

Analyzing the Dynamics of Cryptocurrency Rate of Return through Social Media for Financial Trends

¹ Shahida Razzaq, ² Fiaz Majeed

¹ University of Gujrat, Pakistan

² University of Gujrat, Pakistan

Corresponding Author: shahida.razzaq36@gmail.com

ABBREVIATIONS		
S#	Item	Abbreviation
1	Machine Learning	ML
2	Time Series	TS
3	Long-Short term memory	LSTM
4	Bidirectional Long-Short term memory	Bi-LSTM
5	Autoregressive integrated moving average	ARIMA
6	Seasonal Autoregressive Integrated Moving Average	SARIMA
7	Natural Language Processing	NLP
8	Root Mean Square Error	RMSE
9	Mean Absolute Percentage Error	MAPE
10	Gated Recurrent Unit	GRU
11	Value at Risk	VaR
12	Rate of Returns	RoR
13	Sentiment Analysis	SA

ABSTRACT In recent times, a significant increase in research focus and the consideration of cryptocurrencies as an asset by institutional investors. Due to its volatility and influence on numerous outside variables, it has become a challenge to estimate its values. The proposed work investigates the impact of social media sentiment on the fluctuations related to cryptocurrency market, specifically emphasis on Bitcoin, Ether, and Binance, using data from Twitter and Telegram between July 19, 2023, and July 19, 2024. Firstly, an analysis is conducted using the covariance and correlation between the cryptocurrency return. After analysis, time series models Autoregressive integrated moving average and Seasonal Autoregressive Integrated Moving Average (ARIMA, SARIMA), and machine learning models (LSTM and Bi-LSTM) are employed to analyze and predict cryptocurrency return rate using the data. Finally, Hazards associated with trading in every cryptocurrency are examined with the help of value potential data. The results indicate that Twitter, due to its concise and real-time characteristics, provides more immediate insights into price changes, making it particularly advantageous for short-term trading strategies. Conversely, Telegram encompasses a wider range

of market sentiment, reflecting varied discussions that are beneficial for trend analysis and long-term forecasting, albeit with a higher level of noise. Both LSTM and Bi-LSTM models exhibited robust predictive capabilities, although the outcomes differed among various cryptocurrencies and platforms. Sentiment data from Twitter was found to correlate more closely with short-term price fluctuations, while Telegram's data offered a more holistic perspective on market trends. The study concludes that combining sentiment insights from both platforms can improve predictive accuracy, utilizing Twitter's immediacy alongside Telegram's comprehensive insights, thus facilitating more informed trading and investment choices in the dynamic cryptocurrency market.

INDEX TERMS Cryptocurrency, Machine Learning, Sentiment Analysis, RoR, prediction, Tweeter, Telegram

I. INTRODUCTION

Cryptocurrencies are extremely volatile assets that are a topic of much discussion, in both academic and media circles. An unknown developer "Satoshi Nakamoto," the creator of the original cryptocurrency, issued Bitcoin during the 2008 financial crisis. 14 years later, the price of bitcoin is close to \$21,000 (finance.yahoo.com), having started trading at \$0.0008 (investingnews.com). There are thousands of cryptocurrencies, and the industry grows to an incredible \$1.025 trillion by July 2022([Investopedia.com](https://investopedia.com)). Although Ether, which allows users to create Decentralized Apps and Smart Contracts, is one of the numerous cryptocurrencies that have been created, Bitcoin remains the most well-known cryptocurrency with the largest value (investopedia.com).

Due to the increased interest in cryptocurrencies, numerous research has been conducted in order to predict cryptocurrency prices and determine the key variables that influence these fluctuations [1]. According to simplelearn.com, the quantity of tokens that are created through the "mining" process, validating transactions and the number of tokens that already exist determine a cryptocurrency's supply. A cryptocurrency's supply is based on the number of tokens that have previously been created and those that are created through transaction validation [2].

In comparison to other currencies or gold, cryptocurrencies are incredibly volatile; even the well-known cryptocurrency Bitcoin is around five times more volatile (buybitcoinworldwide.com). An elevated level of asset volatility means higher risks associated with an investment, but it also may offer large potential returns. For instance, if someone bought Bitcoin currency in the month November 2021 and sold it after a year, they would have doubled their investment. Nevertheless, Bitcoin has dropped 70% of its value from its peak in November 2022 (finance.yahoo.com). Numerous studies have attempted to forecast the worth of cryptocurrencies and identify to the key elements influencing their price variations due to their high volatility and expanding popularity. Cryptocurrency-related social media activity is a topic that has been extensively researched [3]. Demonstrate evidence of how microblogging sites affect the stock market's liquidity. Research can be used to determine how private and public opinions affect prices. Now, Cryptocurrency-related social media activity is an extensively researched topic. Jacques Vella [4] Examined how well many neural models performed in forecasting cryptocurrencies price fluctuations using data from Twitter tweets. A social media site called Reddit has been receiving a lot of attention lately because of its impact on cryptocurrency developments [5]. Annotated dataset including sentiment and emotions for predicting the price of bitcoin using reddit posts and [6] study examines the growth of cryptocurrency content on Wikipedia through an analysis of co-edited pages and editorial activity. It finds that, over the majority of the covered period, a straightforward trading strategy based on Wikipedia viewpoints outperforms traditional standard approach in terms of return on investment.

In this study, Standard TS and two ML models with algorithms are used to analysis the relationship between cryptocurrency rate of return and social media sentiment (Telegram and Twitter), focusing on Bitcoin, Ether, and Binance within a specified date range from July 19, 2023, to July 19, 2024. The research computes the rate of return for every cryptocurrency after preprocessing the data, then integrates this with sentiment data to assess correlations. Additionally, (VaR) Values at Risk calculations are incorporated into the analysis to quantify possible losses at different confidence levels. The study hopes to offer practical insights into the relationship between investor sentiment and return on cryptocurrency by using this methodology.

A. Problem Statement

According to preparatory analysis demonstrate that long-run cryptocurrency returns may be pretentious by elements which receive attention on social media along with other trends of the financial industry. However, further research is required to completely comprehend this relationship. Therefore, this analysis intent to research more deeply the impact of attention on social media and other variables associated with the financial market in cryptocurrency returns. By comprehensively examining these ingredients, this investigation aims to contribute valuable observation into the dynamics of the digital market and contribute to the previous knowledge in this field.

B. Research Objectives

The following are the major objectives of the research that are being discussed:

- To investigate every impact of social media awareness and other financial market developments on the long-term return on digital currency investments.
- To understand and explore the relationship between social media attention and cryptocurrency return.
- To evaluate and forecast the relationship between the rate of return of major cryptocurrencies and numerous factors such as Telegram and Twitter sentiment and other financial metrics.
- To enhance a deeper understanding of the dynamic within the market of digital currencies by analyzing the influence of social media and financial market movements.
- To discover information that may be useful for individuals and organizations involved in cryptocurrency investments by identifying potential factors influencing long-term returns.
- To provide information regarding the returns relation between social media and choosing three cryptocurrencies in order to further the field of study on cryptocurrencies.

C. Research Question

- How does social media attention impact the long-term returns of cryptocurrencies or what specific aspects of social media influence these returns?
- To what extent do social media attentions for financial market variables jointly contribute to the fluctuations in cryptocurrency returns?
- What is the relationship between the emotion on social media (Twitter and Telegram) and the rate of return for Binance, Ether, and Bitcoin?
- How can sentiment analysis and time series models and machine learning models impact cryptocurrencies return, and how well different methods can predict and assess financial risks?

D. Outcomes of Research

The outcomes of this research aim to deepen our understanding of the interplay between sentiment

and cryptocurrency returns, evaluate the performance of various predictive time series and machine learning models, and assess financial risk in the context of cryptocurrency investments.

E. Research Contribution

This study aims to analyze the Rate of return (RoR) of three selecting cryptocurrencies and contribute significantly to the field by enhancing our understanding of how social media sentiment impacts cryptocurrency returns.

The study evaluates sentiment analysis tools like VADER, providing insights into their effectiveness in the cryptocurrency context. Therefore, the research helps compare the standard TS (time series) models, ARIMA and SARIMA-with the ML (machine learning) models, such as LSTM and Bi-LSTM, to select the most reliable predictive methodologies. It also applies Value at Risk measures for financial risks to offer a comparison across different cryptocurrencies.

The study supports, through effective visualizations, communication to the audience trends and patterns in sentiment that are significant. It returns, aiding decision-making and providing further understanding of market dynamics. Collectively, these contributions show further knowledge in sentiment analysis, predictive modeling, and risk management of cryptocurrency.

II. LITERATURE REVIEW

In this part, we analysis inclusive literature to focus on the understanding of present impact of social

media and other financial market variables on cryptocurrency return. To address shortcomings in previous research, [7] Kate Murray introduces an extensive framework for comparing bitcoin cryptocurrency price prediction models. Extensive experiments across five popular cryptocurrencies reveal that deep learning approaches, particularly LSTM, consistently outperform other models. The findings highlight LSTM's superiority with average errors RMSE (Root Mean Square Error) of 0.0222 and MAE (Mean Absolute Error) of 0.0173, providing valuable insights for traders and investors, and the dataset and code are shared to encourage reproducibility and more research.

Natural Language Processing (NLP) and Machine Learning (ML) in predicting Bitcoin and Ethereum prices based on social media and news data also treats output forecasting as a classification problem, demonstrating that integrating NLP data significantly improves model performance. Effectively capturing market sentiment are pre-trained models such as Twitter-RoBERTa and BART MNLI, demonstrating the promise of text analysis for economic forecasting [8]. The results show stable profitability in many circumstances, highlighting the long-term influence of NLP data on predicting accuracy. A machine learning-based method for expecting Bitcoin prices that addresses acquisition difficulties brought on by market volatility. The method outperforms previous approaches and shows increased consistency across different cryptocurrencies. It combines secure blockchain architecture with a reinforcement learning algorithm. For mortgage lenders monitoring the complex nature of digital currency investments, the recommended strategy offers a more efficient option [9].

With the help of a combination of DL (deep learning) methods, statistical, and text analytics, this paper presents a methodology for currency trading decision optimization. This approach measures variables such as how changes in price impact posts on social media, the causal effect of prices on

each other, and the attitude of the user on social media towards prices by using big data on the market and social media. The successful trading algorithm is demonstrated by trial results, which show an average gain of 194% excluding transaction costs and 117% after fees.

Significantly exceeding other cryptocurrency types and producing noticeably larger profits, the algorithm is particularly good at predicting price trends for meme coins [10]. Considering the complexities of cryptocurrency exchanges, [11] Victoria Tyshchenko, study enhances the performance of recurrent neural networks (RNNs) in cryptocurrency price prediction through the use of a target variable formula based on the arithmetic mean. Features including daily trade volume, price volatility, and technical indicator indications are all incorporated into the model. The research uses Sortino and Sharpe ratios to assess the models and makes use of both GRU and Long Short-Term Memory (LSTM) networks. GRU-based model showed better risk-adjusted returns, showing increased efficacy on digital currencies exchanges, even if both models' accuracy (69%) was comparable. The results highlight the improved prediction accuracy of recurrent neural networks for cryptocurrency pricing.

Sentiment of the news on the pricing of digital currencies, with particular attention to the NFT market's (Green Satoshi Token) as well as GMT (Green Metaverse Token). The investigation, which spans March to April 2022 and makes used the algorithms like FinBERT and XGBoost, reveals a relationship between token prices and sentiment ratings from financial news. Interestingly, GST's sentiment score foretells price movements earlier than GMT, providing information that helps investors make the decisions well-informed. The study focused the significance of sentiment analysis in comprehending dynamics of the bitcoin market and how it affects pricing, offering useful instruments for identifying market trends [12]. Other study Examines on how sentiment analysis can be used to predict cryptocurrency prices, with a particular emphasis on Ripple, Ethereum, Bitcoin, EOS, and Cardano. A machine learning model, particularly LSTM, is employed for time series data analysis. And demonstrates improved prediction accuracy by assigning weights to sentiment scores based on the influence factor of individuals, notably enhancing LSTM model performance [13].

Additionally, a hybrid model combining the analysis features of dictionary-based text and weight of sentiment scores outperforms alternative models in predicting the prices of Cardano, EOS, and Ethereum. The study emphasizes that an increase in features does not always correlate with enhanced prediction model accuracy [13].

In recent research, [14] develops an end-to-end algorithm to estimate Bitcoin values by fusing historical cryptocurrency price data with sentiment analysis of tweets. The approach, which employs a Bidirectional Encoder Representations from Transformers (BERT)-based Neural Network approach for sentiment and a GRU for price prediction, incorporates metrics such as tweet volume, user following, and verification status. With a mean absolute percentage error of 3.6% for price prediction and 9.45% for sentiment, the results show how social media sentiment may be integrated for effective cryptocurrency forecasting.

In another study [15] over 250,000 daily transactions on over 40 global exchanges, and a market value of \$9 billion USD, traditional methods such as Holt-Winters exponential smoothing become unsatisfactory. The study considers the unique market dynamics and temporary status of Bitcoin in the financial landscape as it uses machine learning, more especially time series analysis, to find trends and predict closing prices.

Using a novel machine learning technique, past market data and sentiment from social media are used to forecast the prices of cryptocurrencies like Ethereum and Solana. Leveraging FinBERT for sentiment analysis on social media tweets, the model integrates daily sentiment scores with cryptocurrency price data. The hybrid LSTM-GRU model demonstrates superior performance in predicting price movements compared to benchmark models. This imaginative method highlights the implication of combining historical data and social media opinion to forecast the price of bitcoin more accurately [10] [16].

[17] Hybrid model that integrates LSTM and Change Point Detection (CPD) to enhance Bitcoin price prediction. Through the identification of sudden market changes, the model dynamically responds, performing better than LSTM alone in MAE, MSE, and RMSE. The method presents more robust and accurate predictions, enabling investors and analysts to make sound decisions in turbulent crypto markets.

[18] The research applies LSTM, GRU, Bi-LSTM, and hybrid LSTM-GRU models with Min-Max scaling for prediction of prices of 20 cryptocurrencies. It provides more accurate results compared to conventional approaches and provides daily buy, sell, or hold suggestions to support decision-making during unstable markets. Other research establishes that for four large cryptocurrencies, LightGBM provides the most lucrative predictions for Bitcoin, Ethereum, and Litecoin but Simple RNN performs best for Ripple. Both ensemble and deep learning models prove effective, with results consistent across time periods [19].

III. Proposed Methodology

Each methodology helps in understanding and predicting cryptocurrency rates of return, providing insights into their relationships, and assessing financial risks. These proposed methodologies show in Figure 1.

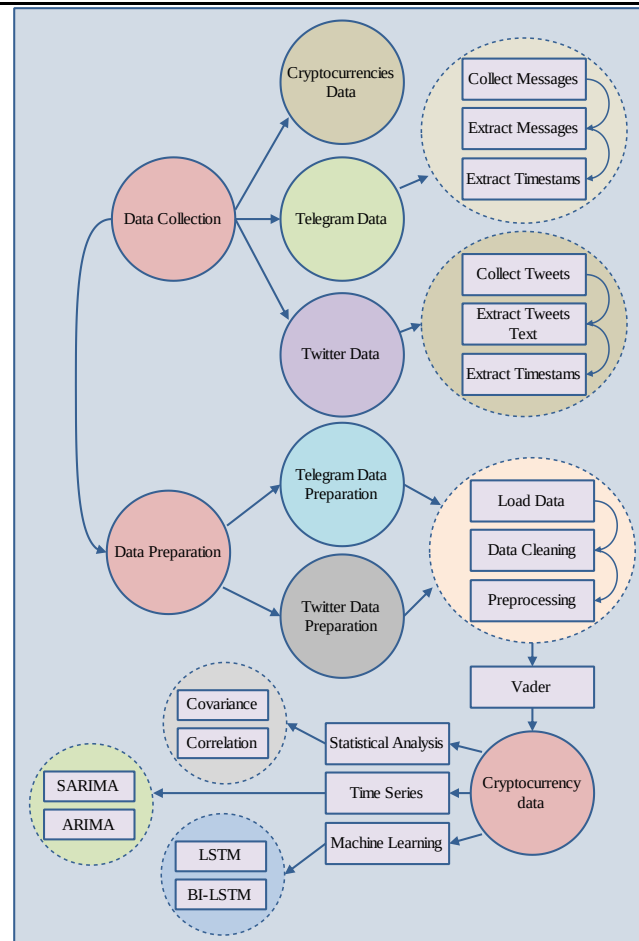


Figure 1. Proposed Methodology

A. Data Collection

The collection of data is an essential step in this analysis, focusing on collecting historical data on digital currency and sentimental information from Twitter and Telegram. Each data source serves a distinct purpose also adds to the comprehensive analysis.

1) Collection of Cryptocurrency Data

The research focuses on three major cryptocurrencies: Bitcoin, Ether, and Binance Coin. The reason for choosing these three digital currencies is because of their large market capitalization, wide usage, and the huge volume of trading that these cryptocurrencies have gained over time, for which any analysis that concerns market behavior would fit well with external factors such as social media sentiment. Prices of Bitcoin, Ether, and Binance Coin have been retrieved from the historical data provided by financial data can be obtained from financial data providers such as Yahoo Finance. Such systems make available most of the datasets comprising date, open, high, low, closing prices, trading volume, and market capitalization. The dataset is collected in the range of dates from 19th July 2023 to 19th July 2024, so that it coincides with the sentiment data collected from Twitter and Telegram.

2) Collection of Twitter Data

For the same purpose, with respect to getting sentiment data from Twitter about Bitcoin, Ether, and Binance Coin, the Twitter API is utilized. With the same accuracy, all relevant words, hashtags (like #Bitcoin, # ETH, Binance), and user mentions are filtered out with respect to the same tweets. Then, it is included tweet text, details about users, timestamps, retweet counts, and other useful metadata. Applying NLP methods like VADER, the tweet text is passed under sentiment analysis. The sentiment indicators involved in this analysis are positive and negative words, emoticons, and sentence structure. All this generates sentiment scores that make it possible to identify the general mood and the overall trend of the sentiment expressed by the users in Twitter towards these three cryptocurrencies.

Further analysis enables the researcher to understand trends, correlations, and even patterns in the data collected; thus, impact of certain news events and market fluctuations becomes clearly visible on the general public's perception of these cryptocurrencies.

It then can be correlated with historical price data to look for any relationships that may exist between sentiment and market movements. Analysis of the Twitter sentiment feed can provide really interesting new insights into how the world sees Bitcoin, Ether, and Binance Coin.

3) Collection of Telegram Data

The third step would involve collecting and analyzing sentiment data about cryptocurrency from Telegram, which is accessed public and private channels/groups via Telegram APIs and web scraping techniques. This data would be based on user messages, timestamp information, channel/group metadata, and relevant information of the users. All text data is cleaned and preprocessed, taking tokenization, stemming, and stop word removal to get ready for sentiment analysis. Sentiment scores extracted from using NLP techniques include sentiment polarity, specifically positive, negative, and neutral. We can use it to find the overall mood and sentiment of the Telegram users toward cryptocurrency besides identifying trends and patterns in sentiment over time.

However, the data collected can be further broken down to observe if there are any relationships between sentiment and other variables like market trends or cryptocurrency projects. Analyzing the sentiment information available within Telegram would give valuable insights on how public views about cryptocurrencies are formed within the Telegram group. Such insights could help investors, traders, and researchers in making intelligent decisions and understanding cryptocurrency dynamics.

B. Data Preparation

Data preparation is considered an intrinsic part of the entire process of analysis because it ensures that data applied for modeling is clean, consistent, and suited for meaningful insights. This stage is divided into loading data, cleaning it from errors or inconsistencies, and preprocessing to make it compatible with machine learning algorithms. Each of these steps in the process of data preparation is elucidated in detail below:

1) Data Loading

Python libraries like Pandas are used to load data from various sources into the analytic environment. Tabular data can be dealt with efficiently while managing Python libraries. This makes possible effective manipulation and analysis of gathered data.

2) Data Cleaning

These include errors, inconsistencies, and missing values, among others to be eradicated from

data once it starts loading.

This might involve operations such as:

- **Handling missing values:** Identify the existence of missing values in datasets using techniques known as imputation or deletion.
- **Data Error Correction:** this involves correcting any typos, formatting inconsistencies, or outliers and to remove duplicate records in order to assure the data's accuracy.
- **Data formatting:** The data should be formatted to have the same format throughout; for instance, by ensuring dates are equalized to the same format or all categorical variables are in the same form.

C. Data Preprocessing

Data preprocessing is an essential step in analysis, which makes sure data should be in some form of useable format for prediction and analysis. To change raw data into a more useable form, there are various approaches that are required.

- **Feature Engineering:**

Feature engineering refers to the process by which engineers either new variables or transform existing ones to improve the predictability of the data. For cryptocurrency data, this could mean computing moving averages, volatility measures and lagged variables to capture the trends and patterns. More features for sentiment data include sentiment polarity, positive, negative, and neutral sentiments, as well as more intensive scores for further understanding market sentiment and likely impact on the crypto price.

- **Normalization:**

Normalization brings different variables, particularly price and sentiment, on a common scale. Scaling values between 0 and 1 allows the model to treat each variable equally instead of being dominated by a more significant magnitude of variables. This can enhance the accuracy and interpretability of the model.

D. Rate of Return Calculations

The RoR or rate of return is a crucial factor in financial asset performance assessment. In our study, we calculated RoR for each of the chose cryptocurrencies in the following manner:

$$RoR = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1)$$

Where P_t represent the price at time t , and P_{t-1} is the previous day price. This formula was applied to every day in the dataset and generated a series for each day's rate of return for Bitcoin, Ether, and Binance. Input derived from this analysis offers insight into profitability over time for each asset, as well as volatility.

E. Sentiment Analysis

It is believed that sentiment analysis is one of the most critical methodologies to identify the psychological factors behind changes in markets as it can sense opinions and emotions in the social media data. The research applies sentiment analysis to analyze the sentiment of the textual data received from Telegram and Twitter. In this approach, the unstructured content is translated into measurable sentiment scores that are valuable for market research and predictive modeling. This process involves the application of NLP to clean and preprocess the text data so that unnecessary items are removed, such as spam or discussions far from the topic.

There are various methods and algorithms applied by the approach, including VADER (Valence Aware Dictionary and sentiment Reasoned) and the tone of cryptocurrency talks, to determine which sentiments are relevant. As a result, the effectiveness of VADER in sentiment analysis

from the social media level is one of its best features because it considers punctuation, emojis, and slang that dominate usage in casual texts. After categorizing text segments into either positive, negative, or neutral categories, these methods assign an intensity score indicating to what extent the intensity of the sentiment is represented. The aggregation of these scores on a daily basis would reflect the overall mood in the market. This would further provide daily sentiment index data in synchronization with the time series present in cryptocurrency prices. This is a major step because now the researcher can directly correlate changes in sentiment with the daily rates of return and provide some insight into how social media sentiment influences market behavior. Then, these aggregated sentiment scores are combined with the cleaned cryptocurrency data into a complete dataset to facilitate analysis of how market sentiment drives price movements. This combined dataset in the forms of other analyses can be performed to work with the productiveness of sentiment modeling and forecasting the return on cryptocurrency.

F. Data Merging

This phase involves data merging, an essential process in which cleaned and preprocessed data of cryptocurrency and sentimental information extracted from Twitter and Telegram must be combined. This step creates a merged dataset where the sentiment scores should be matched with the daily return rates of each cryptocurrency as accurate as possible to be able to perform an intensive analysis on how the market sentiment may influence the price movements.

It first of all triggers the merging process by ensuring that all the datasets share some common thing; it could be a date on which the datasets align. This is because the sentiment data must match in terms of timing with the performance metrics of a particular cryptocurrency if one is to analyze any causal or correlative relationship. It undergoes aggressive preprocessing: cleaning the missing values, standardizing the format, and making date-time consistent in order to avoid any mismatch in

such formats, this might skew the analysis. After that, it may be joined with other datasets: according to the research design, common approaches are time-series merge and left join.

This ensures that the right rate of return for each day is coupled with the appropriate sentiment score, even when days show sparse or inconsistent sentiment data. For example, days that are characterized by missing sentiment data because of outages of the platforms or low user activities are carefully managed to not compromise the integrity of the merged dataset.

It contains a single row for each day, with feature variables including the prices of cryptocurrencies, computed rates of return, and aggregated sentiment scores originating from Twitter and Telegram. The structured format for the data presented herein allows for an in-depth temporal analysis where changes in sentiment can be related directly to shifts in market performance. This therefore means that further analysis, which might involve correlation studies, time series forecasting, and predictive modeling, is based on the merged dataset. In this regard, merger involves the combination of data from diverse sets of sources, thereby providing a full market dynamic view in a manner that isolated datasets would not have been able to provide.

G. Evaluation Metrics

For performance comparison, a time series model and ML models are applied with two measures to compare the prediction accuracy: mean absolute percentage error and root mean squared error.

MAPE is the total of the absolute sum of the prediction error divided by the total of the absolute values. It is calculated as an absolute percentage error and scale invariant. Where n is the nod of fitted points, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(y_i - \hat{y}_i)}{y_i} \right| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

The average amount of prediction mistakes is quantified by the Root Mean Square Error (RMSE). In the above formula, y_i is the actual price, $\hat{y}_i$ is the predicted price, and n is the number of observations.

H. Data Analysis

During the analysis stage, it is important to be able to understand the associations between the sentiment and the performance of market. This stage also helps in checking the accuracy of the models used for forecasting. Various techniques, such as correlation and covariance analysis, time series modeling, and even machine learning methods, come into play at this stage.

1) Correlation and Covariance Analysis

The current research study uses statistical analysis to link cryptocurrency returns with scores of sentiments. With the help of Pearson Correlation, the strength of the connection and direction are derived: positive or negative sentiment and its relation to market fluctuations. For example, if the cryptocurrency prices upsurge, then an increase in sentiment follows the increase in cryptocurrency prices; however, if the relationship is negative, then the opposite would be true.

$$Correlation(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (4)$$

In this case, σ_X and σ_Y refer to the standard deviations of the rate of return series of X and Y , respectively. The coefficient can vary between -1 to 1; such as values close to 1 indicate a strong positive relationship, those around -1 show a strong negative relationship, and values close to 0 imply no linear relationship at all. This also informs us if the two variables move together and how. Thus, it provides more profound insights into their joint variability which is said to be highly crucial to determine whether sentimental data can be a reliable predictor of market trends. The covariance between two assets was computed as follows:

$$Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} \quad (5)$$

Where X and Y produce the two cryptocurrencies return series, $\bar{X}$ and $\bar{Y}$ are their means and n are the observation number. This covariance formula, therefore, provided directional relationship between the two cryptocurrencies.

2) Time Series Models Analysis

The data in the TS is forecasted to be using ARIMA and SARIMA models. These are the models that include ARIMA to identify patterns and then make a forecast in historical information about prices. ARIMA is excellent in investigating the data with trends, seasonality, and randomness-then it enables analysts to express the essential form of the time series and provide accurate forecasts.

•ARIMA Model:

This model is determined by the parameters (p, d, q) , where p show the autoregressive component, d signifies to the differencing component, and q denotes to moving average component. The choice of the model was made by using the Akaike Information Criterion (AIC).

• **SARIMA Model:**

This model expands the ARIMA framework with seasonal parameters (P, D, Q, s), which include seasonal variation. Herein, P, D, and Q refer to seasonal aspects, whereas s refers to the length of the seasonal cycle.

To assess the predictive accuracy of models, the two performances metrics, RMSE and MAPE, have been used. Further, this study is useful in comparison of ARIMA and SARIMA models in capturing both time and seasonality patterns in returns of cryptocurrencies. These models are handy baselines against which more advanced machine learning models like LSTM and Bi-LSTM could be compared to hence giving a better sense of whether the most advanced neural networks do improve performance significantly above traditional statistical methods.

3) Machine Learning Models Analysis

The machine learning LSTM and Bi-LSTM models are used for forecasting the return rates of different cryptocurrencies. Selecting these models is based on their capability to take care of temporal dependencies and sequenced handling of data. In contrast, unlike the previous models, the LSTM and Bi-LSTM networks have learned from the past trend and adaptation for non-linear patterns to yield to effectively fit the dynamic and volatile cryptocurrency market.

The evaluation of the performance of these models reveals the strengths of advanced machine learning techniques in many aspects, proof that such approaches can offer valuable and feasible forecasts of potential future market direction, evaluated with historical data and sentiment analysis.

Model Architecture and Training

- An LSTM model is developed with an input layer. An input layer may be added by one or more LSTM layers, followed by the dense layers along with ends output layer. The hyper parameters would include the number of neurons, dropout rate, and learning rates tuned by grid search.
- Bi-LSTM is similar to the LSTM; however, input sequences were processed in the two directions forward and backward with purpose of maximizing captures of complex patterns.
- **Training and Testing:** These models are trained with 70% of the data and tested with the rest 30%. In back-propagation, an optimizer is used, namely Adam optimizer. Each model is evaluated in terms of RMSE and MAPE metrics.

1. Value at Risk (VaR) Analysis

Value at Risk helps estimate and, thereby, measure the financial risks associated with investments in cryptocurrencies. A widely used methodology of VaR analysis provides for determining the absolute loss value an investor may face in the near future that a given confidence level takes into consideration within the historic volatility of the asset. Especially because of volatile price fluctuations and rapidly changing markets, the adoption of such analysis is substantial in the case of cryptocurrencies. VaR calculation enables investors to evaluate potential downside risks from portfolios, which aids investors in making the proper informed choices over risk management and asset allocation.

It follows that VaR calculates through the study of historical price data and return distributions to give probability that large losses have occurred. This methodology therefore gives a quantitative measure for risk exposure, which is helpful for an investor in ascertaining the possible financial implications of adverse market movements. Other than that, VaR is always administered conjunction combine with other risk assessment measures in order to present a comprehensive view of the risk characteristics of an investment, which can sometimes be very critical in providing insights into the possible weaknesses and strengths of cryptocurrency assets in volatile markets.

$$VaR_{\alpha} = -Quantile_{\alpha}(Returns) \quad (6)$$

The letter alpha can denote the confidence level, and the values it may take can be: 0.90, 0.95 or 0.99, while the term Quantile is represented by the word for the threshold value below which a given percentage of observations lie.

IV. RESULTS AND DISCUSSION

A. *Experimental Setup*

Experimental setup requires ANACONDA NAVIGATOR with Jupyter notebook for implementation and analysis. From this research, the results thus provide an in-depth exploration of the social media sentiment and cryptocurrency returns relationship, as well as the competence of the various predictive models in use. This section presents the output of integrated data, including results of sentiment analysis, predictions of cryptocurrency RoR, and VaR analysis.

This chapter presents the outcome of analysis done on a merged dataset of Bitcoin, Ether, and Binance combined with sentiment data on Twitter and Telegram from July 19, 2023, to July 19, 2024. Key findings including statistical tests, machine learning model results, and VaR estimates are forwarded. A critical assessment of how ARIMA, SARIMA, LSTM, and Bi-LSTM models have been performing at predicting cryptocurrency returns is also conducted.

This brings the datasets of Bitcoin, Ether, and Binance into the standard scope of 230 days, hence forming a single dataset for the cryptocurrencies. The data from Telegram covered the period from December 22, 2022, up to July 20, 2024, which had a total of 8,537 records and are then merged into the final dataset. However, only 8,513 records from the combined cryptocurrency data were successfully matched with the Telegram data for precise analysis.

B. *Data Merging and Integration*

The successful integration of cryptocurrency and sentiment data effectively correlated historical price information with sentiment scores derived from Twitter and Telegram. The datasets were aligned according to a shared date range spanning from July 19, 2023, to July 19, 2024. Although there was a minor variation in the number of overlapping dates, the merging process guaranteed that sentiment scores were precisely aligned with daily cryptocurrency returns.

After filtering, the final datasets for Bitcoin, Ether, and Binance Coin each comprised 16 data points, while the sentiment data encompassed 18 data points. This integration enabled a more thorough analysis by merging financial metrics with sentiment indicators, thereby facilitating the investigation of the impact of sentiment on cryptocurrency returns.

C. *Sentiment Analysis Results*

This SA of the extracted Twitter and Telegram data provided specific insight into market sentiment. Using NLP tools such as VADER, the sentiment expressed in the posts of social media was numerically translated into scores. By this transformation, a direct comparison between sentiment and returns of the cryptocurrency became possible.

The patterns of sentiment scores were striking in the fact that positive sentiment often appeared when prices were on the rise and negative when they were falling. Using the time-series data, spikes in positive social media sentiment usually coincided with higher cryptocurrency returns, while the opposite was true for negative sentiments. These results hint at the fact that probably one of the drivers of market movements is social media sentiment.

Both Telegram and Twitter sentiment analysis of cryptocurrencies and RoRs show in Figures 2, 3,

4, 5, 6 and 7 using dotted graph.

D. Rate of Return Analysis

The RoR calculations also showed outright differences of the daily returns among the three cryptos. On July 20, 2023, Bitcoin has a negative return of -0.4075. Comparatively, Ether returned 0.1038, while Binance had an even higher return of 0.6520. This variation proves intrinsic volatility as well as uniqueness of these digital assets.

Bitcoin Rate of Return

The Bitcoin return rate was highly volatile, showing extreme downfall, for instance, on September 10, 2023, at a -11.44% level. This volatility underscores Bitcoin's responsiveness to market conditions, external developments, and the sentiments of investors.

Ether Rate of Return

Ether's volatility was comparable, with a sharp fall of -11.50% on September 10, 2023, and a high positive return of 0.10% on July 20, 2023. This pattern makes it extremely evident that the top cryptocurrencies are connected in terms of how they respond to shifts in market sentiment and outside disruptions.

Binance Rate of Return

Binance has a similar return on investment curve as well, with peaks and troughs signifying profits and losses within the period of concern. The sharp drop of -10.29% on September 10, 2023, mirrors general downtrends in the cryptocurrency market and indicates that systemic risk factors have been at play and applied equally to all three assets.

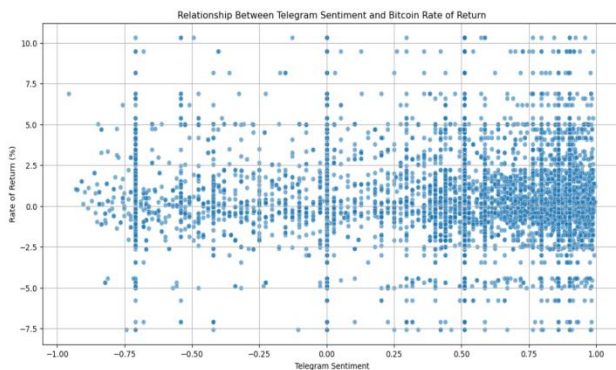


Figure 2: Telegram Sentiment and Bitcoin RoR

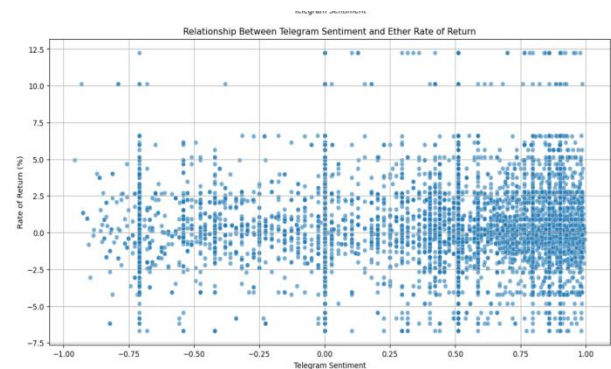


Figure 3: Telegram Sentiment and Ether RoR

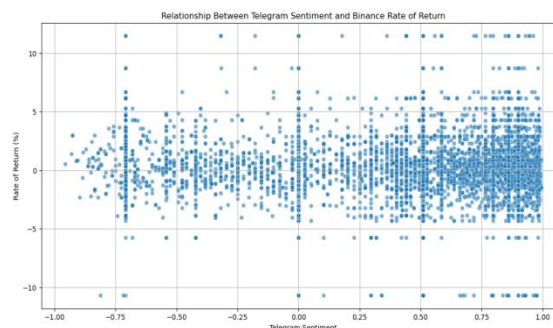


Figure 4: Telegram Sentiment and Binance rate of return

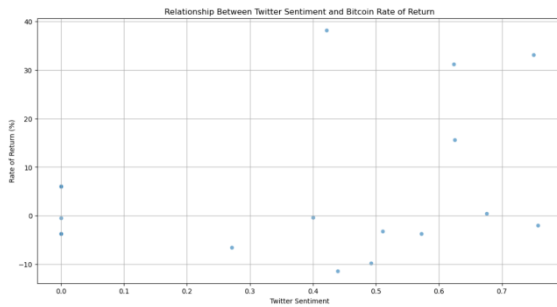


Figure 5: Twitter Sentiment and Bitcoin rate of return

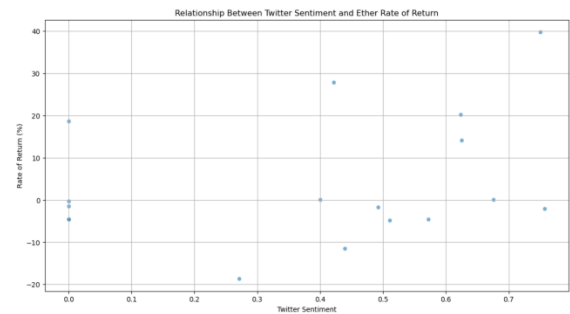


Figure 6: Twitter Sentiment and Ether rate of return

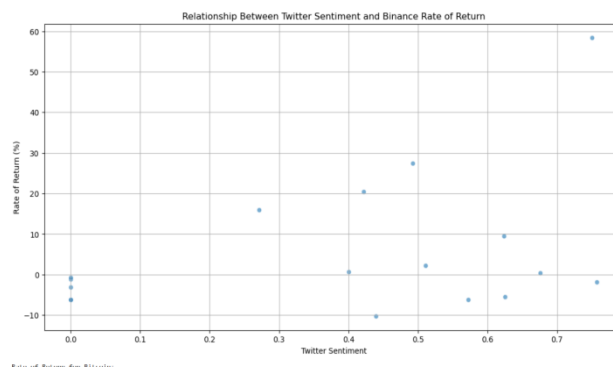


Figure 7: Twitter Sentiment and Binance rate of return

E. Prediction Analysis

1) Correlation and Covariance Analysis

The covariance matrix revealed significant relationships among the returns of various cryptocurrencies. For example, on Twitter table 1 results, Bitcoin and Ether exhibited a high covariance of 209.43, indicating that their returns tend to move in tandem. Additionally, the correlation matrix reinforced these observations, as Pearson's correlation coefficients demonstrated strong positive correlations (0.9094 between BTC and ETH; 0.5893 between ETH and BNB). These statistics highlight the interconnectedness of the cryptocurrencies.

Table 1: The covariance and correlation between RoRs of three cryptocurrencies for Twitter

cryptocurrencies	covariance	Pearsons'r Correlation
BTC-ETH	209.4299	0.90935

BTC-BNB	140.1702	0.53604
ETH-BNB	15.7789	0.58934

The covariance matrix in Telegram table 2 revealed positive correlations among the cryptocurrencies. Specifically, Bitcoin and Ether displayed a covariance of 4.5313, whereas Bitcoin and Binance recorded a covariance of 2.8119.

The correlation matrix reinforced these findings, showing that Bitcoin and Ether had a Pearson correlation coefficient of 0.7339, which signifies a strong positive relationship. In contrast, the correlations between Bitcoin and Binance, as well as Ether and Binance, were more moderate, with coefficients of 0.4952 and 0.4973, respectively.

Table 2: The covariance and correlation between the RoRs of three cryptocurrencies for Telegram

cryptocurrencies	covariance	Pearsons'r Correlation
BTC-ETH	4.531258	0.7339
BTC-BNB	2.811876	0.4952
ETH-BNB	2.880711	0.4973

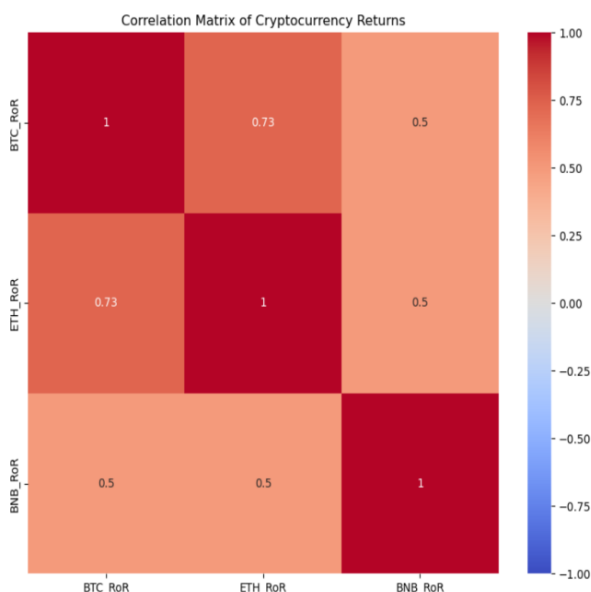
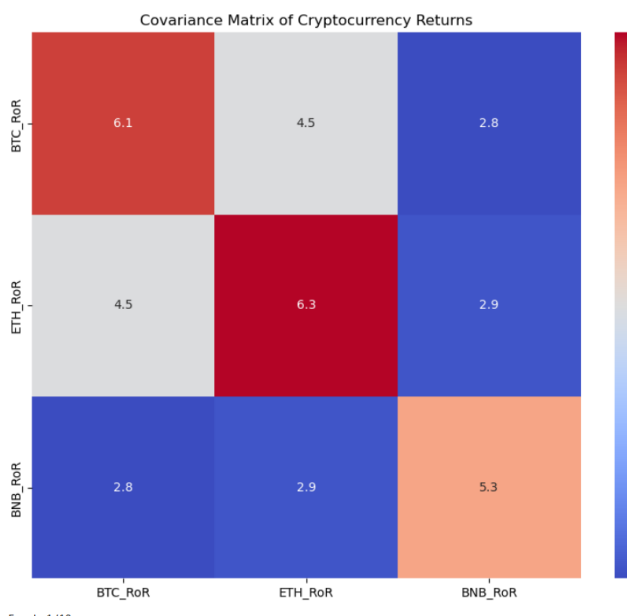


Figure 8: Telegram Correlation Matrix of cryptocurrency returns.

Figure 9: Telegram Covariance Matrix of cryptocurrency returns.

cryptocurrency	Models	RMSE	MAPE
----------------	--------	------	------



2) ARIMA and SARIMA Models performance

Conventional time series models, including ARIMA and SARIMA, demonstrated suboptimal RMSE of 12.7371 and a MAPE of 162.85%, whereas the SARIMA model produced performance. For instance, in Twitter table 3, the ARIMA model applied to Bitcoin yielded an RMSE of 17.6074 and a MAPE of 185.53%. Comparable patterns were noted for Ether and Binance Coin. These findings underscore the inadequacies of traditional models in effectively capturing the non-linear and dynamic characteristics of cryptocurrency returns when contrasted with machine learning approaches.

Table 3: Time Series Models Results for Twitters

BTC	ARIMA	12.7371	162.85%
BTC	SARIMA	17.6074	185.53%
ETH	ARIMA	14.8080	185.25%
ETH	SARIMA	15.5129	319.97%
BNB	ARIMA	16.8110	211.07%
BNB	SARIMA	20.2090	296.68%

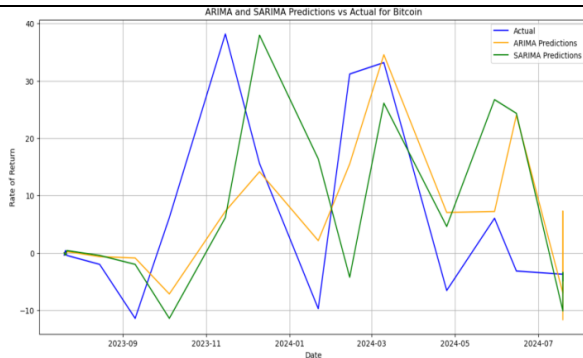


Figure 10: Time Series Models Prediction for Bitcoin using Twitter.

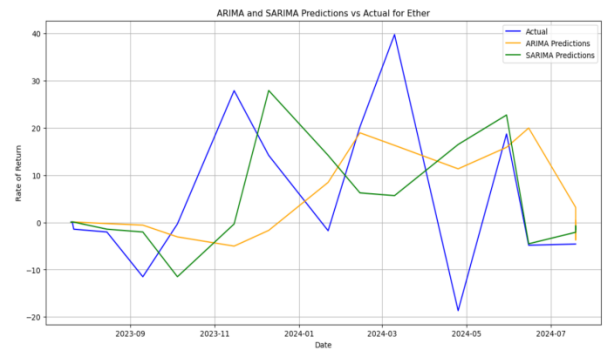


Figure 11: Time Series Models Prediction for Ether using Twitter.

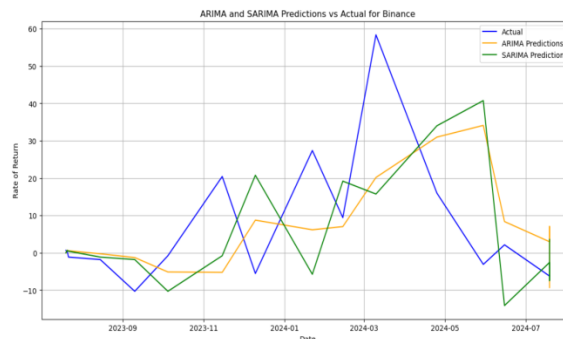


Figure 12: Time Series Models Prediction for Binance using Twitter.

In the Telegram analysis of time series results data display in table 4, both ARIMA and SARIMA models were utilized to predict the rate of return. RMSE of 0.5228 and MAPE of 17.15% were by the ARIMA model for Bitcoin. On the other hand, SARIMA had a slightly less superior RMSE of 0.5235 and an MAPE of 36.46%. For Ether, the model of ARIMA resulted in an RMSE of 0.5690 and MAPE of 15.10%, whereas the SARIMA model presented almost equivalent results with RMSE of 0.5697 and 18.53% MAPE. For Binance, the model of ARIMA achieved an RMSE of 0.5024 and MAPE of 13.01%, whereas the SARIMA model presented results with RMSE of 0.5028 and MAPE of 17.37%. Overall, in all the cryptocurrencies, the ARIMA models outperformed the SARIMA models, which captured the overall tendency of ARIMA models to capture the temporal trend in the data.

Lastly, both ARIMA and SARIMA predictions vs actual prediction show in line graphs 10, 11 and 12 for twitter and figure 13, 14 and 15 for telegram.

Table 4: Time Series Models Results for Telegram

cryptocurrency	Models	RMSE	MAPE
BTC	ARIMA	0.5228	17.15%
BTC	SARIMA	0.5235	36.46%
ETH	ARIMA	0.5690	15.10%
ETH	SARIMA	0.5697	18.53%
BNB	ARIMA	0.5024	13.01%
BNB	SARIMA	0.5028	17.37%

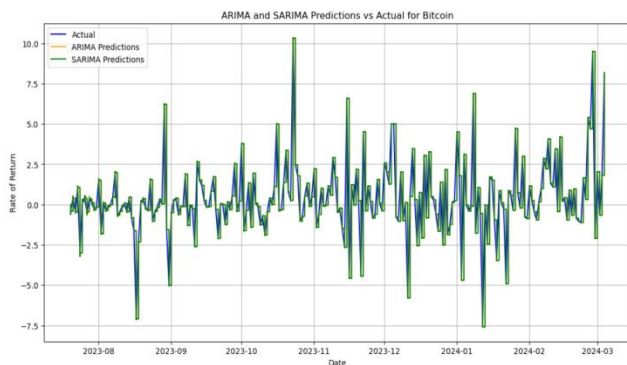


Figure 13: Telegram-Based Time Series Models for Bitcoin Prediction

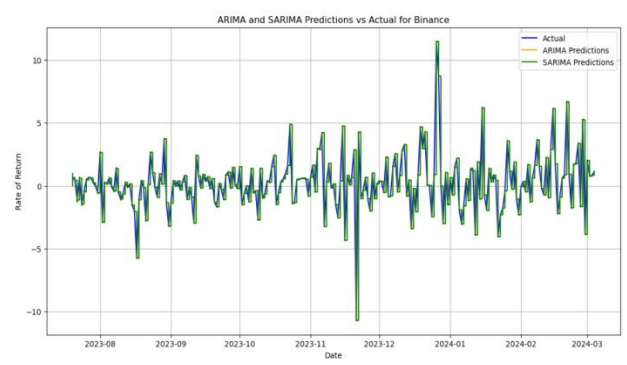


Figure 14: Telegram-Based Time Series Models for Binance

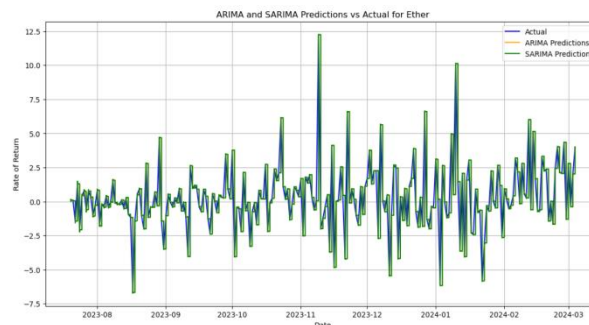


Figure 15: Telegram-Based Time Series Models for Ether

3) Machine learning model Performance

Related to the predictive capabilities of traditional ARIMA and SARIMA models, the evaluation was conducted on the LSTM and Bi-LSTM models. Hence, the following remarkable results were obtained: table number 5 illustrates that the LSTM model exhibited superior performance with a Root Mean Square Error at 3.1917 as well as a Mean Absolute Percentage Error at 85.28% for Bitcoin. For Ether, the LSTMs indicated an RMSE value of 7.3812 and a MAPE value of 161.20%. For Binance Coin, the LSTMs resulted in an RMSE value of 4.8157 and an MAPE value of 77.40%. These reflect that LSTM models work well in extracting the inherent sequential dependencies present in cryptocurrency price data. However, results from the Bi-LSTM model, where it took data in both directions-forwards and backwards-were not consistent. It was observed that for Bitcoin, Bi-LSTM resulted in 6.0261 RMSE and 161.11% MAPE. However, in the case of Ether, the Bi-LSTM performed better than LSTM as it resulted in 1.5492 RMSE and 33.82% MAPE. For Binance Coin, Bi-LSTM almost outperformed LSTM as it gave an RMSE of 1.9397 and a MAPE of 31.20%. The enhanced performance of Bi-LSTM for Ether and Binance Coin is due to the efficiency of bidirectional learning, which could capture more patterns.

Table 5: Machine Learning Models Results for Twitter

cryptocurrency	Models	RMSE	MAPE
BTC	LSTM	3.1917	85.28%
BTC	Bi-LSTM	6.0261	161.11%
ETH	LSTM	7.3812	161.20%
ETH	Bi-LSTM	1.5492	33.82%
BNB	LSTM	4.8157	77.40%
BNB	Bi-LSTM	1.9397	31.20%

The below table 6 indicate the performance for both LSTM and Bi-LSTM was tested in return rate forecasting on Telegram. In the case of Bitcoin, LSTM achieved 0.5402 RMSE and 1.10% MAPE. On the other hand, the Bi-LSTM machine model performed relatively poorly with an RMSE of 0.5665 and a MAPE of 1.73%. For Ether, RMSE was represented by the LSTM model at 0.6393 and MAPE at 0.96% whereas, for comparison, the Bi-LSTM model could show marginally better with the RMSE 0.6219 and a MAPE show 0.58%. Considering Binance, RMSE was recorded by the LSTM model at 0.5421 and the MAPE at 0.52% whereas in comparison, the Bi-LSTM model had an RMSE of 0.5680 and a MAPE of 0.81%. Those results suggest that, though Bi-LSTM models provide minute improvements in performance for Ether, LSTM models generally are better overall.

Table 6: Machine Learning Models Results for Telegram

cryptocurrency	Models	RMSE	MAPE
BTC	LSTM	0.5402	1.10%
BTC	Bi-LSTM	0.5665	1.73%
ETH	LSTM	0.6393	0.96%
ETH	Bi-LSTM	0.6219	0.58%
BNB	LSTM	0.5421	0.52%
BNB	Bi-LSTM	0.5680	0.81%

V. Value at Risk (VaR) Analysis

VaR analysis helped a lot in identifying the risk assets related to holding cryptocurrencies. VaR is the statistical measure used in calculating the potential loss a particular stock or a portfolio would suffer in a specific period (corporatefinanceinstitute.com). The following were the results:

VaR results for Twitter:

Bitcoin Value at Risk was at a -7.85% at 90% confidence level, at -10.10% confidence and reached -11.17% confidence level at 99% confidence. These values reflect the greatest possible loss over the time horizon for various levels of confidence. For Ether, VaR at a 90% confidence level was at -7.50%, while at 95% confidence it was at -12.93% and hit -17.52% at 99% confidence level. Ether was riskier than Bitcoin, especially at a level of confidence at 99%. The VaR for Binance was at a 90% confidence level, while at a 95% level, it became -7.03%, and for 99% level of confidence, it

was -9.64%. Binance Coin presented a risk profile that was much less than both Bitcoin and Ether; however, the risks are high in table 7.

Table 7: VaR analysis of Twitter's three cryptocurrencies

Confidence level (%)	BTC (%)	ETH (%)	BNB (%)
90	-7.85	-7.50	-6.22
95	-10.10	-12.93	-7.03
99	-11.17	-17.52	-9.64

Table 8: VaR analysis of Telegram's three cryptocurrencies

Confidence level (%)	BTC (%)	ETH (%)	BNB (%)
90	-1.77	-2.19	-2.03
95	-2.64	-3.63	-3.00
99	-5.79	-5.83	-5.58

VaR results for Telegram:

The VaR of Bitcoin was found at -1.77% with 90% confidence, at -2.64% when the confidence level reached 95%, and -5.79% for 99% confidence. In the case of Ether, the VaR values were -2.19% at 90% confidence, -3.63% at 95% confidence, and -5.83% at 99% confidence. For Binance, the VaR figures were -2.03% at 90% confidence, -3.00% at 95% confidence, and -5.58% at 99% confidence. These results represent the value-at-risk for each cryptocurrency at different confidence levels. Bitcoin displays relatively lower risk compared to Ether and Binance. The output is given in table 8.

VI. Comparison Discussion

The research findings highlight the significant attention of social media sentiment on the markets of cryptocurrency as well as display the benefits of using advanced machine learning models for predictions. By integrating sentiment data with financial metrics, we gain a deeper understanding of market dynamics, and LSTM and Bi-LSTM models demonstrate better forecasting accuracy than traditional time series models ARIMA and SARIMA.

A comparative analysis of sentiment data from Twitter and Telegram shows that both platforms offer valuable insights; however, Twitter's wider reach tends to reflect market sentiment more

immediately, while Telegram provides a more concentrated, community-driven perspective. Merging data from both platforms enriches our overall understanding of market sentiment. Additionally, the VaR analysis underscores the inherent risks tied to cryptocurrency investments, highlighting the necessity for thorough risk management strategies. These findings add to the expanding knowledge base on cryptocurrency markets and offer practical insights for both investors and researchers.

VII. CONCLUSIONS

The comprehensive analysis of Twitter and Telegram data, along with cryptocurrency rates of returns for Bitcoin, Ether, and Binance from July 19, 2023, to July 19, 2024, revealed important insights into how social media sentiment influences market movements. Both demonstrated promise for prediction, except that Twitter arrived with more brilliant, real-time sentiment because it's more concise, while Telegram gave a more all-encompassing broader sentiment landscape that needs to reflect a further discussion of topics. It has been found that LSTM and Bi-LSTM machine learning models are successful in terms of forecasting cryptocurrency returns, even though results vary based on the type of cryptocurrency with respect to the other time-series models like ARIMA and SARIMA. By comparing the ARIMA and SARIMA models with the ML models, it was found that they were better in individual circumstances but overall failed to be successful. Twitter sentiment was aligned relatively better with short-term price changes and is very handy for trading strategies involving short periods of horizons. Nevertheless, on the contrary, data from Telegram, although noisier, holds a much more wholesome sense of market sentiment, something tremendously helpful in interpreting the trends. Drawing both Twitter and Telegram insights together, this does balance the approach towards improving predictive accuracy through the immediacy of Twitter with the depth of strong evidence in Telegram as advice for trading and investment decisions within the unstable cryptocurrency market.

REFERENCES

- [1] I. A. P. R. P. V. M. E. S. M. A. a. M. S. Ahmed M. Khedr, "Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey," *Semantic scholar*, 2021.
- [2] D. Y. (. Peter Mell (NIST), "Understanding Stablecoin Technology and Related Security Considerations," *NIST*, 2023.
- [3] J. Saleemi, "Microblogging Perceptive and Pricing Liquidity: Exploring Asymmetric Information as a Risk Determinant of Liquidity in the Pandemic Environments," *researchgate*, 2023.
- [4] A. G. J. E. Jacques Vella Critien, "Bitcoin price change and trend prediction," *springer*, 2022.
- [5] Z. Z. M. S. F. M. A. G. F. R. Pavlo Seroyizhko, "A Sentiment and Emotion Annotated Dataset for Bitcoin Price Forecasting Based on Reddit Posts," 2022.
- [6] b. L. A. a. A. B. Abeer ElBahrawya, "Wikipedia and Cryptocurrencies: Interplay Between Collective Attention and Market Performance," *Semantic scholar*, 2019.
- [7] A. R. D. C. A. V. Kate Murray, "On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles," *MDPI*, 2023.
- [8] S. L. W. K. K. H. Vincent Gurgul, "Forecasting Cryptocurrency Prices Using Deep Learning: Integrating Financial, Blockchain, and Text Data," *Researchgate*, 2023.
- [9] J. A. P. V. D. K. O. G. I. C. C. N. L. Thanga Mariappan, "A Forecasting Approach to Cryptocurrency Price Index Using Reinforcement Learning," *Applied Science*, 2023.
- [10] D. C. ., C. C. ., F. M. a. P. T. Loris Belcastro, "Enhancing Cryptocurrency Price Forecasting by Integrating Machine Learning with Social Media and Market Data," 2023.
- [11] S. A. O. N. S. K. V. K. Victoria Tyshchenko, "DEVELOPMENT OF RECURRENT NEURAL NETWORKS FOR PRICE FORECASTING AT CRYPTOCURRENCY EXCHANGES," *Researchgate*, 2023.
- [12] T. C. Y. W. Eik Den Yeoh, "Predicting Price Trends Using Sentiment Analysis: A Study of StepN's SocialFi and GameFi Cryptocurrencies," *Semantic Scholar*, 023.
- [13] F. F. a. B. Amiri, "Cryptocurrency Price Prediction Model Based on Sentiment Analysis and Social Influence," *IEEE*, 2023.
- [14] H. G. B. a. S. N. B, "CRYPTOCURRENCY PRICE PREDICTION USING TWITTER SENTIMENT ANALYSIS," *IEEE*, 2023.
- [15] B. A. B. D. ., K. K. S. Ch. Shivani, "Forecasting cryptocurrency prices using Recurrent Neural Network and Long Short-term Memory," *ScienceDirect*, 2022.
- [16] S. ABBA SUGANDA GIRSANG, "Cryptocurrency Price Prediction Based Social Network Sentiment Analysis Using LSTM-GRU and FinBERTABBA SUGANDA GIRSANG1," *IEEE*, 2022.
- [17] C. L. W. S. Zheshi Chen, "Bitcoin price prediction using machine learning: An approach," 2019.

-
- [18] N. M. V. S. P. I. S. R. M. A. K. G. Akila V, "A Cryptocurrency Price Prediction Model using Deep Learning," in *E3S Web of Conferences*, 2023.
 - [19] V. P. G. G. I. B. L. S Kavithra, "Cryptocurrency price prediction through integrated forecasting techniques," 2024.
 - [20] M. Z. A. P. H. K. Y. Ahmed Bouteska, "Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods," vol. 92, 2024.