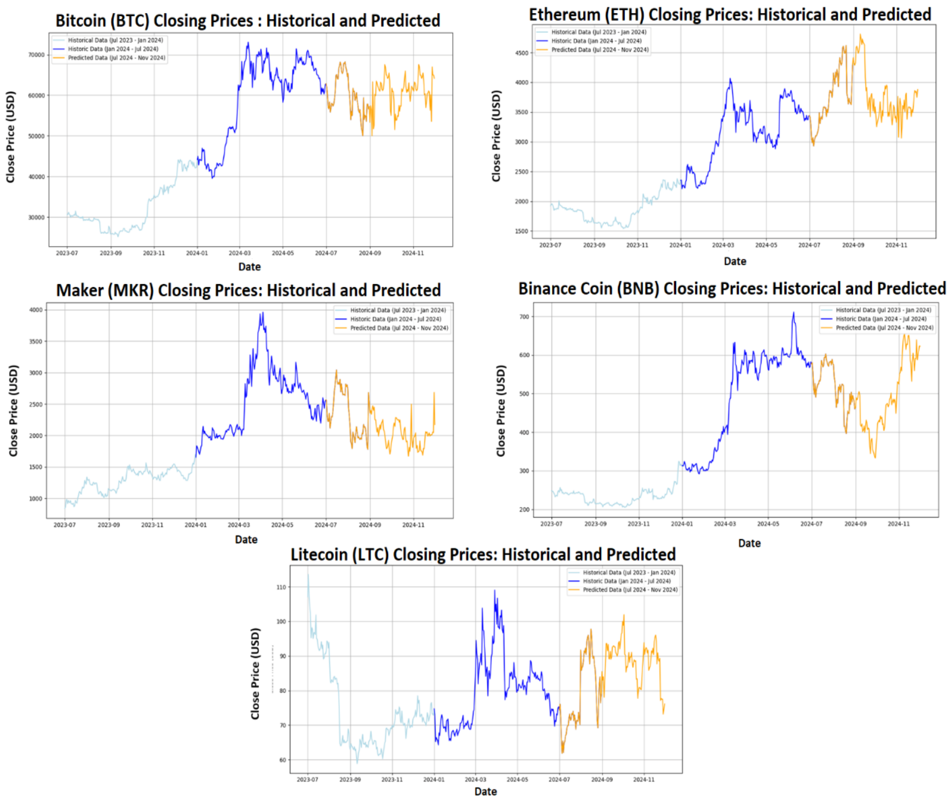

Figure 7 90-days forecasting of Bitcoin, Ethereum, Maker, Binance Coin and Litecoin utilizing LSTM Model

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

The average absolute deviation of predictions from actual prices was minimized, indicating a close fit between predicted and observed values. RMSE highlighted the model's sensitivity to larger deviations and confirmed the model's overall precision. The low percentage error reflected the model's consistency across the test period. The LSTM model forecasted upward and downward trends with significant precision. The results aligned closely with historical price patterns.

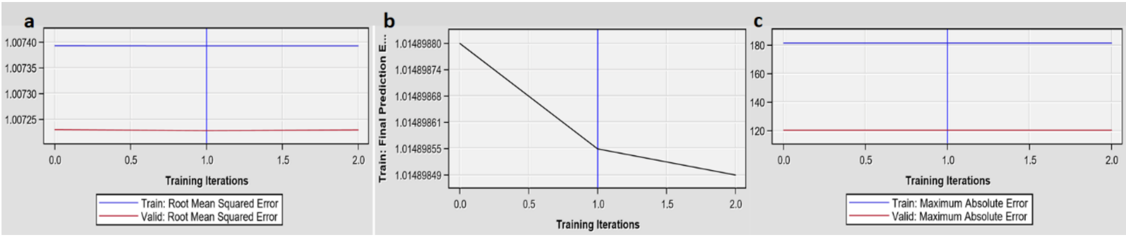

Figure 8 Root Mean Square Error (RMSE), Maximum Absolute Error (MAE) and Final Prediction Error of LSTM Model

DISCUSSION

Cryptocurrencies have emerged as a significant financial innovation and has revolutionized the landscape of digital assets. The volatility and fluctuating nature of cryptocurrencies have been a subject of extensive debate in the literature. Studies by Agyei and Adam (2022) and Obeng et al. (2023) highlight the extreme price volatility of cryptocurrencies and the absence of intrinsic value (Agyei et al., 2022; Obeng, 2023). Further, research by Mpinyuri et al. (2019) demonstrated the systemic risk posed by cryptocurrencies due to their lack of regulation and susceptibility to market manipulation (Mpinyuri, 2019). On the other hand, studies such as Habib and Sharma (2022) and Santos (2017) argue that blockchain technology offer unique advantages including lower transaction costs, transparency, and enhanced security (Dos Santos, 2017; Habib et al., 2022). Despite these benefits, empirical analyses indicate that the unpredictable price swings and regulatory uncertainties limit their acceptance as stable financial instruments (Borio et al., 2001). Thus, practical integration of cryptocurrencies into mainstream finance remains fraught with challenges. Forecasting cryptocurrency trends holds significant potential benefits, particularly in enhancing decision-making for investors and financial institutions. Previous studies such as those by Fang and Ventre (2022) demonstrate that accurate trend predictions can aid investors in mitigating risks associated with the high volatility of digital assets (Fang et al., 2022). Moreover, Gosh et al. (2019) also argue that predictive models provide better understanding into market dynamics (Ghosh et al., 2019). Additionally, the work of Marzo et al. (2022) highlights the growing importance of reliable forecasting in stabilizing the cryptocurrency ecosystem (Marzo et al., 2022).

This study aims to predict the 90-days market trend of top cryptocurrencies in the market. The Exploratory Data Analysis identified significant trends and patterns among the 50 analyzed cryptocurrencies and top 5 cryptocurrencies in the market were shortlisted. Bitcoin is the first recognized cryptocurrency and serves as the foundation of the entire digital asset market. Introduced by Nakamoto (2008), Bitcoin revolutionized the concept of money by enabling person to person transactions without the interference of banks (Nakamoto & Bitcoin, 2008). Its decentralized nature has positioned Bitcoin as a "digital gold" (Taskinsoy, 2021). As noted by Riggs (2022), Bitcoin's status as the flagship cryptocurrency solidifies its importance as a key indicator of market trends and a gateway for broader blockchain adoption (Riggs & Vyas, 2022). Ethereum, as highlighted by Oliva (2020), revolutionized blockchain technology by introducing smart contracts (Oliva et al., 2020). Binance Coin (BNB) was basically launched as a token for reduced trading fees on the Binance platform but now has evolved into a multi-functional asset that powers the Binance Smart Chain (Cernera et al., 2023). Litecoin, often referred to as the "silver to Bitcoin's gold," offers faster transaction times and a more efficient mining process (Novak, 2023). Maker, as noted by (Nedashkovskiy, 2021) is a cornerstone of the DeFi ecosystem that provide a decentralized stablecoin (DAI).

Following data collection and preprocessing, feature engineering and z-score normalization was done. Feature engineering enhances the predictive power of models by capturing essential market dynamics, such as moving averages, volatility indicators, and price momentum. These engineered features allow models to better understand underlying patterns. Z-score normalization standardizes data by scaling features to a common scale. This process is particularly important in cryptocurrency data where values can vary widely, leading to biased model training if not addressed. Normalization ensures that each feature contributes equally to the model and prevent any one variable from disproportionately influencing the results (Kappal, 2019).

Multiple models were selected to evaluate their predictive capabilities, including HP Forest, Decision Tree, Default Regression, Forward Regression, Stepwise Regression, Backward Regression, and Neural Network. Decision Trees and HP Forest are robust non-linear models that excel in capturing complex interactions within the data. These models effectively handle high-dimensional datasets without overfitting (Jain & Chauhan, 2019). These regression techniques are especially useful in identifying the most influential predictors among engineered features. Neural Networks included for their exceptional ability to capture non-linear patterns and learn from large volumes of data through deep learning architectures (LeCun et al., 2015). The comparative analysis of these models revealed that the Neural Network outperformed others so this model was considered as the optimal model for the 90-day forecasting task.

LSTM model has been used previously to forecast different aspects of financial economics. The predicted future fluctuations of the selected five cryptocurrencies suggest a diverse range of market behaviors. The relatively stable future trends observed for Bitcoin imply continued market confidence and its role as a store of value (Giraldo). In contrast, the slightly downward trends projected for Ethereum and Binance Coin may indicate market corrections following their rapid growth phases, influenced by evolving regulatory landscapes and competitive pressures. The fluctuating trends observed for Litecoin and Maker highlight their susceptibility to broader market sentiment and lower liquidity compared to larger-cap assets. Maker's volatility is particularly tied to its integral role within the DeFi space (Boukas, 2023; Mounter). For investors, these anticipated fluctuations provide critical insights for strategic portfolio adjustments.

The study's findings align with the Adaptive Market Hypothesis (AMH), which suggests that market efficiency is dynamic and evolves with changing market conditions. The successful application of LSTM models in forecasting suggests that machine learning can significantly enhance traditional financial analysis tools. This represents a shift towards more sophisticated financial technologies (FinTech) that uses big data and AI to improve decision-making processes.

CONCLUSION

This study investigated the use of machine learning techniques for predicting cryptocurrency prices. The research began with the historical data collection, preprocessing, normalization, scaling and model training. Comprehensive Exploratory Data Analysis (EDA) of 50 cryptocurrencies, identified the top performers through log time series analysis, bar graphs, correlation matrices, and closing value line graphs. The implementation of the LSTM model demonstrated strong predictive accuracy in forecasting 90-day price movements. Performance metrics such as MAE, RMSE, and MAPE were employed to validate the model's robustness. The study's results have significant implications for financial economics and the broader financial markets. They highlight the transformative role of machine learning models in refining risk management and optimizing investment portfolios. The ability to forecast price movements with high accuracy can provide investors and financial institutions with a more informed decision-making in a rapidly evolving market landscape. Future research should focus on integrating other financial indicators and assessment of the impact of external factors such as regulatory changes and macroeconomic events on cryptocurrency prices.

STATEMENTS & DECLARATIONS

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

1. Agyei, S. K., Adam, A. M., Bossman, A., Asiamah, O., Owusu Junior, P., Asafo-Adjei, R., & Asafo-Adjei, E. (2022). Does volatility in cryptocurrencies drive the interconnectedness between the cryptocurrencies market? Insights from wavelets. Cogent Economics & Finance, 10(1), 2061682.
2. Akbulaev, N., Mammadov, I., & Hemdullayeva, M. (2020). Correlation and Regression Analysis of the Relation between Ethereum Price and Both Its Volume and Bitcoin Price. Journal of Structured Finance, 26(2), 46-56.
3. Awotunde, J. B., Ogundokun, R. O., Jimoh, R. G., Misra, S., & Aro, T. O. (2021). Machine learning algorithm for cryptocurrencies price prediction. In Artificial intelligence for cyber security: methods, issues and possible horizons or opportunities (pp. 421-447). Springer.

4. Borio, C., Furfine, C., & Lowe, P. (2001). Procyclicality of the financial system and financial stability: issues and policy options. BIS papers, 1(3), 1-57.
5. Boritz, J. E., & No, W. G. (2020). How significant are the differences in financial data provided by key data sources? A comparison of XBRL, Compustat, Yahoo! Finance, and Google Finance. Journal of Information Systems, 34(3), 47-75.
6. Boukas, I. (2023). Exploring dynamics in cryptocurrencies' exchange rates.
7. Cernera, F., La Morgia, M., Mei, A., & Sassi, F. (2023). Token Spammers, Rug Pulls, and Sniper Bots: An Analysis of the Ecosystem of Tokens in Ethereum and in the Binance Smart Chain (BNB). 32nd USENIX Security Symposium (USENIX Security 23)
8. Chlioumi, A. (2022). The new distributed ledger technologies in banking transactions and transaction banking platforms.
9. Cohen, G. (2023). Trading cryptocurrencies using algorithmic average true range systems. *Journal of Forecasting*, *42*(2), 212-222.
10. Daniswara, D. A., Widjanarko, H., & Hikmah, K. (2022). The Accuracy Test of Technical Analysis of Moving Average, Bollinger Bands, and Relative Strength Index on Stock Prices of Companies Listed in Index Lq45. *Indikator*, *6*(2), 411842.
11. Dos Santos, R. P. (2017). On the philosophy of bitcoin/blockchain technology: is it a chaotic, complex system? *Metaphilosophy*, *48*(5), 620-633.
12. Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., & Schmidt-Thieme, L. (2021). Do we really need deep learning models for time series forecasting? *arXiv preprint arXiv:2101.02118*.
13. Fang, F., Ventre, C., Basios, M., Kanthan, L., Martinez-Rego, D., Wu, F., & Li, L. (2022). Cryptocurrency trading: a comprehensive survey. *Financial Innovation*, *8*(1), 13.
14. Fleischer, J. P., von Laszewski, G., Theran, C., & Parra Bautista, Y. J. (2022). Time series analysis of cryptocurrency prices using long short-term memory. *Algorithms*, *15*(7), 230.
15. Florencio, F., & Moreno, E. D. (2021). Benchmarking the Keras API on GPU: the use of tensorflow and CNTK libraries as back-end. *International Journal of High Performance Computing and Networking*, *17*(1), 19-27.
16. Ghosh, I., Jana, R. K., & Sanyal, M. K. (2019). Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms. *Applied Soft Computing*, *82*, 105553.
17. Ghosh, P., Neufeld, A., & Sahoo, J. K. (2022). Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters*, *46*, 102280.
18. Giraldo, J. Trust in Bitcoin as a store of value.
19. Habib, G., Sharma, S., Ibrahim, S., Ahmad, I., Qureshi, S., & Ishfaq, M. (2022). Blockchain technology: benefits, challenges, applications, and integration of blockchain technology with cloud computing. *Future Internet*, *14*(11), 341.
20. Huyen Chau, N. T., & Doan, T. P. (2024). Data Processing and Feature Engineering for Stock Price Trend Prediction. In *Machine Learning and Other Soft Computing Techniques: Biomedical and Related Applications* (pp. 183-197). Springer.
21. Jain, S., & Chauhan, D. (2019). Standard Multiple Regression Analysis Model for Cell Survival/Death Decision of JNK Protein Using HT-29 Carcinoma Cells.
22. January, B. (2020). *Cryptocurrencies and the Blockchain Revolution: Bitcoin and beyond*. Millbrook Press.
23. Kappal, S. (2019). Data normalization using median median absolute deviation MMAD based Z-score for robust predictions vs. min–max normalization. *London Journal of Research in Science: Natural and Formal*, *19*(4), 39-44.
24. Karn, P. K., Ardekani, I., & Abdulla, W. H. (2024). Generalized Framework for Liquid Neural Network upon Sequential and Non-Sequential Tasks. *Mathematics*, *12*(16), 2525.
25. Khan, Z. A., Hussain, T., Ullah, A., Rho, S., Lee, M., & Baik, S. W. (2020). Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework. *Sensors*, *20*(5), 1399.
26. Latif, N., Selvam, J. D., Kapse, M., Sharma, V., & Mahajan, V. (2023). Comparative performance of LSTM and ARIMA for the short-term prediction of bitcoin prices. *Australasian Accounting, Business and Finance Journal*, *17*(1), 256-276.
27. Lechner, M., & Hasani, R. (2020). Learning long-term dependencies in irregularly-sampled time series. *arXiv preprint arXiv:2006.04418*.
28. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

29. Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia Cirp*, *99*, 650-655.
30. Liu, Y., Gong, C., Yang, L., & Chen, Y. (2020). DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. *Expert Systems with Applications*, *143*, 113082.
31. Lubogo, I. (2022). *Digital money*. Jescho publishing house.
32. Ma, Z., Liu, Z., Zhao, Y., Zhang, L., Liu, D., Ren, T., Zhang, X., & Li, S. (2020). An unsupervised crop classification method based on principal components isometric binning. *ISPRS International Journal of Geo-Information*, *9*(11), 648.
33. Malik, V. (2016). The history and the future of Bitcoin. *Praha: Bankovní institut vysoká škola Praha*.
34. Market, C. (2006). Technical analysis of the currency market.
35. Marzo, G. D., Pandolfelli, F., & Servedio, V. D. (2022). Modeling innovation in the cryptocurrency ecosystem. *Scientific Reports*, *12*(1), 12942.
36. Mounter, D. Exploring the Evolution and Future Perceptions of Cryptocurrencies in Financial Markets.
37. Mpinyuri, E. B. (2019). *Beyond cryptocurrencies: financial applications of blockchain technology*. University of Johannesburg (South Africa).
38. Mudassir, M., Bennbaia, S., Unal, D., & Hammoudeh, M. (2020). Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. *Neural computing and applications*, 1-15.
39. Nabila, S., Usman, M., Indryani, N., & Kurniasari, D. (2021). Dynamic modeling data time series by using constant conditional correlation-generalized autoregressive conditional heteroscedasticity. Journal of Physics: Conference Series,
40. Naghib Moayed, A., & Habibi, R. (2020). Crypto-Currency Price Prediction with Decision Tree Based Regressions Approach. *Journal of Algorithms and Computation*, *52*(2), 29-40.
41. Nakamoto, S., & Bitcoin, A. (2008). A peer-to-peer electronic cash system. *Bitcoin.–URL: https://bitcoin. org/bitcoin. pdf*, *4*(2), 15.
42. Nedashkovskiy, A. (2021). Understanding DeFi ecosystem and how can it change or transform existing financial system.
43. Novak, I. (2023). *A systematic analysis of cryptocurrencies* University of Zagreb. Faculty of Economics and Business].
44. Obeng, J. (2023). *Impact of cryptocurrency volatility on stock market performance* University of Education, Winneba].
45. Oliva, G. A., Hassan, A. E., & Jiang, Z. M. (2020). An exploratory study of smart contracts in the Ethereum blockchain platform. *Empirical Software Engineering*, *25*, 1864-1904.
46. Peng, Y., Albuquerque, P. H. M., Kimura, H., & Saavedra, C. A. P. B. (2021). Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators. *Machine Learning with Applications*, *5*, 100060.
47. Quaranta, L., Calefato, F., & Lanubile, F. (2021). KGTorrent: A dataset of python jupyter notebooks from kaggle. 2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR),
48. Riggs, W., & Vyas, V. (2022). Current State of Blockchain and Cryptocurrency for Major US Cities. *Available at SSRN 4337656*.
49. Rouf, N., Malik, M. B., & Arif, T. (2021). Predicting the stock market trend: an ensemble approach using impactful exploratory data analysis. International Conference on Information, Communication and Computing Technology,
50. Rusdiana, S., Yuni, S. M., & Khairunnisa, D. (2020). Comparison of Rainfall Forecasting in Simple Moving Average (SMA) and Weighted Moving Average (WMA) Methods (Case Study at Village of Gampong Blang Bintang, Big Aceh District-Sumatera-Indonesia. *Journal of Research in Mathematics Trends and Technology*, *2*(1), 21-27.
51. Sial, A. H., Rashdi, S. Y. S., & Khan, A. H. (2021). Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python. *International Journal*, *10*(1), 277-281.
52. Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L., & Cheng, Z. (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, *186*, 106682.
53. Sung, S.-H., Kim, J.-M., Park, B.-K., & Kim, S. (2022). A Study on Cryptocurrency Log-Return Price Prediction Using Multivariate Time-Series Model. *Axioms*, *11*(9), 448.
54. Taskinsoy, J. (2021). Bitcoin: A New Digital Gold Standard in the 21st Century? *Available at SSRN 3941857*.
55. Truong, D. (2024). *Data Science and Machine Learning for Non-Programmers: Using SAS Enterprise Miner*. CRC Press.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 10 October 2024
DOI: 10.47191/ijcsrr/V7-i10-15, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

56. Urolagin, S., Sharma, N., & Datta, T. K. (2021). A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting. *Energy*, *231*, 120963.
57. Xin, R., Liu, H., Chen, P., & Zhao, Z. (2023). Robust and accurate performance anomaly detection and prediction for cloud applications: a novel ensemble learning-based framework. *Journal of Cloud Computing*, *12*(1), 7.
58. Zollanvari, A. (2023). Supervised Learning in Practice: the First Application Using Scikit-Learn. In *Machine Learning with Python: Theory and Implementation* (pp. 111-131). Springer.
59. Zu, X., & Wei, K. (2021). A simple gated recurrent network for detection of power quality disturbances. *IET Generation, Transmission & Distribution*, *15*(4), 751-761.