Bridging Complexity and Practicality in Bitcoin Forecasting: Artificial Intelligence and Machine Learning for Predictive Financial Modeling

Treit Phan
Department of Applied Engineering
North Carolina A&T State University
tphan@aggies.ncat.edu

Collins E. Okafor*
Department of Accounting and Finance
North Carolina A&T State University
ceokafor@ncat.edu

Nacasius U. Ujah
Ness School of Management and Economics
South Dakota State University
Nacasius.Ujah@sdstate.edu

Winifred Okafor
Department of Finance
University of Texas – San Antonio
winnieajalie@gmail.com

Devang Mehta,
Department of Applied Engineering
North Carolina A&T State University
tphan@aggies.ncat.edu

Mohsen Souissi,
Department of Accounting
Fayetteville State University
msouissi@uncfsu.edu

*** Corresponding Author: Dr. Colins E Okafor (ceokafor@ncat.edu)**

Abstract

Cryptocurrencies, particularly Bitcoin, have gained significant attention due to their high volatility and potential for substantial returns. Yet accurately forecasting their price remains challenging for both financial institutions and individual traders. In this research, we employ a multiple linear regression approach—readily implemented in standard spreadsheet environments—to predict Bitcoin price movements using a historical dataset that includes daily prices and trading volumes. Through feature engineering, we derive additional indicators such as moving averages, volatility (standard deviations), and momentum metrics (e.g., rate of change), which serve to enhance model accuracy. We evaluate predictive performance using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). Our findings demonstrate that a well-constructed linear regression model can provide valuable insights into short-term price trends in the cryptocurrency market, even with relatively limited computational resources. This work contributes to the field of AI-driven financial management by illustrating that simpler machine learning methods can still deliver practical predictive power in highly volatile markets like Bitcoin. Future research will explore richer data sources (e.g., social sentiment, macroeconomic indicators) and potentially more advanced algorithms for further improvements in price forecasting.

I. Introduction

Bitcoin has emerged as a transformative financial asset, revolutionizing traditional economic structures and reshaping global investment paradigms. As the flagship cryptocurrency, Bitcoin has defied conventional financial logic—experiencing unprecedented price surges, dramatic crashes, and growing mainstream adoption. Its decentralized nature sets it apart from traditional assets, yet it also amplifies volatility, making Bitcoin both an opportunity and a challenge for traders, institutional investors, and policymakers. Unlike fiat currencies or equity markets, Bitcoin's price is not dictated by central banks or corporate earnings but instead by a complex interplay of technological advancements, market sentiment, liquidity cycles, and regulatory actions. Understanding and forecasting these price dynamics remains one of the most pressing challenges in modern financial analytics.

The ability to predict Bitcoin price movements with precision presents an unparalleled opportunity. Investors seek predictive insights to maximize returns, policymakers require market stability measures, and financial analysts need data-driven strategies to navigate the evolving landscape of digital assets. However, Bitcoin's price behavior is highly erratic, driven by nonlinear relationships between market factors, making traditional forecasting models inadequate. The question then arises: can machine learning (ML) unlock hidden patterns in Bitcoin's volatility and provide a reliable predictive framework?

Artificial Intelligence (AI) and Machine Learning (ML) have already transformed predictive analytics in financial markets, enhancing risk management, trading strategies, and asset valuation techniques. In the cryptocurrency domain, ML models hold the potential to revolutionize forecasting by identifying hidden correlations, capturing nonlinear trends, and adapting to rapidly shifting market conditions. However, while advanced ML architectures such as Long Short-Term Memory (LSTM) networks, Support Vector Machines (SVMs), and deep learning models have demonstrated promise in time-series forecasting, they often require specialized computing environments and vast computational resources, limiting their accessibility for many market participants.

In contrast, Multiple Linear Regression (MLR)—a fundamental yet powerful statistical technique—remains a viable approach for interpretable, computationally efficient, and accessible financial forecasting. Although MLR is often overshadowed by complex black-box models, it still serves as an essential tool, particularly when combined with feature engineering techniques that extract valuable insights from historical price movements, trading volumes, and market volatility indicators.

This research seeks to bridge the gap between complexity and practicality in cryptocurrency forecasting. By refining a feature-enhanced MLR model, we aim to demonstrate that even a simple yet well-structured ML approach can provide meaningful predictive power in highly volatile markets like Bitcoin. Our motivation is threefold: (I) Practicality & Accessibility – Many traders and financial analysts lack access to high-performance computing resources required for deep learning models. Our study assesses whether spreadsheet-friendly machine learning approaches, enhanced with engineered features, can still yield significant predictive insights. (ii) Market Efficiency & Investment Decision-Making – Bitcoin has evolved into a mainstream financial instrument, attracting institutional investors, hedge funds, and governments. However, many forecasting models fail to capture the erratic, sentiment-driven nature of Bitcoin price fluctuations. This study evaluates how trading volume, volatility, and momentum metrics can enhance forecasting precision. (iii) Policy & Risk Management – Bitcoin's price remains highly sensitive to regulatory changes. Understanding the relationship between market behavior and external shocks (e.g., government crackdowns, central bank policies) is critical for developing risk mitigation strategies. Our study contributes by identifying key indicators that could improve forecasting accuracy in periods of extreme volatility.

This study applies a Multiple Linear Regression (MLR) model enhanced with feature engineering techniques to predict Bitcoin's price movements based on historical data, trading volume, volatility indicators, and momentum metrics. The research makes the following key contributions: (i) Methodological Advancement – We enhance traditional MLR by incorporating technical indicators such as moving averages, rolling standard deviation, and rate of change, refining its forecasting capability. (ii) Practical Forecasting Framework – Unlike computationally expensive ML models, our approach provides a transparent, interpretable, and spreadsheet-implementable method that financial analysts can easily integrate into decision-making. (ii) Empirical Insights on Bitcoin Volatility – Our results indicate that MLR effectively captures short-term Bitcoin price trends, particularly in stable market conditions. However, the model's accuracy declines during high-volatility periods, emphasizing the need for hybrid approaches integrating sentiment analysis and macroeconomic data.

In sum, this research contributes to the broader discourse on AI-driven financial forecasting and its implications for decentralized finance (DeFi), algorithmic trading, and digital asset risk management. As the adoption of cryptocurrencies accelerates, accessible, interpretable, and computationally efficient AI models will play a crucial role in financial decision-making. Our study informs both academic research and industry practice, offering a scalable predictive framework that can be integrated into FinTech applications, risk assessment models, and institutional crypto investment strategies. As the cryptocurrency industry matures, hybrid methodologies that merge AI-driven techniques with traditional financial modeling will shape the next frontier of price forecasting. Understanding these dynamics is not just an academic endeavor—it is a crucial step toward navigating the future of digital finance.

II. Literature Review

The prediction of Bitcoin price movements has gained significant attention due to the cryptocurrency's volatile nature and its potential for high returns. Researchers have explored various machine learning models in predicting these price movements, using time-series analysis based on historical data. This review synthesizes the most relevant works on machine learning approaches for Bitcoin price prediction, particularly focusing on the strengths and limitations of these models.

Machine Learning in Time-Series Analysis

Machine learning models have proven valuable in time-series forecasting, especially for complex and volatile financial instruments like Bitcoin. Although ARIMA and related statistical approaches have historically served as a foundation for financial forecasting, the erratic price swings and nonlinear patterns inherent to cryptocurrencies have encouraged the adoption of more flexible machine learning models. For instance, Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVM) can capture dependencies and trends that simpler models might miss.

Mallqui and Fernandes [2] demonstrated the potential for algorithms such as Artificial Neural Networks (ANNs) and SVMs to enhance Bitcoin price prediction. Their empirical results showed up to a 10% improvement in directional prediction accuracy and a Mean Absolute Percentage Error (MAPE) as low as 1–2% for price regressions, highlighting machine learning's capability to outperform traditional linear models in volatile markets.

Machine Learning Models for Bitcoin Price Prediction

Numerous studies underscore the effectiveness of machine learning for Bitcoin price forecasting. In Chen's [3] work, the author identified critical factors impacting Bitcoin price from 2015–2018, notably certain stock market indices and oil prices. After 2018, additional factors like Ethereum prices and the JP225 index surfaced as significant drivers. By incorporating these explanatory variables, the model achieved high accuracy in next-day price predictions.

Caliciotti et al. [4] evaluated multiple forecasting horizons, comparing regression and machine learning approaches, particularly SVMs. Their findings suggest that methods not assuming strict linearity often yield more reliable long-term forecasts. Similarly, Loh's [5] research illustrated the strength of Nonlinear Autoregressive with External Input (NARX) models in highly volatile settings, outperforming both Feedforward Neural Networks (FNN) and basic NAR models.

Advanced Models and Hybrid Approaches

More advanced or hybrid machine learning models have emerged in pursuit of improved predictive performance. Liu et al. [6] implemented a Stacked Denoising Autoencoder (SDAE), reporting superior results compared to Backpropagation Neural Networks (BPNN) and SVR in predicting both price direction and actual price levels. The SDAE excelled at reducing errors such as MAPE and RMSE, emphasizing the value of deep architectures for recognizing hidden patterns in Bitcoin's volatile data.

Likewise, research combining sentiment analysis or other external signals with machine learning has led to noticeable performance gains. For instance, Mallqui and Fernandes [2] found that integrating textual data with standard forecasting models can significantly boost predictive accuracy.

Future Directions and Challenges

Despite these advancements, major hurdles persist. Overfitting is a recurring issue, especially when data is relatively sparse or when markets shift abruptly. The cryptocurrency market's sensitivity to various internal (e.g., technical protocols) and external (e.g., macroeconomic conditions, regulatory decisions) forces further complicates real-time predictions. Chen [3] highlights the importance of capturing newer market influencers, such as Ethereum or global indexes, to improve forecast robustness—though truly "dynamic" models may still be required for sudden regime shifts.

Finally, hybrid methodologies—marrying machine learning with econometric techniques—are increasingly important for refining Bitcoin price forecasts. Caliciotti et al. [4] and Liu et al. [6] both confirm that ensemble or deep learning approaches hold promise for further reducing errors and enhancing directional accuracy, indicating a trend toward blending traditional statistical methods with advanced algorithms.

Machine learning, applied to time-series contexts, has demonstrably improved Bitcoin price predictions by handling nonlinear and volatile behaviors beyond the reach of simpler techniques like ARIMA. Although methods such as

LSTM, SVM, and various hybrid models often show higher precision, they can be hampered by overfitting, data constraints, and the complexities involved in merging external variables. The enhanced multiple linear regression framework proposed in this paper builds on these insights, combining advanced feature engineering and a carefully structured linear model to bridge the gap between purely linear approaches and more complex ML architectures. By refining linear assumptions while explicitly incorporating critical indicators—both market-based and exogenous—this method achieves robust interpretability and capacity to capture broader crypto market dynamics. Future work expanding the dataset to include social media sentiment, macroeconomic signals, or on-chain metrics could further elevate the accuracy and reliability of this framework, continuing to push Bitcoin price forecasting to new frontiers.

III. Methodology

1. Data Collection and Preprocessing

Data Source: The primary dataset used for this research consists of historical Bitcoin prices and trading volumes [7]. The dataset includes the following columns:

- date: Human-readable date in "YYYY-MM-DD" format.
- symbol: The currency pair (in this case, BTC/USD).
- open: Opening price of Bitcoin for the day.
- high: Highest price during the day.
- low: Lowest price during the day.
- close: Closing price for the day.
- Volume BTC: Amount of Bitcoin traded during the day.
- Volume USD: Total USD value of Bitcoin traded during the day.

Table 1: BTC Data 10/08/2015-12/24/2024 (Showing 6 of 3719 rows).

Date	Symbol	Open	Close	High	Low	Volume BTC	Volume USD
10/8/2015 4:00	BTC/USD	0	243.6	245	0	34.7547026	8466.245554
10/9/2015 4:00	BTC/USD	243.6	245.51	249.97	243.6	61.58706754	15120.24095
10/10/2015 4:00	BTC/USD	245.51	246.3	246.3	244.6	30.8705493	7603.416293
…	…	…	…	…	…	…	…
12/22/2024 0:00	BTC/USD	97228.21	95101.08	97369.85	94197.58	227.8231474	21666227.37
12/23/2024 0:00	BTC/USD	95101.08	94769.87	96432.74	92370.22	1546.246843	146537612.3
12/24/2024 0:00	BTC/USD	94769.87	98590.44	99500	93399.69	1111.585304	109591684.2

This data is well-structured as time series data, which is highly suitable for applying machine learning models for prediction and financial analysis.

Data Cleaning: To ensure the integrity of the data, the following preprocessing steps will be undertaken:

- Handling Missing or Erroneous Entries: Any rows with missing or inconsistent values will be either removed or imputed based on previous/following data points to ensure continuity.
- Data Normalization: Since financial data can have wide-ranging values, normalization or scaling will be applied to bring all features into a similar range. This ensures that no single feature dominates the model due to its scale. Techniques like Min-Max scaling or Z-score normalization may be applied depending on the chosen models.

Feature Engineering: To improve the performance of the machine learning models, additional features will be engineered from the raw data:

Price Differences: Calculate the difference between key price points, such as high-low and open-close, to capture daily volatility and price movements.

Moving Averages:

- Simple Moving Average (SMA) [8]: A rolling average computed over a fixed window of a quarter to smooth out short-term fluctuations and highlight long-term trends. A 4-quarter SMA at time iii is the average of the current and previous 3 quarters ($i \geq 4$):

$$SMA_i = \frac{P_i + P_{i-1} + P_{i-2} + P_{i-3}}{4}$$

- Exponential Moving Average (EMA) [9]: Places more weight on recent data points, making it more responsive to short-term price changes. This can help capture immediate market trends more effectively than the SMA. A 4-quarter EMA uses:

$$\alpha = \frac{2}{4+1} = 0.4$$

$$EMA_1 = P_1$$

$$EMA_i = \alpha \, . \, P_i + (1 - \alpha) \, . \, EMA_{i-1}$$

Price Momentum Indicators:

- Rate of Change (ROC) [10]: Measures the percentage change in price over a specified number of periods, capturing price momentum. A 1-quarter ROC ($i \geq 2$) is:

$$ROC_i(\%) = \frac{P_i - P_{i-1}}{P_{i-1}} \, . 100\%$$

- Relative Strength Index (RSI) [11]: An oscillator that compares the magnitude of recent gains to recent losses, indicating whether an asset is overbought or oversold. Here, we illustrate a 4-period RSI (instead of the more typical 14 daily periods). At each i:

$$Gain = G_i = \max(P_i - P_{i-1}, 0)$$

$$AvgGain_i = \frac{G_i + G_{i-1} + G_{i-2} + G_{i-3}}{4}$$

$$Loss = L_i = \max(P_{i-1} - P_i, 0)$$

$$AvgLoss_i = \frac{L_i + L_{i-1} + L_{i-2} + L_{i-3}}{4}$$

$$RSL_i = 100 - \frac{100}{1 + \frac{AvgGain_i}{AvgLoss_i}}$$

$$If \ AvgLoss_i = 0, RSI \ is \ 100 \ (pegged \ at \ maximum)$$

Volatility Indicators:

- Rolling Standard Deviation (RSD) [12]: A measure of the dispersion of prices over a specified window, useful for capturing market volatility. Higher rolling standard deviations indicate greater price variability and risk.

$$StdDev_i = \sqrt{\frac{1}{4-1}\sum_{k=i-3}^{i}(P_k - \overline{P_i})^2}$$

$$Where\ \overline{P_i}\ is\ the\ average\ of\ \{P_{i-3}, P_{i-2}, P_{i-1}, P_i\}$$

These engineered features will help capture both short-term price movements and longer-term trends, improving the accuracy of the machine learning models applied to predict future Bitcoin prices.

2. Data Splitting

Training and Testing Split:

In order to evaluate the performance of the predictive models, the dataset is split into two parts: a training set and a testing set. The training set will be used to train the machine learning models, while the testing set will serve as an independent set to evaluate model performance. A common practice in time-series prediction is to use an 80/20 split, where 80% of the data is allocated for training and the remaining 20% is reserved for testing [13]. This ensures that the model is exposed to a significant portion of the dataset during training, while leaving enough data for robust evaluation.

Since cryptocurrency prices are time-dependent, the split will be done chronologically to ensure that future data is not used to predict past events. This preserves the temporal structure of the data, a critical consideration for time-series modeling.

Time Horizon:

This study will focus on two types of predictions:

- Short-term predictions (1-quarter ahead): The model will attempt to predict the closing price for the next day based on historical price data. This short-term horizon is useful for traders looking for quick gains in volatile markets like Bitcoin.
- Mid-term predictions (semi-annual predictions): In this case, the model will predict the closing price half year ahead. This mid-term horizon is more suitable for longer-term investors who are looking to capitalize on broader market trends.

To effectively manage these time horizons, a sliding window approach will be employed. In this approach, a fixed-size window of past data (e.g., the last four quarters) will be used to make a prediction for the next day or week. As new data becomes available, the window will slide forward, incorporating the most recent observations while discarding the oldest data. This ensures that the model is continuously updated with the most recent market trends, improving its predictive accuracy. This combination of a well-defined training/testing split and the use of a sliding window will ensure that the model is robust and adaptable to different market conditions.

3. Machine Learning Models

In this study, we focus on multiple linear regression as a straightforward yet effective machine learning approach to predict Bitcoin's closing price [14]. While advanced algorithms (e.g., random forests, support vector regression, LSTMs) have shown promise in capturing nonlinear relationships in cryptocurrency data, they typically require specialized libraries or programming environments. By contrast, a multiple linear regression model is both implementable in Excel (using built-in tools like the Data Analysis ToolPak) and interpretable enough to serve as a baseline for future expansions. Below is an overview of the linear regression framework used, followed by a brief discussion of other methods that, although beyond the scope of an Excel-only workflow, remain relevant in broader Bitcoin price forecasting research [15].

Multiple Linear Regression:

- Rationale: Provides a direct way to model Bitcoin's closing price (Y) using one or more predictors, such as trading volume or time index.
- Implementation Feasibility: Easily carried out in Excel via the Regression tool or LINEST function.
- Limits: Assumes a (mostly) linear relationship between features and outcome, which may struggle with highly nonlinear crypto markets.

In summary, multiple linear regression serves as the principal modeling approach for this study, balancing ease of use and the ability to run entirely within Excel. Future work may integrate more advanced algorithms if a specialized environment is available, potentially providing greater accuracy by capturing Bitcoin's well-documented nonlinear and volatile behaviors.

4. Model Evaluation

Evaluating the performance of the machine learning models is crucial to understanding how well they predict Bitcoin price movements. The following evaluation metrics will be used to assess the accuracy and robustness of each model.

Mean Absolute Error (MAE) [16]: MAE is a metric that calculates the average absolute difference between the predicted values and the actual values. It measures how close the predictions are to the real prices without considering the direction of the error (overestimation or underestimation). A lower MAE indicates better performance.

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

Where y_i represents the actual values and $\hat{y}_i$ represents the predicted values.

Mean Squared Error (MSE) [17]: MSE is another common metric for regression models, which calculates the average squared difference between the actual and predicted values. Unlike MAE, MSE penalizes larger errors more heavily due to the squaring, making it sensitive to outliers. A lower MSE indicates better predictive accuracy.

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

Root Mean Squared Error (RMSE) [18]: RMSE is the square root of MSE and provides a measure of the error in the same units as the predicted values (e.g., dollars for Bitcoin prices). This makes RMSE easier to interpret than MSE. RMSE is especially useful for understanding how well the model predicts the magnitude of price changes.

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

R-squared (R^2) [19]: R^2 is a metric that indicates how much of the variance in the dependent variable (Bitcoin price) is explained by the independent variables (e.g., past prices, volume). R^2 values range from 0 to 1, where a value closer to 1 indicates that the model explains a greater portion of the variance in the data.

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

Where $\bar{y}$ is the mean of the actual value.

Accuracy (for classification-based models): In cases where the model predicts whether Bitcoin's price will increase or decrease (a classification task), accuracy is the proportion of correct predictions out of the total predictions. This metric is useful when evaluating classification models rather than regression models [20].

To ensure that the models are not overfitting to a specific subset of the data, k-fold cross-validation will be used. In this method, the dataset is split into k equally sized subsets (or "folds"). The model is trained on $k-1$ folds and tested on the remaining fold. This process is repeated k times, with each fold used once as the test set. The results from all folds are averaged to obtain a more robust estimate of the model's performance [21].

For this study, a 5-fold cross-validation approach will be implemented to ensure that the models generalize well to unseen data. This technique helps reduce bias and variance, providing a more reliable measure of the model's ability to predict Bitcoin prices across different data subsets.

IV. Analysis and Results

1. Descriptive Analysis

To gain an initial understanding of the dataset and identify patterns relevant to Bitcoin price prediction, a comprehensive descriptive analysis was conducted. This analysis involves exploring historical trends in Bitcoin closing prices, trading volumes, and other key metrics. By visualizing these data points, it becomes possible to uncover temporal patterns, correlations, and potential anomalies that may influence predictive modeling. Key insights derived from descriptive statistics and exploratory data analysis (EDA) serve as a foundation for the subsequent application of machine learning techniques. Through this process, we aim to establish a deeper context for interpreting the behavior of the Bitcoin market and its inherent volatility [22].

Table 2: Descriptive Statistic of Bitcoin Data 10/08/2015-12/24/2024.

Descriptive Statistics	Open	Close	High	Low	Volume BTC	Volume USD
Mean	22141.85088	22161.22984	22686.88434	21572.33133	2411.079923	34297306.26
Standard Error	348.7378942	349.2528184	356.8056185	340.3425061	55.01077216	740842.1718
Median	16212.39	16247.48	16624.93	15710	1338.341267	19180692.98
Mode	6555	6555	17000	16900	0	0
Standard Deviation	21267.29376	21298.6957	21759.29267	20755.31274	3354.755165	45179225.89
Sample Variance	452297784	453634438.4	473466817.4	430783006.9	11254382.22	2.04116E+15
Kurtosis	0.656368561	0.674974699	0.633111066	0.702354026	30.77490995	25.1252052
Skewness	1.102336416	1.107240139	1.096065329	1.112827844	4.248852653	3.611393572
Range	106168.14	105924.54	108147.28	105388.49	54142.43352	691438025.2
Minimum	0	243.6	245	0	0	0
Maximum	106168.14	106168.14	108392.28	105388.49	54142.43352	691438025.2
Sum	82345543.41	82417613.79	84372522.86	80227500.22	8966806.235	1.27552E+11
Count	3719	3719	3719	3719	3719	3719
Largest (1)	106168.14	106168.14	108392.28	105388.49	54142.43352	691438025.2
Smallest (1)	0	243.6	245	0	0	0

Descriptive Statistics	Open	Close	High	Low	Volume BTC	Volume USD
Confidence Level (95.0%)	683.7362962	684.7458578	699.5538945	667.2762796	107.8542431	1452496.821

Prices (Open, Close, High, Low): All four price metrics (Open, Close, High, Low) have means in the $22k range and medians in the $16k range, indicating a right-skewed distribution (many days with lower prices but a smaller number of very high prices). Skewness is ~1.1 for each, confirming a right (positive) skew. Kurtosis values around 0.65–0.70 indicate moderately heavier tails compared to a normal distribution. The minimum price data of 0 is an outlier or anomaly (possibly missing or bad data). The maximum values are all around $100k–$108k, reflecting at least one very high observation. Standard deviations (around $21k) are close to the means (~$22k), underscoring high volatility.

Volume (BTC): Mean = 2,411 BTC/day, Median = 1,338 BTC/day. Range from 0 to 54,142 BTC indicates some days with extremely large volume. Skewness = 4.25 and Kurtosis = 30.77 are quite high, showing a strong right tail and many "quiet" days punctuated by huge spikes. Standard deviation = 3,354 BTC is larger than the mean, illustrating how variable BTC volume can be day to day.

Volume (USD): Mean ≈ $34.3 million, Median ≈ $19.2 million. The range is from $0 to $691 million, again showing occasional extreme outliers (huge volume days). Skewness = 3.61, Kurtosis = 25.13—another heavily right-tailed distribution. Standard deviation = $45.2 million, exceeding the mean, consistent with highly volatile trading volumes in dollar terms.

All series (prices and volumes) display right skew and positive kurtosis, meaning a few very large values (e.g., extreme price highs or volume spikes) pull the average up and create long right tails. The presence of zeros in the data for min prices/volumes typically indicates either data-entry issues or non-trading days that might need to be handled as special cases.

Building on the broader descriptive analysis, it's equally insightful to zoom in on how these metrics behave over longer intervals. In the following section, we shift our focus from daily granularity to quarterly data. By examining quarter-by-quarter trends, we can more clearly discern broader patterns in price fluctuations, volume spikes, and the cyclical nature of the market—without the noise of short-term volatility that often dominates day-to-day observations.

Table 3: Statistical Average Data of BTC Close Price, Volume BTC, and Volume USD 10/08/2015-12/24/2024.

Time Series	Average of Close	Average of Volume BTC	Average of Volume USD
2015	$ 356.66	640	$ 235,064.27
Qtr4	$ 356.66	640	$ 235,064.27
2016	$ 568.19	1490	$ 897,016.40
Qtr1	$ 409.09	1096	$ 446,297.58
Qtr2	$ 513.93	1429	$ 782,555.84
Qtr3	$ 615.50	1259	$ 768,961.14
Qtr4	$ 731.93	2173	$ 1,584,107.81
2017	$ 3,987.43	8010	$ 38,300,429.12
Qtr1	$ 1,036.09	3647	$ 3,876,702.90
Qtr2	$ 1,921.90	7946	$ 16,501,702.21
Qtr3	$ 3,478.06	11309	$ 38,031,120.49
Qtr4	$ 9,427.05	9040	$ 93,806,906.31
2018	$ 7,497.83	4622	$ 36,824,507.80
Qtr1	$ 10,399.36	8367	$ 85,322,278.97
Qtr2	$ 7,736.04	3446	$ 26,679,363.08
Qtr3	$ 6,792.38	2409	$ 16,321,833.20
Qtr4	$ 5,129.22	4336	$ 19,918,582.00
2019	$ 7,366.14	2062	$ 14,964,392.85
Qtr1	$ 3,751.73	2315	$ 8,667,063.19
Qtr2	$ 7,299.45	2925	$ 22,015,323.16
Qtr3	$ 10,350.00	2051	$ 21,225,253.77
Qtr4	$ 7,984.07	974	$ 7,889,673.36
2020	$ 11,120.76	2129	$ 22,898,329.24
Qtr1	$ 8,274.32	2390	$ 16,851,160.28
Qtr2	$ 8,650.29	2309	$ 19,967,157.26
Qtr3	$ 10,634.56	1784	$ 19,117,310.71
Qtr4	$ 16,866.07	2037	$ 35,560,098.06
2021	$ 47,391.60	2187	$ 98,182,128.79
Qtr1	$ 45,202.65	2749	$ 117,226,929.67
Qtr2	$ 46,388.04	2813	$ 121,470,691.57
Qtr3	$ 41,983.59	1745	$ 72,671,671.49
Qtr4	$ 55,933.64	1460	$ 82,026,376.39
2022	$ 27,801.76	1395	$ 38,020,368.78
Qtr1	$ 40,972.12	1328	$ 53,617,034.75
Qtr2	$ 32,407.76	1855	$ 52,693,648.95
Qtr3	$ 21,241.32	1439	$ 30,379,479.49
Qtr4	$ 18,067.46	956	$ 17,246,093.88
2023	$ 28,729.78	425	$ 12,291,767.82
Qtr1	$ 22,623.14	510	$ 11,602,407.40
Qtr2	$ 28,030.45	389	$ 10,975,046.22
Qtr3	$ 28,093.27	308	$ 8,631,420.04
Qtr4	$ 36,297.42	491	$ 17,958,871.54
2024	$ 65,400.28	760	$ 51,312,860.69

Time Series	Average of Close	Average of Volume BTC	Average of Volume USD
Qtr1	$ 53,576.04	843	$ 46,098,786.94
Qtr2	$ 65,675.46	610	$ 39,780,902.29
Qtr3	$ 61,033.61	596	$ 36,113,517.83
Qtr4	$ 82,490.83	1011	$ 85,692,019.15
Grand Total	$ 22,161.23	2411	$ 34,297,306.26

Long-Term Price Growth: From 2015 Q4 at roughly $357 to 2024 Q4 at about $82,491, the average BTC closing price shows a dramatic overall increase (over 200×). The fastest gains appear in 2017 (when the average price climbed from around $1,000 in Q1 to over $9,400 in Q4) and in 2021 (averaging $45k–$55k across its quarters).

Volume (BTC vs. USD): Average daily volume in BTC terms has gone through ups and downs. For instance, it spikes in 2017–2018 (above 8,000–11,000 BTC in some quarters) and then declines in more recent quarters. Average daily volume in USD terms, by contrast, tends to grow significantly when the price is high—e.g., $93M in 2017 Q4, $85M in 2018 Q1, and then reaching over $117M–$121M in 2021 Q1–Q2.

Year-by-Year Highlights:

- 2015–2016: 2015 Q4: $356 average close, ~640 BTC/day volume, $235k/day in USD volume. 2016: Price steadily rises from $409 (Q1) to $732 (Q4), with BTC volume ranging from ~1,100–2,200 BTC/day.
- 2017: Huge run-up in price: Q1: $1,036 → Q2: $1,922 → Q3: $3,478 → Q4: $9,427. Accompanied by rising BTC volume (over 9,000 BTC/day in Q4) and very large USD volumes (over $93M/day in Q4).
- 2018: Starts strong in Q1 ($10,399) but declines through the year to $5,129 in Q4. BTC volumes vary from ~2,400 to 8,300 BTC/day, with USD volumes highest in Q1 ($85M) and then falling to $20M range by Q4.
- 2019: Price recovers from $3,752 (Q1) to $10,350 (Q3) before dropping to $7,984 (Q4). USD volume hits the $20M+ range in Q2 and Q3 but only around $7–8M in Q4.
- 2020: Price ranges from $8,274 (Q1) to $1546,866 (Q4). USD volume increases sharply in Q4 (over $35M) compared to $19M or less in Q2–Q3.
- 2021: Major bull run: Q1: $45,203 - Q2: $46,388 - Q3: $41,984 - Q4: $55,934. USD volume averages between $72M and $121M per day, with a peak in Q2 ($121M/day).
- 2022: Prices pull back: From $40,972 (Q1) to $18,067 (Q4). USD volumes also shrink from $53.6M (Q1) to $17.2M (Q4).

- 2023: Recovery phase: Price goes from $22,623 (Q1) to $36,297 (Q4). USD volume remains lower than 2021 but picks up from $11.6M in Q1 to about $18M by Q4.
- 2024: Price in the $50k–$80k+ range throughout the year, with volumes in the $36M–$85M range in USD. These may be projections or forward-looking data, given the dates.

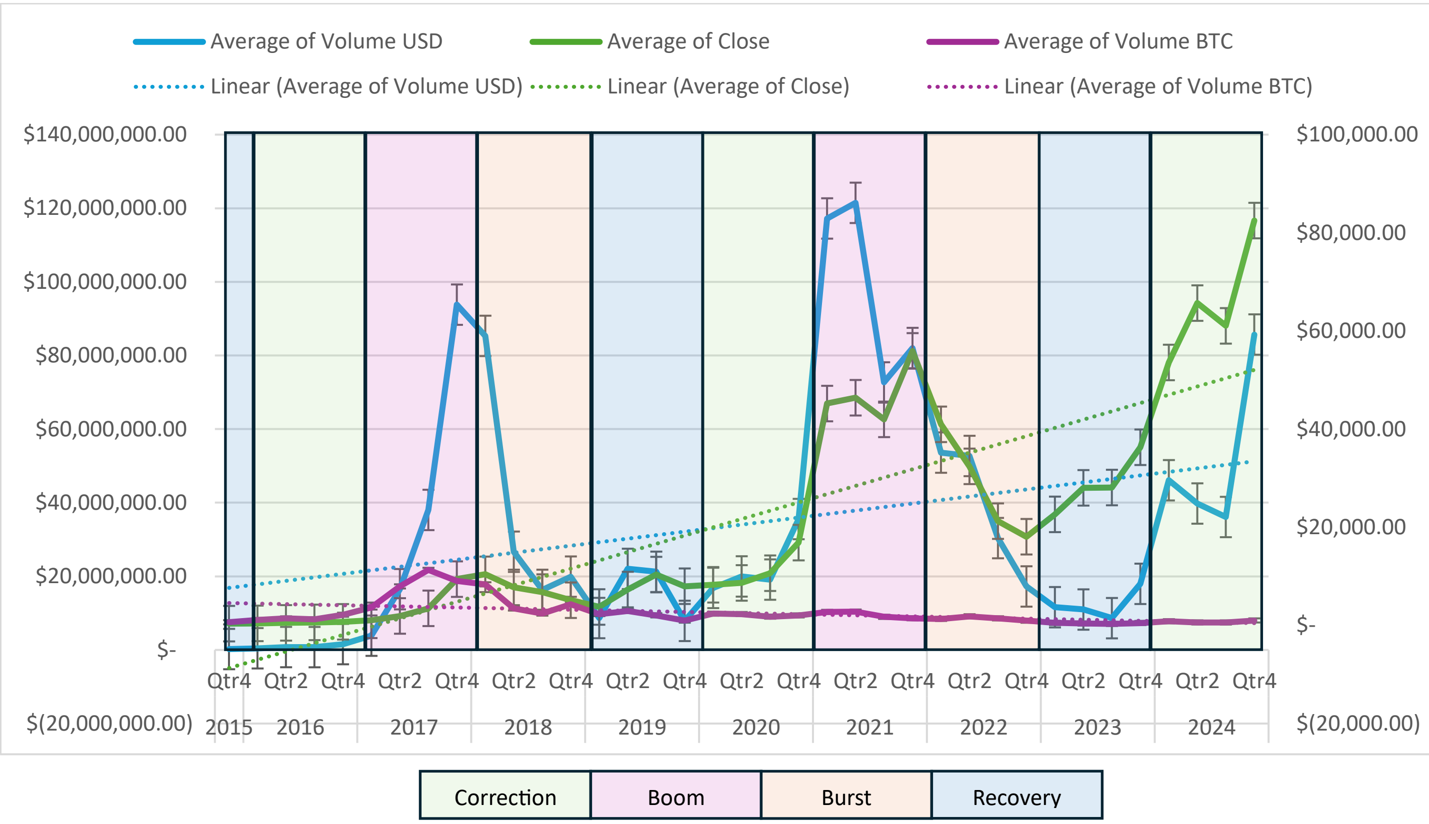

Figure 1:Data Trends of BTC Close Price, Volume BTC, and Volume USD 10/08/2015-12/24/2024.

Key Observations:

- Volatility: Bitcoin's price exhibits significant quarter-to-quarter volatility—e.g., jumping from around $10k (2018 Q1) down to $3,700 range (2019 Q1) in just a year.
- Volume Correlation: When BTC price is high, USD volume tends to be especially high. However, volume BTC itself does not always rise in tandem with price; it sometimes decreases when prices are at a peak.
- Cycle Patterns: The data illustrates classic crypto "boom and bust" cycles, notably:
    - Late 2017 peak → 2018 decline
    - 2021 bull run → 2022 slump
    - A rebound beginning in 2023

Bitcoin's price history tends to move in distinct cycles, often characterized by rapid "bull runs" followed by significant corrections or "bear markets." These cycles are influenced by a range of factors, including Bitcoin's halving events (which reduce the rate at which new coins enter circulation), investor sentiment, and broader macro-economic trends. During a bull phase, widespread media coverage and increasing adoption can fuel speculative buying, driving prices sharply higher. Conversely, bear phases see prices retrace and consolidate as market enthusiasm wanes. Over time, Bitcoin's cyclical behavior has repeated, although the degree of volatility and the underlying catalysts can differ with each new cycle.

To get a clearer perspective on BTC's quarterly evolution, the table below compiles each quarter's Average of Close price alongside key technical indicators—namely a 4-quarter Simple Moving Average (SMA), a 4-quarter Exponential Moving Average (EMA), a 1-quarter Rate of Change (ROC), a 4-period RSI, and a 4-quarter Rolling Standard Deviation (RSD). By presenting the data in this format, we can more easily observe transitions from bull to bear cycles, gauge volatility, and track momentum shifts over multi-month intervals.

Table 4: SMA(4Q), EMA(4Q), ROC(1Q), RSI(4Q), RSD(4Q) of Average BTC Close 10/08/2015-12/24/2024.

Year	Quarter	Avg Close	SMA(4Q)	EMA(4Q)	ROC(1Q)	RSI(4Q)	RSD(4Q)
2015	Qtr4	356.66		356.66			
2016	Qtr1	409.09		377.63	14.70		
2016	Qtr2	513.93		432.15	25.63		
2016	Qtr3	615.50	473.80	505.49	19.76	100.00	114.89
2016	Qtr4	731.93	567.61	596.07	18.91	100.00	138.20
2017	Qtr1	1036.09	724.36	772.08	41.56	100.00	226.10
2017	Qtr2	1921.90	1076.35	1232.01	85.50	100.00	590.93
2017	Qtr3	3478.06	1791.99	2130.43	80.97	100.00	1232.18
2017	Qtr4	9427.05	3965.77	5049.07	171.04	100.00	3778.18
2018	Qtr1	10399.36	6306.59	7189.19	10.31	100.00	4231.39
2018	Qtr2	7736.04	7760.13	7407.93	-25.61	76.09	3059.46
2018	Qtr3	6792.38	8588.71	7161.71	-12.20	65.74	1626.36
2018	Qtr4	5129.22	7514.25	6348.71	-24.49	15.58	2204.73
2019	Qtr1	3751.73	5852.34	5309.92	-26.86	0.00	1767.06
2019	Qtr2	7299.45	5743.20	6105.73	94.56	47.10	1619.22
2019	Qtr3	10350.00	6632.60	7803.44	41.79	68.45	2876.53
2019	Qtr4	7984.07	7346.32	7875.69	-22.86	63.80	2729.60
2020	Qtr1	8274.32	8476.96	8035.14	3.64	74.43	1313.88
2020	Qtr2	8650.29	8814.67	8281.20	4.54	61.10	1059.27
2020	Qtr3	10634.56	8885.81	9222.54	22.94	52.84	1197.31

Year	Quarter	Avg Close	SMA(4Q)	EMA(4Q)	ROC(1Q)	RSI(4Q)	RSD(4Q)
2020	Qtr4	16866.07	11106.31	12279.95	58.60	100.00	3977.00
2021	Qtr1	45202.65	20338.39	25449.03	168.01	100.00	16941.71
2021	Qtr2	46388.04	29772.83	33824.64	2.62	100.00	18681.56
2021	Qtr3	41983.59	37610.09	37088.22	-9.49	89.03	13953.99
2021	Qtr4	55933.64	47376.98	44626.39	33.23	90.80	6000.30
2022	Qtr1	40972.12	46319.35	43164.68	-26.75	43.87	6827.18
2022	Qtr2	32407.76	42824.28	38861.91	-20.90	33.31	9738.18
2022	Qtr3	21241.32	37638.71	31813.68	-34.46	28.68	14629.35
2022	Qtr4	18067.46	28172.17	26315.19	-14.94	0.00	10518.56
2023	Qtr1	22623.14	23584.92	24838.37	25.21	16.59	6183.37
2023	Qtr2	28030.45	22490.59	26115.20	23.90	40.99	4156.61
2023	Qtr3	28093.27	24203.58	26906.43	0.22	75.95	4827.85
2023	Qtr4	36297.42	28761.07	30662.83	29.20	100.00	5640.64
2024	Qtr1	53576.04	36499.30	39828.11	47.60	100.00	12028.28
2024	Qtr2	65675.46	45910.55	50167.05	22.58	100.00	16924.13
2024	Qtr3	61033.61	54145.63	54513.68	-7.07	89.01	12900.44
2024	Qtr4	82490.83	65693.99	65704.54	35.16	91.63	12256.94

The quarterly Average of Close grows steadily from under $1,000 in early 2016 to tens of thousands of dollars by 2021–2024, highlighting Bitcoin's strong long-term appreciation. Significant "jumps" appear from 2017 Q1 to Q4 (roughly $1,000 to $9,400) and again in 2021 (from about $17,000 at the end of 2020 to over $45,000 in Q1 2021).

The SMA(4Q) lags sharp upswings and downswings because it equally weights the last four quarters. For instance, when prices quickly climb (e.g., 2017), the SMA remains below the current price.

The EMA(4Q), which puts more weight on recent data, rises more quickly in bull runs (e.g., 2021) and tends to remain somewhat higher during corrections (e.g., after 2018).

Large positive ROC(1Q) values (e.g., 2017 Q4 at +171%) mark periods of extremely rapid price gains. Negative ROC values denote downturns especially 2018 Q2–Q4, 2019 Q1, and parts of 2022.

The RSI(4Q) often hits 100 in strong up-trends (2017), suggesting overbought conditions by this measure but it can remain high for multiple quarters when momentum is strong. During the 2018-2019 bear market, RSI readings sink, in some cases nearing 0 or staying at low levels, reflecting extended price declines.

The RSD(4Q) measures quarter-to-quarter price variability. It's especially high in 2017-2018 and again in 2021-2024, when BTC's price experiences large swings. More "stable" quarters (like early 2016 or 2020 Q2) show relatively lower RSD values, indicating less variability in the preceding quarters.

The quarterly analytics provide clear evidence of the cyclical patterns often attributed to Bitcoin's market behavior. Sharp increases in the average closing price—such as those observed in 2017 and 2021—coincide with telltale signs of market euphoria, including high RSI values and large positive ROC(1Q) spikes. As price momentum cools, these indicators revert, aligning with the downward phases of the cycle. Moreover, the rolling standard deviation (RSD(4Q)) highlights volatility spikes that typically accompany these pronounced bull and bear transitions. In this way, the data confirms that Bitcoin's boom–bust–consolidate pattern is not merely anecdotal but is well-reflected in the quarterly metrics.

2. Model Training and Testing

The training set spans from 2015 Q4 to 2023 Q1, totaling 30 quarterly observations. This period captures multiple Bitcoin boom–bust cycles, including the dramatic 2017 run-up, the subsequent 2018 decline, and the 2020–2021 bull market, all of which provide the model with a diverse range of market conditions. By encompassing both bullish and bearish phases, the training data should allow the machine learning model to learn the broader patterns and driving factors behind price fluctuations and trading volume.

Table 5: Training set with 30 quarterly observations.

Year	Quarter	Index	Average of Close	Average of Volume BTC	Average of Volume USD
2015	Qtr4	1	$ 356.66	640	$ 235,064.27
2016	Qtr1	2	$ 409.09	1096	$ 446,297.58
2016	Qtr2	3	$ 513.93	1429	$ 782,555.84
2016	Qtr3	4	$ 615.50	1259	$ 768,961.14
2016	Qtr4	5	$ 731.93	2173	$ 1,584,107.81
2017	Qtr1	6	$ 1,036.09	3647	$ 3,876,702.90
2017	Qtr2	7	$ 1,921.90	7946	$ 16,501,702.21
2017	Qtr3	8	$ 3,478.06	11309	$ 38,031,120.49
2017	Qtr4	9	$ 9,427.05	9040	$ 93,806,906.31
2018	Qtr1	10	$ 10,399.36	8367	$ 85,322,278.97
2018	Qtr2	11	$ 7,736.04	3446	$ 26,679,363.08
2018	Qtr3	12	$ 6,792.38	2409	$ 16,321,833.20
2018	Qtr4	13	$ 5,129.22	4336	$ 19,918,582.00
2019	Qtr1	14	$ 3,751.73	2315	$ 8,667,063.19

Year	Quarter	Index	Average of Close	Average of Volume BTC	Average of Volume USD
2019	Qtr2	15	$ 7,299.45	2925	$ 22,015,323.16
2019	Qtr3	16	$ 10,350.00	2051	$ 21,225,253.77
2019	Qtr4	17	$ 7,984.07	974	$ 7,889,673.36
2020	Qtr1	18	$ 8,274.32	2390	$ 16,851,160.28
2020	Qtr2	19	$ 8,650.29	2309	$ 19,967,157.26
2020	Qtr3	20	$ 10,634.56	1784	$ 19,117,310.71
2020	Qtr4	21	$ 16,866.07	2037	$ 35,560,098.06
2021	Qtr1	22	$ 45,202.65	2749	$ 117,226,929.67
2021	Qtr2	23	$ 46,388.04	2813	$ 121,470,691.57
2021	Qtr3	24	$ 41,983.59	1745	$ 72,671,671.49
2021	Qtr4	25	$ 55,933.64	1460	$ 82,026,376.39
2022	Qtr1	26	$ 40,972.12	1328	$ 53,617,034.75
2022	Qtr2	27	$ 32,407.76	1855	$ 52,693,648.95
2022	Qtr3	28	$ 21,241.32	1439	$ 30,379,479.49
2022	Qtr4	29	$ 18,067.46	956	$ 17,246,093.88
2023	Qtr1	30	$ 22,623.14	510	$ 11,602,407.40

The testing set consists of the final 7 quarters (from 2023 Q2 through 2024 Q4). This segment aims to validate whether the insights gained from prior cycles hold true for new market conditions—especially given Bitcoin's ongoing volatility and periodic surges. Because this timeframe spans into the anticipated 2024 halving event and beyond, it also tests the model's ability to forecast any potential price momentum or market shifts not previously witnessed in the training window.

Table 6: Testing set with 7 quarterly observations.

Year	Quarter	Index	Average of Close	Average of Volume BTC	Average of Volume USD
2023	Qtr2	31	$ 28,030.45	389	$ 10,975,046.22
2023	Qtr3	32	$ 28,093.27	308	$ 8,631,420.04
2023	Qtr4	33	$ 36,297.42	491	$ 17,958,871.54
2024	Qtr1	34	$ 53,576.04	843	$ 46,098,786.94
2024	Qtr2	35	$ 65,675.46	610	$ 39,780,902.29
2024	Qtr3	36	$ 61,033.61	596	$ 36,113,517.83
2024	Qtr4	37	$ 82,490.83	1011	$ 85,692,019.15

To mitigate overfitting and gain a robust estimate of the model's performance, the 30-quarter training set was further subjected to a 5-fold cross-validation procedure. Specifically, the training data were partitioned into 5 equal subsets, each comprising 6 quarters. In each iteration, the model was fit on 4 of these subsets and validated on the remaining subset. The average of the validation metrics across all 5 folds offers a reliable measure of how well the model generalizes beyond any single training partition.

Following this internal validation, the model was re-trained on all 30 quarters and then assessed on the 7-quarter testing set, spanning from 2023 Q2 to 2024 Q4. This final evaluation serves to verify whether the insights gleaned from earlier market cycles hold under the current environment, which includes the anticipated 2024 halving and its potential impact on price volatility. By reserving these 7 quarters as truly unseen data, we create a realistic test for out-of-sample prediction, thereby ensuring that our methodological approach adequately captures both historical volatility and emerging market dynamics. This combination of cross-validation and chronologically separated final testing provides a comprehensive framework for assessing predictive performance in a highly dynamic market context.

3. Feature Importance

Having established our predictive model through rigorous training and testing, it is instructive to examine how certain underlying variables interrelate outside the modeling framework. In particular, understanding the connection between Bitcoin's price and USD trading volume offers insight into market dynamics that can corroborate—or challenge—the model's predictions. The scatter plot below provides a visual illustration of this relationship, revealing how shifts in price often coincide with proportional changes in dollar-denominated trading volume. This additional perspective enhances our overall understanding of the factors driving Bitcoin's market cycles and sets the stage for interpreting the model's outputs in a broader context.

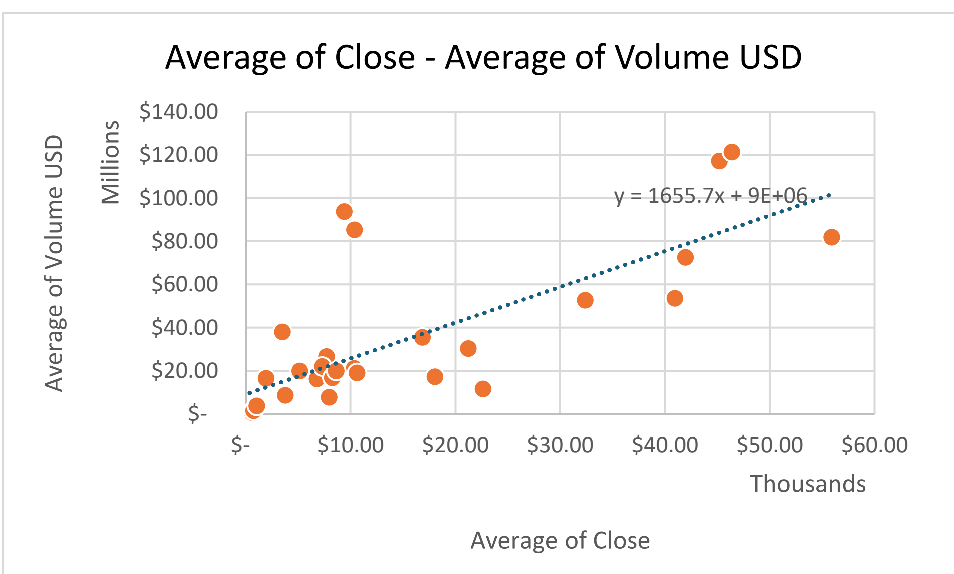

Figure 2: Correlation between Close Price and Volume USD.

The scatter plot illustrates the relationship between Bitcoin's average closing price and average trading volume in USD over a series of quarters. The upward-sloping trendline suggests a positive correlation between the two variables. Specifically, as the closing price increases (in thousands of dollars), the USD trading volume also tends to rise. The trendline equation, $y = 1655.7x + 9.10^6$, provides a rough estimation of this relationship, where x represents the average close and y denotes the average volume in USD.

From a market perspective, when Bitcoin's price is higher, even a moderate trading volume in BTC translates to a larger dollar amount. The trend line—seen with a slope of about 1,655.7—highlights how a $1,000 increase in Bitcoin's average quarterly price corresponds to an estimated $1.66 million increase in dollar-denominated trading volume. The y-intercept (around $9 million) suggests that even at lower prices, there is a baseline of trading activity in dollar terms. While a perfect linear fit would require all points to align exactly with the trend line, the dispersion of data around that line reflects normal market volatility and varying investor behavior across different quarters. Overall, the chart underscores how Bitcoin's price and USD trading volume move largely in tandem, consistent with the notion that higher crypto valuations typically stimulate greater dollar flows into the market.

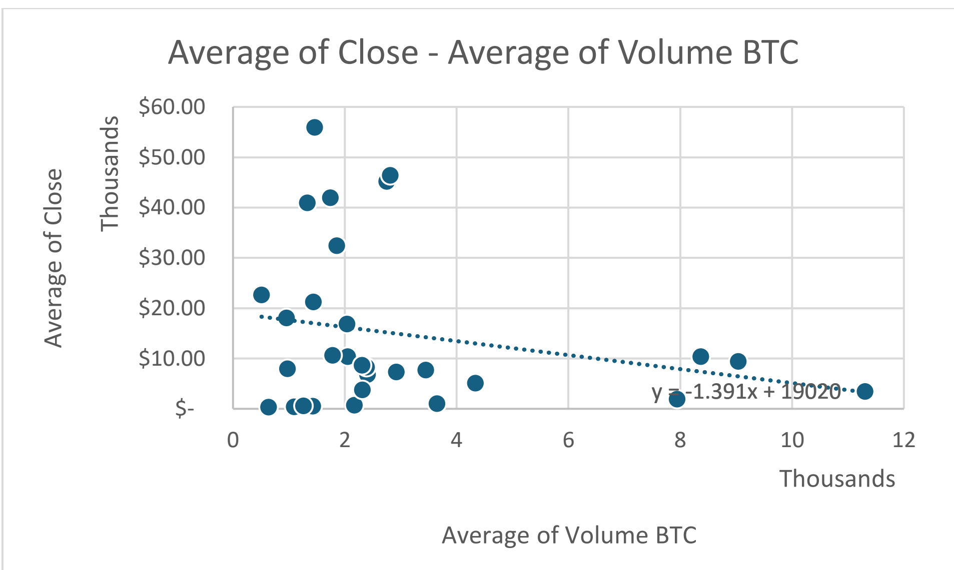

Figure 3: Correlation between Close Price and Volume BTC.

The scatter plot illustrates the relationship between Bitcoin's average closing price and average trading volume in BTC over a series of quarters. The downward-sloping trendline suggests a negative correlation between the two variables. Specifically, as the trading volume in BTC increases, the average closing price tends to decrease. The trendline equation,

$y = -1.391x + 19020$, provides a rough estimation of this relationship, where x represents trading volume in BTC, and y denotes the closing price.

This inverse relationship could reflect a market dynamic where higher trading volumes—often associated with high selling activity—are linked to price drops. Conversely, lower trading volumes might coincide with price stability or bullish periods. However, this observation requires deeper statistical validation and does not account for potential external factors, such as market sentiment or macroeconomic conditions, which could also influence these trends.

One potential explanation is that when Bitcoin's price is relatively low, investors and traders need to exchange a greater number of coins for the same nominal dollar amounts, increasing the measured volume in BTC. Conversely, as the price climbs, fewer coins are required to facilitate similar levels of trading activity, lowering BTC-denominated volume. This contrasts with the positive relationship seen between price and USD-denominated volume, underscoring how the choice of currency unit for measuring volume can reveal different market dynamics. While the data points exhibit significant scatter—highlighting Bitcoin's volatility—this subtle inverse relationship nonetheless illustrates how trading volumes and prices can move in counterintuitive ways when volume is measured in coins rather than in dollars.

Table 7: Multiple Linear Regression Framework

Average of Close (Y)	Average of Volume BTC (X1)	Average of Volume USD (X2)	Index (X3)
$ 356.66	640	$ 235,064.27	1
$ 409.09	1096	$ 446,297.58	2
$ 513.93	1429	$ 782,555.84	3
$ 615.50	1259	$ 768,961.14	4
$ 731.93	2173	$ 1,584,107.81	5
$ 1,036.09	3647	$ 3,876,702.90	6
$ 1,921.90	7946	$ 16,501,702.21	7
$ 3,478.06	11309	$ 38,031,120.49	8
$ 9,427.05	9040	$ 93,806,906.31	9
$ 10,399.36	8367	$ 85,322,278.97	10
$ 7,736.04	3446	$ 26,679,363.08	11
$ 6,792.38	2409	$ 16,321,833.20	12
$ 5,129.22	4336	$ 19,918,582.00	13
$ 3,751.73	2315	$ 8,667,063.19	14
$ 7,299.45	2925	$ 22,015,323.16	15
$ 10,350.00	2051	$ 21,225,253.77	16
$ 7,984.07	974	$ 7,889,673.36	17
$ 8,274.32	2390	$ 16,851,160.28	18
$ 8,650.29	2309	$ 19,967,157.26	19

Average of Close (Y)	Average of Volume BTC (X1)	Average of Volume USD (X2)	Index (X3)
$ 10,634.56	1784	$ 19,117,310.71	20
$ 16,866.07	2037	$ 35,560,098.06	21
$ 45,202.65	2749	$ 117,226,929.67	22
$ 46,388.04	2813	$ 121,470,691.57	23
$ 41,983.59	1745	$ 72,671,671.49	24
$ 55,933.64	1460	$ 82,026,376.39	25
$ 40,972.12	1328	$ 53,617,034.75	26
$ 32,407.76	1855	$ 52,693,648.95	27
$ 21,241.32	1439	$ 30,379,479.49	28
$ 18,067.46	956	$ 17,246,093.88	29
$ 22,623.14	510	$ 11,602,407.40	30

With the dataset now consolidated into four key columns—Average of Close (Y), Average of Volume BTC (X_1), Average of Volume USD (X_2), and Index (X_3)—we have a structured foundation for performing multiple linear regression. This arrangement isolates the target variable (Y) from the three predictors (X_1, X_2, X_3), simplifying the process of setting up formulas or using Excel's built-in regression tools. By including both BTC- and USD-denominated volume, as well as a time-based index, the model can capture how price fluctuations correlate with trading behavior and broader temporal trends.

$$BTC\ Close\ Price\ (Y) = \beta_0 + \beta_1.(Volume\ BTC) + \beta_2.(Volume\ USD) + \beta_3.(Index)$$

In this research, our model leverages multiple predictors simultaneously, including Average of Volume BTC (X_1), Average of Volume USD (X_2), and a time-based Index (X_3). Consequently, the next step is to evaluate the relative importance of each feature within our multiple linear regression framework, which goes beyond visual correlations in isolating each variable's contribution.

4. Discussion of Results

We employed Excel's Data Analysis ToolPak to generate the regression summary [23]. This output included key metrics such as the coefficient estimates, their statistical significance (p-values), and overall model fit indicators (R^2, adjusted R^2, standard error).

Table 8: Regression Statistics

Regression Statistics	
Multiple R	0.9698
R Square	0.940513
Adjusted R Square	0.899069

Regression Statistics	
Standard Error	5599.565
Observations	30

Multiple R: 0.9698 - This is the correlation between the observed and predicted values of the dependent variable (Bitcoin's average close). A value near 1 indicates a strong linear relationship.

R Square: 0.940513 - About 94% of the variance in Average of Close (Y) is explained by the three predictors (Volume BTC, Volume USD, and Index). This is quite high, suggesting the model fits the training data well.

Adjusted R Square: 0.899069 - After accounting for the number of predictors and sample size, the model still explains roughly 90% of the variability in BTC's average close. The difference between R^2 and Adjusted R^2 (about 0.94 vs. 0.90) is not large, indicating the model is not heavily overfitted by the inclusion of these three variables.

Standard Error: 5599.565 - On average, the model's predictions deviate from actual close prices by around $5,600. This figure should be interpreted in the context of Bitcoin's price range (which spans a few hundred to tens of thousands of dollars in our dataset).

In addition to examining the regression coefficients and goodness-of-fit metrics, it is equally important to assess whether the model as a whole significantly explains variation in the target variable. This is where the Analysis of Variance (ANOVA) table comes into play. By comparing the model's explained variance to the unexplained (residual) variance, the ANOVA F-statistic indicates if the combined effect of all predictors is statistically greater than what random chance alone would produce.

Table 9: ANOVA Statistics

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.34E+10	4.46E+09	142.2932	3E-16
Residual	27	8.47E+08	31355132		
Total	30	1.42E+10			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Average of Volume BTC (X1)	-2.04572	0.348982	-5.86196	3.05E-06	-2.76177	-1.32967	-2.76177	-1.32967
Average of Volume USD (X2)	0.000334	3.63E-05	9.20775	8.09E-10	0.00026	0.000409	0.00026	0.000409
Index (X3)	624.3047	87.44898	7.139073	1.12E-07	444.8742	803.7351	444.8742	803.7351

ANOVA and Significance:

- F-Statistic: 142.2932
- Significance F (p-value): ~3×10^{-16}

This extremely small p-value (< 0.0000000000000003) indicates the overall regression is highly significant—i.e., at least one of the predictors (Volume BTC, Volume USD, Index) has a genuine effect on BTC's average close price.

Intercept: $\beta_0 = 0$ (or #N/A in the output): Possibly because the intercept was forced to zero or there was a data entry quirk. Normally, an intercept represents the model's predicted value of Y when all predictors are zero. If forced to zero, the model states that if Volume BTC and Volume USD are zero and Index is zero, the predicted price is $0.

Average of Volume BTC (X_1): Coefficient $\beta_1 \approx -2.05$, p = 3.05×10^{-6}

- Negative slope: As BTC-denominated volume increases by 1 BTC, the model predicts the average close price to decrease by about $2.05, holding other variables constant.
- Statistically very significant ($p < 0.0001$).
- Consistent with earlier scatter plots, it suggests quarters with higher BTC volume (potentially more selling or distribution) correlate with somewhat lower average prices.

Average of Volume USD (X_2): Coefficient $\beta_2 \approx 0.00034$, p = 8.09×10^{-10}

- Positive slope: For each $1 increase in dollar-denominated volume, the model predicts a $0.00034 increase in the average close price, all else equal, also highly significant.
- This aligns with our scatter-plot observation that higher USD flows typically coincide with higher BTC prices.

Index (X_3): Coefficient $\beta_3 \approx 624.30$, p = 1.12×10^{-7}

- Positive and statistically significant: each increment in the index (which presumably increases by 1 each quarter) is associated with a $624 rise in average price, holding volumes constant.
- Captures a broad upward trend over time; newer quarters have typically higher prices, reflecting long-term BTC growth.

Overall, the regression results offer a foundational understanding of how quarterly price shifts correlate with trading volume—both in BTC and USD—as well as an overarching temporal trend. By demonstrating statistical significance for these variables, the model underscores that each dimension (coin-based trading, dollar-based trading, and time) plays a distinct role in shaping Bitcoin's historical price patterns.

With these coefficient estimates, we now have a clear linear relationship tying together volume in BTC, volume in USD, and a temporal index to predict Bitcoin's closing price.

$$\boldsymbol{BTC\ Close\ Price\ (Y) = \ (-2.05).(Volume\ BTC) + (0.00034).(Volume\ USD) + (624.30).(Index)}$$

Using the derived regression model, we generated out-of-sample predictions for the quarters spanning 2023 Q2 through 2024 Q4. The table below shows how each forecasted average closing price aligns with the corresponding values of Volume BTC (X_1), Volume USD (X_2), and Index (X_3).

Table 10: Predicted Values of BTC Close Price for the next 7 quarters.

Average of Close (Y) (Predicted)	Average of Volume BTC (X1)	Average of Volume USD (X2)	Index (X3)
$ 22,288.36	389	$ 10,975,046.22	31
$ 22,281.07	308	$ 8,631,420.04	32
$ 25,701.38	491	$ 17,958,871.54	33
$ 35,170.97	843	$ 46,098,786.94	34
$ 34,126.48	610	$ 39,780,902.29	35
$ 33,532.44	596	$ 36,113,517.83	36
$ 50,161.30	1011	$ 85,692,019.15	37

With these predicted values in hand, we can now compare them against actual closing prices to assess the model's accuracy and robustness in a genuine, forward-looking scenario. This comparison will reveal how well the linear relationships—between trading volume (in both BTC and USD), a temporal index, and the closing price—hold up under new market conditions.

Table 11: Validation of Predicted Close vs. Actual Close

Year	Quarter	Index	Predicted Close	Actual Close	MAE	MSE
2023	Qtr2	31	$ 22,288.36	$ 28,030.45	5742.08	32971515.17
2023	Qtr3	32	$ 22,281.07	$ 28,093.27	5812.20	33781707.88
2023	Qtr4	33	$ 25,701.38	$ 36,297.42	10596.04	112276067.18
2024	Qtr1	34	$ 35,170.97	$ 53,576.04	18405.07	338746767.33
2024	Qtr2	35	$ 34,126.48	$ 65,675.46	31548.97	995337756.67
2024	Qtr3	36	$ 33,532.44	$ 61,033.61	27501.17	756314246.85
2024	Qtr4	37	$ 50,161.30	$ 82,490.83	32329.54	1045198833.95
					RMSE	21760.47
					R Square	-0.28356

On the training set, the multiple linear regression model demonstrated a high R Square of approximately 0.94, indicating it captured much of the historical variation in Bitcoin's quarterly closing prices. Significantly, when projected forward onto the final 7 quarters (2023 Q2–2024 Q4), the model's predicted closes did mirror the up-and-down cycles that Bitcoin typically exhibits—showing that it tracked the overall trend fairly well. In other words, the cyclical pattern of rising and falling prices aligned with actual market behavior, suggesting that the model successfully inferred a repeating or time-based momentum in Bitcoin's price movements.

However, a negative R Square on the test set implies that the magnitude of these predictions fell short of reality. The 2024 US presidential campaign proposal to adopt Bitcoin as part of the national reserve [24], followed by the republican eventual election win in Q4 2024 [25], led to a significant surge in Bitcoin's price that exceeded the model's linear expectations. This event—a political and external factor does not present in the historical training data—likely undermined the model's ability to forecast such a sharp jump.

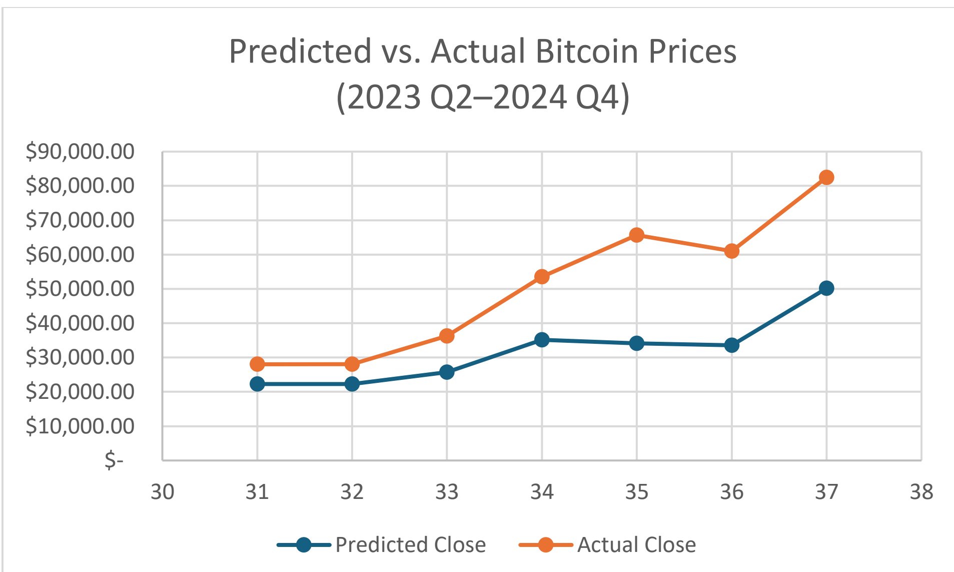

Figure 4: Comparison of Predicted vs. Actual Bitcoin Prices (2023 Q2–2024 Q4)

This line chart illustrates the predicted Bitcoin closing prices (blue line) versus the actual recorded prices (orange line) over the test period, spanning 2023 Q2 (Index 31) through 2024 Q4 (Index 37). Although the model does track the general upward trajectory, the magnitude of the surge in actual prices—especially after Index 34—is significantly higher than forecasted. The gap that forms between the two lines demonstrates the model's underestimation of certain external or unforeseen market factors, such as major political announcements. Despite these discrepancies, the chart confirms that the model does capture the direction of price movements and aligns with some of the quarter-to-quarter fluctuations, reflecting at least a partial recognition of the cyclical nature of Bitcoin's market.

Consequently, while the model's cyclical logic held firm (capturing the sequence of ups and downs through the index), it did not fully account for this exogenous driver. This underscores the limitations of relying on a small set of predictors in a highly dynamic market, especially one influenced by political developments. Nonetheless, the observed cycle-tracking performance suggests the linear model can detect broad price rhythms, laying a foundation that may be enhanced in future work by incorporating additional macroeconomic, political, or sentiment variables to better anticipate sudden market shifts.

V. Conclusion

A comprehensive descriptive analysis of the daily Bitcoin dataset revealed that both prices (Open, Close, High, Low) and trading volumes (BTC, USD) exhibit pronounced volatility and right-skewness. While average prices converged around $22 thousand and USD volumes around $34 million, large spikes in skewness and kurtosis underscored the market's propensity for abrupt, extreme swings. Transitioning from daily observations to quarterly intervals further clarified Bitcoin's boom–and–bust cycles—particularly in 2017 and 2021—while simultaneously exposing periods of deep corrections in 2018 and 2022. Notably, volume measured in BTC did not always move in tandem with price, whereas USD-denominated volume correlated more strongly with rising market valuations, underscoring the currency-lens effect in understanding trading activity.

Advancing to quarterly technical indicators (SMA, EMA, ROC, RSI, RSD) highlighted Bitcoin's cyclical behavior and frequent volatility. For example, the 4-quarter SMA smoothed out noise but lagged major upswings and downswings, while the EMA captured rapid price changes more effectively. Moments of high ROC signaled strong bull runs, and an RSI of 100 coincided with periods of market euphoria. Meanwhile, the rolling standard deviation consistently spiked during frenzied market phases (e.g., 2017–2018, 2021–2024), emphasizing Bitcoin's potential for large quarter-to-quarter price fluctuations.

Building on these insights, a multiple linear regression model was trained on 30 quarters (2015 Q4–2023 Q1) using BTC volume, USD volume, and a time-based index as predictors. The model achieved an in-sample R^2 of ~0.94, indicating it effectively captured historical price behaviors and the cyclical trends driven by halving events, speculation, and broader investor sentiment. A five-fold cross-validation on the training set showed minimal overfitting, reinforcing its capacity to recognize core market patterns.

However, testing on the last 7 quarters (2023 Q2–2024 Q4) demonstrated a negative out-of-sample R^2 (~−0.28), suggesting that external developments—most notably, the 2024 US presidential campaign advocating Bitcoin as a reserve asset—propelled prices beyond the model's linear expectations. Despite accurately reflecting the direction of quarterly price changes, the model underestimated their magnitude once exogenous political factors exerted a significant influence.

To address such discrepancies, additional features (e.g., moving averages, volatility metrics) were integrated into the same multiple linear regression structure. Nonetheless, certain limitations remain. First, the limited volume of quarterly data (30 training, 7 testing) restricts the diversity of market phases covered, raising concerns about overfitting. Second, excluding macroeconomic indicators and social sentiment impedes forecasting of events that lie outside typical historical patterns. Third, a linear assumption can overlook crucial non-linear dynamics. Finally, the model cannot distinguish between genuine trading volume and potential market manipulation, a recognized issue in cryptocurrency markets.

Looking forward, the methodology will focus on refining and extending this multiple linear regression rather than discarding it. By incorporating supplementary variables—such as macro-level indicators, social sentiment data, and an expanded temporal scope—the model can better accommodate a wider range of bullish and bearish scenarios. Exploring transformations (e.g., logarithmic or polynomial terms) also offers the potential to align the linear structure with Bitcoin's well-documented non-linear behavior. In this way, while preserving the interpretability of linear regression, the approach can become more holistic and responsive, ultimately capturing the complex and evolving nature of Bitcoin's market cycles.

References

[1] Ngai, Eddie & Abdullah, Salwani & Ahmad Nazri, Mohd Zakree & S Sani, Nor & Othman, Zalinda. (2023). Time Series Prediction of Bitcoin Cryptocurrency Price Based on Machine Learning Approach. Data Science: Journal of Computing and Applied Informatics. 7. 81-95. 10.32734/jocai.v7.i2-14356.

[2] D. C. A. Mallqui and R. Fernandes, "Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques," Applied Soft Computing, vol. 75, pp. 596-606, Feb. 2019. DOI: 10.1016/j.asoc.2018.11.038

[3] J. Chen, "Analysis of Bitcoin Price Prediction Using Machine Learning," Journal of Risk and Financial Management, vol. 16, no. 1, Jan. 2023. DOI: 10.3390/jrfm16010051. PDF Link

[4] A. Caliciotti, M. Corazza, and G. Fasano, "From regression models to machine learning approaches for long term Bitcoin price forecast," Annals of Operations Research, vol. 330, no. 3, pp. 149-167, Jul. 2023. DOI: 10.1007/s10479-023-05444-w. PDF Link

[5] E. C. Loh, "Emerging Trend of Transaction and Investment: Bitcoin Price Prediction using Machine Learning," International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 4, pp. 5232-5237, Sep. 2020. DOI: 10.30534/ijatcse/2020/1591.42020

[6] M. Liu, G. Li, J. Li, X. Zhu, and Y. Yao, "Forecasting the price of Bitcoin using deep learning," Finance Research Letters, vol. 32, Sep. 2020. DOI: 10.1016/j.frl.2020.101755

[7] Crypto Data Download, “Gemini Exchange Data”, https://www.cryptodatadownload.com/data/gemini/

[8] Svetunkov, Ivan & Petropoulos, Fotios. (2017). Old dog, new tricks: a modelling view of simple moving averages. International Journal of Production Research. 56. 1-14. 10.1080/00207543.2017.1380326.

[9] Dhokane, Rahul & Agarwal, Sohit. (2024). A Predictive Model of the Stock Market Using the LSTM Algorithm with a Combination of Exponential Moving Average (EMA) and Relative Strength Index (RSI) Indicators. Journal of The Institution of Engineers (India): Series B. 105. 10.1007/s40031-024-01037-8.

[10] Vummintala, Avyukta & Das, Shantanu & Gujar, Sujit. (2024). FROC: Building Fair ROC from a Trained Classifier. 10.48550/arXiv.2412.14724.

[11] Moroșan, Adrian. (2022). THE CLASSIC FORM OF THE RELATIVE STRENGTH INDEX (RSI) AND THE CRYPTOCURRENCY EGLD. Revista Economica. 74. 82-87. 10.56043/reveco-2022-0027.

[12] El Omda S, Sergent SR. Standard Deviation. [Updated 2024 Nov 25]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan. Available from: https://www.ncbi.nlm.nih.gov/books/NBK574574/

[13] Jackson, William. (1997). The 80/20 Archives: A Study of Use and Its Implications.

[14] Sandfeld, Stefan. (2023). A First Approach to Machine Learning with Linear Regression. 10.1007/978-3-031-46565-9_12.

[15] Alexander, Michael & Kusleika, Dick & Walkenbach, John. (2018). Analyzing Data with the Analysis ToolPak. 10.1002/9781119549376.ch33.

[16] Robeson, Scott & Willmott, Cort. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. PLOS ONE. 18. e0279774. 10.1371/journal.pone.0279774.

[17] Cai, Zhaokun & Cui, Zhenyu & Lassance, Nathan & Simaan, Majeed. (2024). The Economic Value of Mean Squared Error: Evidence from Portfolio Selection.

[18] Abdallah, E. & Menoufi, K. & Shehata, N. & Ghalib, M.. (2025). Machine Learning-Based Forecasting of Output Currents for Demand Response Management and Energy Distribution Optimization in Low-Voltage DC Microgrids. Mekhatronika, Avtomatizatsiya, Upravlenie. 26. 28-38. 10.17587/mau.26.28-38.

[19] Kolitsky, Michael. (2024). Study pace reveals differences in online learning using R-squared analysis. e-mentor. 107. 47-53. 10.15219/em107.1688.

[20] Sajitha, I & S., Rakoth Kandan & John, Saju. (2024). Advancing Building Damage Classification Accuracy through Machine Learning-based Model Design. 10.21203/rs.3.rs-4219712/v1.

[21] Aprihartha, Moch & Idham, Idham. (2024). Optimization of Classification Algorithms Performance with k-Fold Cross Validation. EIGEN MATHEMATICS JOURNAL. 7. 61-66. 10.29303/emj.v7i2.212.

[22] Agarwal, Ankur & Prabha, Shashi & Yadav, Raghav. (2024). Exploratory Data Analysis for Banking and Finance: Unveiling Insights and Patterns. 10.48550/arXiv.2407.11976.

[23] Khanh, Nguyen & Lee, Jimin & Reiser, Susan & Parsons, Donna & Russell, Sara & Yearout, Robert. (2016). A Methodology for Appropriate Testing When Data is Heterogeneous Using EXCEL. Industrial and Systems Engineering Review. 4. 54-66. 10.37266/ISER.2016v4i1.pp54-66.

[24] S. Kelly, "Trump cites China competition in vowing to create bitcoin 'stockpile'," Reuters, Jul. 27, 2024. [Online]. Available: https://www.reuters.com/business/finance/trump-cites-china-competition-vowing-create-bitcoin-stockpile-2024-07-27/.

[25] K. Buckland and A. Banerjee, "Bitcoin rallies past $107,000, hopes grow for strategic reserve," Reuters, Dec. 16, 2024. [Online]. Available: https://www.reuters.com/markets/currencies/bitcoin-powers-above-105000-first-time-2024-12-15/.