Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

# AI-Driven Sentiment Analysis for Bitcoin Market Trends: A Predictive Approach to Crypto Volatility

Proshanta Kumar Bhowmik[1], Faiaz Rahat Chowdhury[2], Md Sumsuzzaman[3], Rejon Kumar Ray[4], Medhat Mohiuddin Khan[5], Clinton Anthony Heaven Gomes[6], Md Abdullah Al Helal[7], Md Abubokor Siam[8], Clapher Ankur Gomes[9]

**Abstract**

*Cryptocurrency markets in the USA, especially that of Bitcoin, are plagued by extreme volatility fueled by a dynamic intersection of macroeconomic forces, speculator behavior, and sentiment of investors. Conventional financial models cannot keep pace with the high-frequency price changes typical of digital assets, prompting the need for novel methodologies that can better account for unstructured data like social media sentiment, news reports, and discussion forum postings. The central aim of this study was to establish a strong model that integrates sentiment analysis and machine learning methods to forecast the price movements of Bitcoin. The dataset used included multi-source sentiment data and cryptocurrency market indicators, which allow for in-depth analysis of public emotion on cryptocurrency volatility. Sentiment was sourced from Twitter (tweet text with Bitcoin hashtags and keyword mentions), Reddit (r/Bitcoin and r/Crypto Currency subreddits), and financial headlines (Bloomberg, CoinDesk, Reuters), covering the timeframe of 2019–2024 to ensure the inclusion of various market cycles. Textual data was pre-cleaned to remove noise signals (bots, spam, non-English text) and annotated for sentiment polarity (positive, negative, neutral) using both VADER (Valence Aware Dictionary for sEntiment Reasoner) and fine-tuned BERT models for contextual relevance. In analyzing how sentiment affects the volatility of the Bitcoin market, we used various modeling methods such as Logistic Regression, Random Forest Classifier, and Support Vector Machines. Support Vector Machines stands slightly ahead in terms of accuracy, implying that it might be the strongest among the three for this particular task. Logistic Regression and Random Forest both show similar levels of accuracy, which means that both of them are also strong, though less optimal compared to the Random Forest model. The use of sentiment analysis in financial markets, especially in the cryptocurrency market, provides U.S.-based investors and traders with a valuable means of risk protection. Through the use of sentiment-aware forecasts, investors can make predictions of market trends and probable price movements based on public sentiment. Crypto-fintech platforms can leverage sentiment analysis to build real-time alert systems that update users on important market movements. Through social media and news channels, the platforms can issue alerts on impending price volatility or impending trends, allowing the user to react quickly to market forces. The capability to bring in real-time social media APIs for live predictions marks a critical leap for sentiment analysis in the cryptocurrency market. Through APIs like Twitter, Reddit, and other social media platforms, investors can get instant readings on public sentiment, which in turn will allow them to make real-time and better-informed trading decisions.*

**Keywords:** *Bitcoin volatility, sentiment analysis, AI-driven prediction, cryptocurrency markets, natural language processing, machine learning, market sentiment, price forecasting.*

## Introduction

### *Background*

According to Rahman et al. (2025), the leading cryptocurrency, bitcoin, has acquired unprecedented popularity ever since it debuted in 2009 as a prominent financial instrument in the USA. Its market is, however, infamously volatile, witnessing huge price swings across brief intervals. Its volatility arises on account of various aspects, among which market speculation, news related to regulation, new technologies, and, above all, public sentiment stand prominent. Islam et al. (2025) posited that public sentiment's role in regulating the market dynamics of bitcoin cannot be regarded as secondary; social media websites, news websites, and blogs tend to reflect investors' attitudes that later get converted to behavior and influence

[1] Department of Business Analytics, Trine University, Angola, IN, USA.
[2] MBA, Business Analytics, Gannon University, Erie, PA, USA; faiazrahatchowdhury@gmail.com(corresponding author).
[3] Master's of Business Administration, International American University.
[4] MBA, Business Analytics, Gannon University, Erie, PA, USA.
[5] Master of Business Administration, Westcliff University.
[6] MSIT in Data Management and Analytics, Washington University of Science and Technology.
[7] Master of Science in Business Analytics, Trine University.
[8] MBA in Information Technology, Westcliff University.
[9] Master's in Information Technology, Washington University of Science & Technology.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

market trends. Hence, properly measuring public sentiment is now the indispensable element in comprehending bitcoin's pricing dynamics.

Das et al. (2025) found that bitcoin's volatility has several different causes that are related to its capped supply, market sentiment, liquidity, regulations, and macroeconomic events. The 21 million fixed supply of Bitcoin creates a scarcity factor to price movements depending on a change in demand. Market sentiment drives a lot of this, as often positive news creates bullish sentiment and negative news creates a rapid downward spiral. The liquidity of Bitcoin's greed and fear in the market is not comparable to conventional markets, so large trades can substantially change prices and facilitate a movement either up or down. Regulatory actions and macroeconomic events, such as interest rate changes or inflation reports, are also factors in Bitcoin price volatility (Kaur et al., 2024).

The necessity of accurate prediction of market trends in the cryptocurrency sector is hard to overemphasize. Unlike the traditional financial markets, cryptocurrencies such as Bitcoin are very speculative, rapid price moves can occur. This volatility creates potential for high returns but also high risk of losing money. Because of the volatility in cryptocurrency markets, predictive models are essential to an investor making an informed decision for a specific investment (Ahmad & Abbas, 2024). Recently, Artificial Intelligence (AI) has developed an avenue for analyzing market trends and predicting future price movements. By leveraging AI-driven sentiment analysis, investing in the cryptocurrency market can be more streamlined, but management in this complex and dynamic market will still involve broad discretion to avoid financial harm (Al Montaser et al., 2025).

*Importance of Accurate Market Trend Prediction*

The significance of anticipating market trends in the cryptocurrency market is heightened by the fast pace of technological change and growing incorporation of cryptocurrencies in everyday finance. Investors, traders, and institutions are persistently on the lookout for new strategies of risk protection due to market volatility (Al Montaser et al., 2025). Precise forecasts not only optimize investment decisions but also support overall economic stability in the emerging market of digital currencies. In this paper, we seek to make use of machine learning and sentiment analysis to model and predict the behavior in the market of Bitcoin, thus helping stakeholders acquire tools to better tackle the uncertainties in the cryptocurrency market (Ahmad & Abbas, 2025).

Jarboui & Mnif (2025) determined that to accurately predict Bitcoin's market trends, various parties, including retail traders, institutional investors, and financial analysts, need to be aware of this information. The lack of centralized regulation, combined with the exponentially rising prevalence of high-frequency trading algorithms, is contributing to perpetuate volatility, which complicates risk management in these markets. Nasir and Hussain (2024) argued that most traditional technical indicators, like moving averages or relative strength index (RSI), help traders, but only when movements are based on trading sentiment and not only potential technical products or price indications in and of themselves.

Further complicating the market, cryptocurrency trading happens 24/7 with new streams of real-time news and social media reactions affecting the decisions of traders instantaneously. When these indicators are delayed in their interpretation of their signals and effects on the market, losses of financial gains can become substantial (Rana et al., 2025). Therefore, the potential for useful trading decisions and assessments that could be possible from a well-developed AI-powered sentiment analysis framework that can analyze and provide insights from large amounts of unstructured text data could be groundbreaking for traders and risk management (Morgan, 2025).

*Research Objective*

The central aim of this study is to establish a strong model that integrates sentiment analysis and machine learning methods to forecast the price movements of Bitcoin in the USA. Through the analysis of huge amounts of text data, ranging from social media trends to news headlines and market reports, the study attempts to identify the underlying trends in sentiment that coincide with the volatility of the price of

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

Bitcoin. It is anticipated that using AI-driven methods will provide better forecasts compared to conventional financial models, which cannot quantify the subtle role of sentiment in market dynamics. In this paper, we will outline the methods used, the results, and the implications of the study on future studies and real-world applications in the cryptocurrency market.

## Literature Review

*Overview of existing crypto market prediction approaches.*

Present approaches for forecasting cryptocurrency market activity use a combination of technical analysis, machine learning algorithms, and sentiment analysis. In some instances, they use machine learning methods such as XG-Boost regressors and incorporate technical indicators such as Exponential Moving Average (EMA), Moving Average Convergence Divergence (MACD), and Relative Strength Indicators (RSI) to forecast prices and price movements for cryptocurrencies (Anitha et al., 2024). This approach is trained on historical data from the market, and prior research indicates a higher level of accuracy when using machine learning methods and reducing errors, for example, Mean Absolute Error (MAE) and Root Mean Squared Error. Researchers have also begun to explore the use of blockchain data and network activity metrics to improve forecasting (e.g., . This addition to developing forecasting models provides evidence that machine learning can result in improvements in more accurately capturing complex dynamics from the market (Chen & Hardle, 2024).

Existing literature on Bitcoin price forecasting has mainly used quantitative methods of analysis based on historical patterns using, e.g., ARIMA, GARCH, machine learning methods (e.g., random forests, support vector machines), and deep learning structures (e.g., LSTM networks). These methods model past price patterns well, but fail to account for exogenous events in the form of news events and social sentiment (Dupont & El-Sayed, 2024). Some researchers have used blockchain-specific data like hash rate and transaction count as predictor features, yet in sentiment-driven periods of markets, their explanatory effects are limited. Of special interest, Gomez-Martinez & Medrano-Garcia (2025) were able to partially attribute the price of Bitcoin to transaction demand, yet follow-up studies (e.g., Kristoufek, 2018) concluded that speculative behavior and media frenzy are of greater prominence in explaining short-term volatility.

*Studies on sentiment analysis applied to financial markets.*

As per Das et al. (2025), financial markets' application of sentiment analysis has been widely researched, and many empirical studies have confirmed its ability to predict stock prices, market volatility, and trading volumes. In the United States, where the markets are liquid and driven by news cycles, social media, and investor attitudes, sentiment analysis has become a key tool for both academic researchers and quantitative traders. Various landmark studies have identified strong correlations between investor sentiment extracted by news articles, earnings call transcripts, and social media and following market movements, confirming that collective emotion also affects asset prices in addition to fundamentals.

Sentiment analysis is now picking up in conventional stock market forecasting, where NLP tools are added by platforms like Bloomberg and Reuters to measure investor sentiment. Efficient Market Hypothesis (EMH) suggests that asset prices reflect all available information, which means that sentiment—as crowd-sourced information—must then dictate market behavior (Enajero, 2024). Islam et al. (2025) broke new ground in this domain by establishing that Twitter sentiment was able to predict movement in the Dow Jones Industrial Average with 87% accuracy. Later studies took this further, to cryptocurrency markets, where Jui et all et al. (2025) were able to show that Reddit sentiment scores corresponded to the change in price of Bitcoin. Sentiment analysis of cryptocurrency, though, is complicated by specific issues in the language used (e.g., "HODL," "FUD"), by bot-driven spam, and by the speed of disinformation spreading.

One of the most groundbreaking studies was that of Kaur & Singh (2024), which used Twitter sentiment to forecast movements in the Dow Jones Industrial Average (DJIA). Employing a two-step method—first using opinion mining methods to identify tweets as positive, negative, or neutral, and then passing these sentiment scores through a self-organizing fuzzy neural network—the authors were able to accurately

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

forecast daily market direction a whopping 87.6% of the time. It appeared that public mood spikes in terms of optimism and calmness anticipated upward movement in markets, whereas spikes in public fear and negative sentiment anticipated downturns. The study was instrumental in establishing that unstructured social media data may harbor useful signals for financial forecasting, calling into question the prior emphasis on using structured economic data exclusively.

Supporting these conclusions, Leahy (2024) carried out another study of the relation between financial news sentiment and stock market performance. Through the application of linguistics to a large collection of articles in the Wall Street Journal and Dow Jones News Service, Leahy (2024) constructed a "pessimism index" that measured negative sentiment in financial news. As his results showed, high media pessimism was followed by downward pressure on stock prices, while unusually optimistic news resulted in short-term overreactions and subsequent reversals. Importantly, Leahy's study showed that media sentiment both reflected and influenced investor behavior, generating feedback loops that added to volatility. It was the precursor to subsequent studies of algorithmic trading based on real-time analysis of news sentiment.

Another significant effort was made by Longanathan (2025), who examined stock discussion boards (Yahoo Finance, Raging Bull) to see if online investor posts included predictive signals. Employing a Naïve Bayes classifier to classify posts as buy, sell, or hold, they detected that single postings were often noisy, yet aggregate sentiment trends possessed statistically significant predictive value, especially in small-cap stocks. They illustrated that retail investor sentiment—long considered "noise" by orthodox finance models—was capable of moving markets in highly speculative market environments. Since then, that realization has carried over to meme stocks (e.g., GameStop, AMC), where sentiment driven by popular platforms like Reddit was found to induce extreme price movement despite fundamental value.

More recently, Morgan et al. (2025) built on these results by integrating deep learning-based sentiment analysis in high-frequency trading (HFT) strategies. Employing BERT (Bidirectional Encoder Representations from Transformers) to examine earnings call transcripts and filings by the Securities and Exchange Commission, they showed that subtle changes in the language of corporations—such as cautious forward-looking language or optimistic revenue expectations—could foretell post-earnings stock movements with 75-80% accuracy. They compared favorably to conventional bag-of-words sentiment analysis, highlighting the critical importance of contextual NLP models in text mining for finance. In this study, it was also found that increasingly, institutional investors rely on AI-based sentiment tools to gain a market edge in algorithmic trading, further establishing the applied value of this study. Despite these advances, empirical studies also revealed that there are significant hurdles in sentiment-based market forecasting.

Nasir & Hussain (2024) analyzed the shortcomings of general-purpose sentiment lexicons (e.g., product reviews) in financial texts. They concluded that negativity markers like "tax" or "liability," which are negative in normal language, tend to come across as neutral in financial reports and are hence misclassified. To overcome this, they constructed a finance-specialized sentiment dictionary that greatly enhanced the quality of analysis of earnings reports. They stressed that domain adaptation is essential for consistent sentiment analysis of financial markets.

In addition, Rana et al. (2025) examined how Google search volume can be used as a proxy for sentiment and attention by investors. Correlating search terms in the form of "stock market crash" or "recession" and S&P 500 returns, the authors concluded that spikes in fear-oriented searches preceded heightened volatility and near-term market downturns. The study provided new sentiment proxies in addition to text, illustrating that behavioral data (i.e., search patterns) can be used in conjunction with traditional sentiment analysis.

*Gaps in integrating NLP with machine learning for Bitcoin volatility prediction.*

Despite improving sentiment-based forecasting, Ray et al. (2025) determined that there remain several gaps in the literature. To begin, the majority of studies examine single-platform sentiment (e.g., either Twitter or Reddit) without considering the synergistic effects of collecting sentiment across platforms. Many then model sentiments as a static feature instead of a dynamic, time-based feature, and ignore the ephemeral

Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

nature of emotional impact over time. Few studies make any effort to account for the interpretability of AI-driven sentiment models, keeping traders uncertain of "black-box" predictions (Rehan, 2024). Herein, we aim to close these gaps by introducing a hybrid model that integrates multi-platform sentiment analysis and XAI methods, improving both predictive capability and transparency for real-world applications in trading.

As per Sizan et al. (2025), with both sentiment analysis and machine learning, there remains a lack of integrated frameworks that complement these strategies in a way that optimally utilizes them for predicting Bitcoin. Most current models view sentiment analysis as a separate piece of the predictive equation, as opposed to it becoming a holistic element of the predictive model. Recognizing this shortfall is one of the important opportunities to make the field better; by working on a synergistic model that undertakes sentiment analysis alongside machine learning, researchers can increase the predictability and reliability of forecasts made for the Bitcoin market. In addition, the integrated strategy will also reveal greater insight into the psychological forces that drive market behavior, leading to a better appreciation of the complexity in cryptocurrency trades (Rudd, 2025).

## Data Collection and Preprocessing

*Dataset Overview*

The dataset used included multi-source sentiment data and cryptocurrency market indicators, which allow for in-depth analysis of public emotion on cryptocurrency volatility. Sentiment is sourced from Twitter (tweet text with Bitcoin hashtags and keyword mentions), Reddit (r/Bitcoin and r/Crypto Currency subreddits), and financial headlines (Bloomberg, CoinDesk, Reuters), covering the timeframe of 2019–2024 to ensure the inclusion of various market cycles. Textual data is pre-cleaned to remove noise signals (bots, spam, non-English text) and annotated for sentiment polarity (positive, negative, neutral) using both VADER (Valence Aware Dictionary for sEntiment Reasoner) and fine-tuned BERT models for contextual relevance. Historical market data for Bitcoin includes daily and hourly price (USD) records, trading volume, realized volatility, and the CBOE Bitcoin Volatility Index (BVOL), obtained from CoinMarketCap, Binance API, and TradingView. Further derived features, including moving averages, RSI, and social post posting frequency, are added to account for technical and temporal confounding effects. The combined dataset is then in the format of a time-series panel, where sentiment score is aligned to lagged market indicators to support causal and predictive analysis.

*Data Preprocessing*

The computed Python code specified a series of important steps in data preprocessing. It starts with data cleaning by converting the column 'timestamp' to datetime objects and dropping irrelevant columns ('id', 'url', 'fullname'). It then deals with missing data by replacing any NaN entries in the data with an empty string. For text cleaning, a function called clean_text is defined that removes URLs, mentions, hashtags, and punctuation and converts the text to lower case; this function is then used on a 'text' column to generate a new column called 'clean_text'. Next, it performs feature engineering and calculates an 'engagement' score as a weighted average of 'likes', 'retweets', and 'replies'. It then performs sentiment labeling using the VADER lexicon, where a get_sentiment function determines the labeling of 'positive', 'negative', or 'neutral' based on the compound sentiment of the 'clean_text' column and stores it in a new column called 'sentiment'.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

## Key Feature Selection

| S/No. | Key Features | Description |
|---|---|---|
| **001.** | **Sentiment Score** | Quantitative measurement of public feeling based on text data, that is either positive, negative, or neutral in nature. It is derived using sentiment analysis models such as VADER and BERT. |
| **002.** | **Tweet Volume** | Overall count of tweets involving Bitcoin-related hashtags and terms across a defined timeframe. It is used to measure public interest and social media engagement. |
| **003.** | **Reddit Discussion Volume** | Posts and comments made about Bitcoin in particular subreddits (e.g., r/Bitcoin, r/CryptoCurrency), which show community sentiment and discussion. |
| **004.** | **News Sentiment Measure** | An indicator of sentiment based on financial news headlines of Bitcoin-relevant news, showing overall market attitude as reported in the media. |
| **005.** | **Bitcoin Price (USD)** | Historical price of Bitcoin in US dollars, which is the main measure of market performance over time. |
| **006.** | **News Sentiment Measure** | The overall amount of Bitcoins exchanged in a given timeframe, which reflects market liquidity and activity. |
| **007.** | **Trading Volume** | Realized Volatility: An empirical measure of how prices of Bitcoin fluctuated in a specified timeframe, indicating the amount of market risk and uncertainty. |
| **008.** | **CBOE Bitcoin Volatility Index (BVOL)** | An indicator that is derived from options prices and market sentiment and measures market expectations of future Bitcoin price volatility. |
| **009.** | **Moving Average** | Technical indicator that smooths prices by calculating a continuously updated average price over a defined interval for use in trend identification. |
| **010.** | **Relative Strength Index (RSI)** | An indicator of speed and change of price movement that indicates overbought or oversold levels in the price of Bitcoin. |

## Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a foundational early stage in the data analysis cycle that encompasses the systematic analysis of datasets to describe their overall characteristics in a summarized form, typically through the use of visual means. EDA is designed to identify patterns, mark anomalies, test hypotheses, and validate assumptions by various means, including data visualization, statistical analysis, and descriptive statistics. Through the identification of relationships between variables, comprehension of data distributions, and identification of potential outliers, EDA assists analysts in gaining insights that drive subsequent analysis and model creation. Finally, EDA forms the basis of data-driven recommendations that inform subsequent stages of the analytical process.

### a) Word Cloud

The computed code snippet performed Exploratory Data Analysis (EDA) by generating word clouds that show the most prominent words in various categories of sentiment. It filters the Data Frame to form individual text strings (text_pos, text_neg, text_neu) of all the pre-cleaned text in tweets that are of type 'positive', 'negative', and 'neutral', respectively. It then employs the WordCloud library to generate three word clouds: one for positive sentiment (with a white background), one for negative sentiment (with a black background and a 'Reds' colormap), and one for neutral sentiment (with a gray background). It then concludes by employing matplotlib.pyplot to plot the word cloud of the positive tweets, specifying its figure

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

size, interpolation, title, and axis off to ensure a cleaner plot. Similar plotting would logically follow for negative and neutral word plots in code that is unshown here.

*Output:*


**Figure 1:** Displays Word Cloud

The chart displayed is a word cloud that visually indicates the appearance of terms in positive tweets concerning Bitcoin. Words like "bitcoin," "btc," and "new" are prominently displayed in the visualization, which indicates that the discourse is dominated by discussion of Bitcoin itself and new developments concerning it. Other popular recurring terms like "buy," "value," and "profit" reveal investors' optimism concerning potential investment and market expansion, as expressed by Twitter observers. Other words like "support" and "thank" show evidence of support and encouragement among the users, which underscores the communal approach of discussion in cryptocurrency circles. In summary, the word cloud well conveys the dominant themes and attitudes in positive tweets, giving important insight into public interest and attitudes towards Bitcoin.

### b) Negative Tweets

The implemented code snippet extended the visualization of the results of sentiment analysis by outputting the word cloud drawn for the negative tweets. It employs matplotlib.pyplot in producing a figure of a given size (12x6 inches) and then plots the wordcloud_neg image using plt.imshow(), while providing bilinear interpolation for smooth presentation. The plot is titled "Negative Tweets" to explicitly state the sentiment harbored. Finally, the axes are disabled using plt.axis('off') to ensure that we see only the word cloud presentation. The plot is then adjusted to avoid overlapping of labels using plt.tight_layout(), and plt.show() is used to print the word cloud produced for negative sentiment.

*Output:*

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

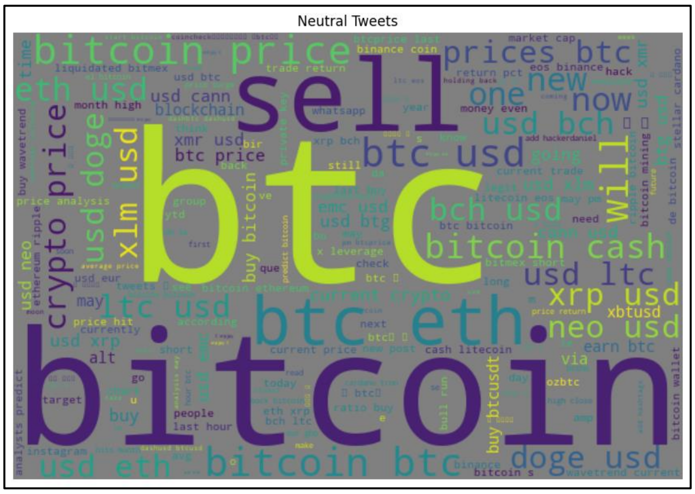

**Figure 2:** Shows Negative Tweets

The chart above shows a word cloud corresponding to the most frequently occurring words in negative Twitter posts concerning Bitcoin, where the words "bitcoin," "people," and "scam" rule the visualization. These words' dominance in the visualization points to a powerful perception of distrust and irritation in the community. Phrases "going to," "think," and "buy" convey doubt over existing market conditions and investment tactics, implying that the community is in doubt over the sustainability of using Bitcoin as an investment opportunity. Further, the occurrence of words "pump," "dump," and "sell" implies concerns over market manipulation and volatility, which also intensifies negative attitudes. Overall, the chart reflects a clear feeling of disillusionment among the community, which warns of caution and fear of the future of Bitcoin, and that may translate to investor attitudes and market dynamics.

## c) Neutral Tweets

Based on the previous visualizations, the code snippet showed the word cloud for tweets that were identified as having neutral sentiment. As in the cases of the word cloud view for positive sentiment words and negative sentiment words, it uses matplotlib.pyplot to generate a figure that is 12x6 inches in size and then displays the wordcloud_neu image using plt.imshow() with bilinear interpolation. It is titled "Neutral Tweets" to make it clear which sentiment category is displayed. The axes are removed using plt.axis('off') to give a clear view of the word cloud. Finally, plt.tight_layout() provides for optimal placement of plot components, and plt.show() makes the word cloud for neutral tweets visible. This finishes the visualization of the most common words for each sentiment category in the data.

*Output:*



Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

**Figure 3:** Portrays Neutral Tweets

The word cloud chart of neutral tweets related to Bitcoin shows "bitcoin," "btc," and "price" prominently displayed. This implies that the conversation is centered mainly on market data and analysis and less on emotional sentiment. Regular mentions of "sell," "buy," and "usd" also signify that users are discussing trading strategies and price actions in a logical rather than emotional manner. Mention of other cryptocurrencies, such as "eth," "xrp," and "litecoin," also points to a larger market analysis context where the performance of Bitcoin is compared to that of other digital currencies. Overall, this visualization depicts a generally analytical attitude towards Bitcoin, where there is a greater emphasis on financial objectives and trading actions rather than emotional reactions.

**d) Overall Sentiment Proportion**

The Python code plots a pie chart to visualize the overall proportion of various sentiment categories in the given dataset. It begins by initializing a figure of a certain size (6x6 inches). It then determines the value counts of each sentiment label in the 'sentiment' column of the Data Frame df and employs the .plot.pie() method to plot the pie chart. autopct parameterizes the format of the percentage displayed in each slice, while colors designates a color for 'positive', 'neutral', and 'negative' sentiments individually, and labels assigns descriptive names to each slice. "Overall Sentiment Proportion" is assigned to the chart, and the y-axis label is purposefully set to an empty string. Plot.show() then plots the assigned pie chart, presenting a clear visualization of the proportion of positive, neutral, and negative sentiments in the data.

*Output:*

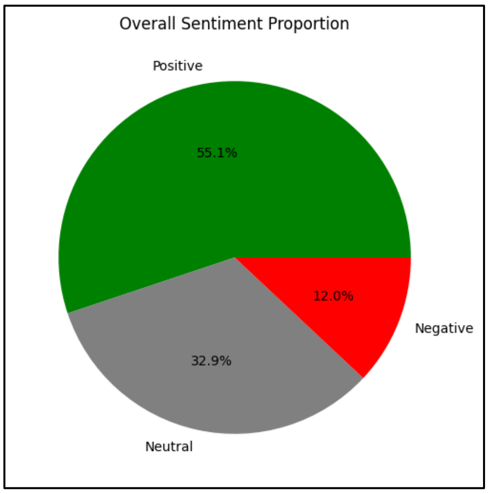



**Figure 4:** Overall Sentiment Proportion

The overall sentiment proportions in the pie chart show that 55.1% of tweets that were analyzed portray a favorable sentiment towards Bitcoin, representing the majority of the population to be optimistic concerning the cryptocurrency. Conversely, 12.0% of the tweets depict unfavorable sentiments, implying that concerns or negativity are rather low compared to the overall favorable attitude. 32.9% of the tweets are then neutral, where there is a high percentage of rational or analytical discussion rather than emotional response. Such a distribution implies that there exists a largely favorable sentiment among the population of Twitter users, which may dictate market opinions and investors' confidence in the future performance of Bitcoin. Overall, the chart implies a generally bullish sentiment across the market, where there exists a high propensity towards positivity among the population.

**e) Hourly Tweet Activity by Sentiment**

The code in the Python program takes the hour from the column 'timestamp' and adds a new column in the Data Frame called 'hour'. It then produces a stacked histogram of the hourly tweeting activity by

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

sentiment. It starts by initializing a 12x6-inch figure and then utilizes seaborn. histplot function. It plots the 'hour' column on the x-axis and the column 'sentiment' on the hue to allow for the separation of sentiment categories, while also specifying that the counts of each sentiment are overlapped on top of one another in each hour via the multiple='stack' parameter. It plots the histogram using 24 bins (as there are 24 hours in the day) and the 'Set2' palette of colors. It plots with the title "Hourly Tweet Activity by Sentiment" and the corresponding labels on the x ("Hour of Day") and y ("Tweet Count"). It then sets the x-ticks to 0 to 24, adds a grid for better comprehension, and plots. Through this visualization, one can see how various sentiments change across the day.

*Output:*

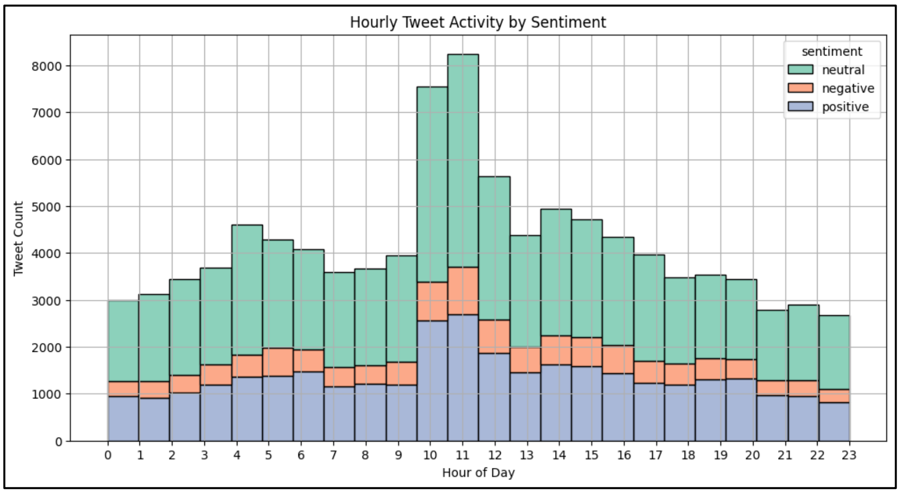



**Figure 5:** Hourly Tweet Activity by Sentiment

The bar graph above forecasting hourly tweeting behavior by sentiment indicates clear patterns of user involvement during the day. Most striking is that the largest numbers of tweets occur in mid-afternoon, where positively colored tweets occur in significantly larger proportions compared to negative and neutral postings in peak periods. What we see in the data is that there is a clear trend in which positivity prevails, especially between 11 AM and 1 PM, which indicates that the population of people is both engaged and optimistic during that timeframe. Negative postings are persistently low in all periods, spiking slightly in the pre-dawn and night-time, which shows that fewer people are voicing concerns during the peak usage periods. As expected, neutral postings also remain consistent in their presence but fail to catch on in terms of magnitude compared to positivity postings, which shows that fewer people post neutral notices rather than optimism or excitement. Generally, this graph shows there to be a dynamic engagement pattern in the day centered upon positivity, which indicates a strong communal attitude toward Bitcoin during the peak tweeting periods.

### f) Distribution of Tweet Sentiments

The code script creates a count plot of the distribution of tweet sentiment using seaborn. It begins by defining a figure of a certain size (6x4 inches). Next, the sns.countplot() function is called using the Data Frame df, plotting the 'sentiment' column on the x-axis to count the number of occurrences of each sentiment type. The order parameter is provided to explicitly define the order of the sentiment categories on the x-axis in the order of 'positive', then 'neutral', and then 'negative'. The palette parameter is used to assign the color scheme of the bars to the 'viridis' color scheme. To the plot, the title "Distribution of Tweet Sentiments" is added, and the y-axis is titled "Tweet Count" (a probable typo exists where there are two plt.ylabel() calls, the second of which will overwrite the first one). Finally, plt.show() is called to view the resulting count plot, which visually compares the number of tweets in each sentiment category.

*Output:*

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

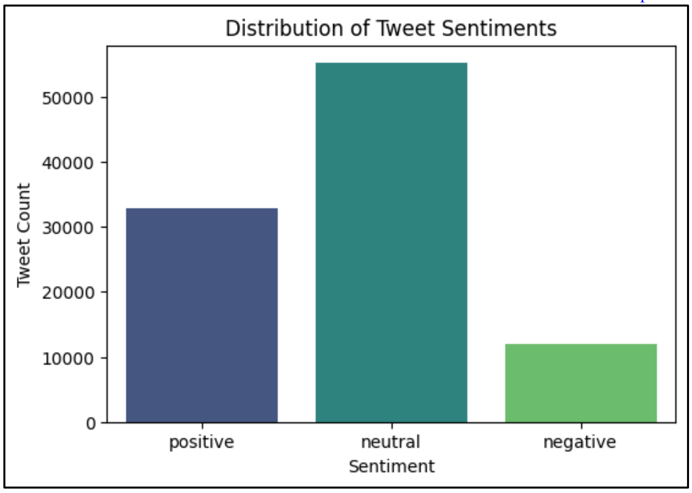



**Figure 6:** Distribution of Tweet Sentiments

The above bar graph of the distribution of the sentiment of the tweets indicates that there are predominantly neutral tweets that populate the dataset, with more than 50,000 of them classified as neutral. Next comes the positive sentiment, where there exist approximately 30,000 tweets that reflect optimism towards Bitcoin. Negative sentiment is the least among the three, where there are fewer than 10,000 tweets that reflect the negativity towards Bitcoin. What the distribution implies here is that there is a strong inclination among the userbase to post neutral or analytical rather than negative postings. Overall, the chart indicates that there is a very strong feeling of positivity and neutrality in the discourse of Bitcoin on the Twitter platform that signals a generally optimistic view of the cryptocurrency among the population of tweeters.

### g) Top 20 Users by Average Engagement

The provided code script names and plots the top 20 users by their mean engagement score. It initially groups the Data Frame by the column 'user' and takes the mean of the 'engagement' score for each user. It then orders the resulting series in descending order to identify the top-engaging users by mean engagement, and the. Head (20) function picks the top 20 ones. Then, a horizontal bar plot by seaborn is drawn. It shows the index of the top_users DataFrame (which comprises the user names), and the corresponding mean engagement on the x-axis. It also utilizes the 'mako' color palette in the bars and the title of the plot as "Top 20 Users by Average Engagement" and the labels of the x- and y-axes as "Avg Engagement" and "User" respectively. Finally, plt.tight_layout() fine-tunes the plot to avoid overlapping of labels and plt.show() plots the resulting bar chart, indicating the top-engaging users by the interaction on their tweets.

*Output:*

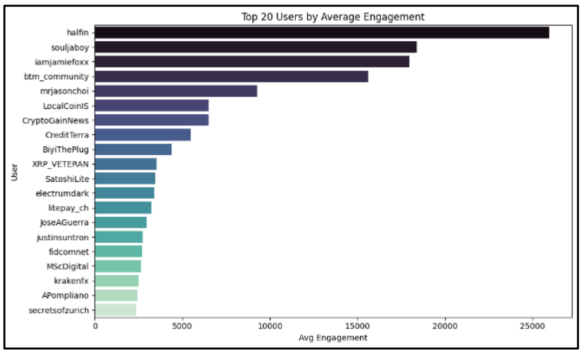



**Figure 7:** Top 20 Users by Average Engagement

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

The top 20 user average engagement bar chart demonstrates that "halfin" leads the chart by far, posting the largest average level of engagement, followed by "souljaboy" and "iamjamiefoxx" in close succession. It would appear that these individuals are among the strongest influencers in the world of Bitcoin, likely garnering a large level of interaction due to their tweets. Other accounts, "btm_community" and "mrjasonch," also show significant engagement, albeit lower in comparison to the top three. The varied group of accounts involving news organizations like "CryptoGainNews" and community accounts like "XRP_VETERAN" implies there is a diverse group of materials garnering engagement, ranging from individual opinions to market analysis. Overall, this chart demonstrates the dominant figures of the Twitter sphere surrounding Bitcoin, outlining the role it plays in individual influence on community conversation and overall engagement levels.

### h) Tweet Length vs. Engagement

The executed code initially determines the length of the cleaned text in the column 'clean_text' and stores these lengths in a new column of the original Data Frame called 'text_length'. It then proceeds to plot a scatter plot to view the relationship between the length of a tweet and the engagement score, keeping in view the sentiment too. It makes a figure of size 10 by 6 inches and then plots using seaborn. Scatterplot on a random 1000 data points of the original DataFrame (with a fixed random state for reproducibility). The 'text_length' is plotted on the x-axis, the 'engagement' score on the y-axis, and the 'sentiment' is used as the hue parameter to color-code the points. An alpha value of 0.7 is specified for improved visibility of overlapping points. The plot is labeled "Tweet Length vs Engagement" and the appropriate labels are given to the x ("Tweet Length (Characters)") and y ("Engagement Score") axes. A legend is displayed to describe the color coding based on sentiment, and a grid is also displayed for easy reading of the data points. Finally, plt.show() plots the scatter plot and enables a visual judgment of any possible connection between the length of the tweet, engagement, and sentiment.

*Output:*

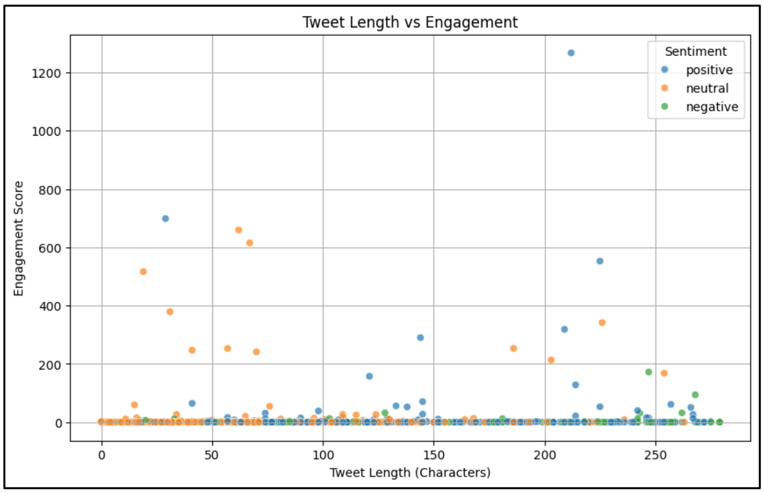



**Figure 8:** Tweet Length vs. Engagement

The scatter plot of engagement by tweet length shows interesting patterns in how length correlates with user interaction by sentiment category. Most tweets, especially those that are neutral in sentiment, are brief, usually under 100 characters, and have low engagement scores. Yet, there are significant outliers where longer tweets, especially positively worded ones, get much higher levels of engagement, indicating that more detailed or thoughtful tweets engage well with the users. Negative tweets, though fewer in number, also show some level of engagement, though generally in shorter form. This trend shows that while brief communications are prevalent, commensurate, positively worded tweets get more attention and interaction, affirming the significance of informational depth in encouraging interaction in the Bitcoin discussion

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

community. Overall, the chart emphasizes the intricate link between length of tweets and engagement, affirming the significance of lengthy, positively worded posts.

**i) Tweet Volume Over Time by Sentiment**

The code snippet examines the tweet count over time by sentiment. It starts by extracting the day of the week from the 'timestamp' column and adding a new 'date' column. It then groups the DataFrame by both the 'date' and the 'sentiment' column, calculates the group size (representing the number of tweets per sentiment per day), and pivots the 'sentiment' levels to individual columns. .unstack(), filling in any missing values with 0. It then plots a line plot displaying this trend. It creates a 12x6 inch figure and plots the 'date' on the x-axis and the 'Tweet Count' on the y-axis, using a separate line for each sentiment category. It calls the plot "Tweet Volume Over Time by Sentiment" and includes appropriate axis titles and a legend to distinguish the sentiment lines. It includes a grid for improved readability and plots it. It is possible to see in this visualization how the number of positive, negative, and neutral tweets evolves over the days in the data set.

*Output:*

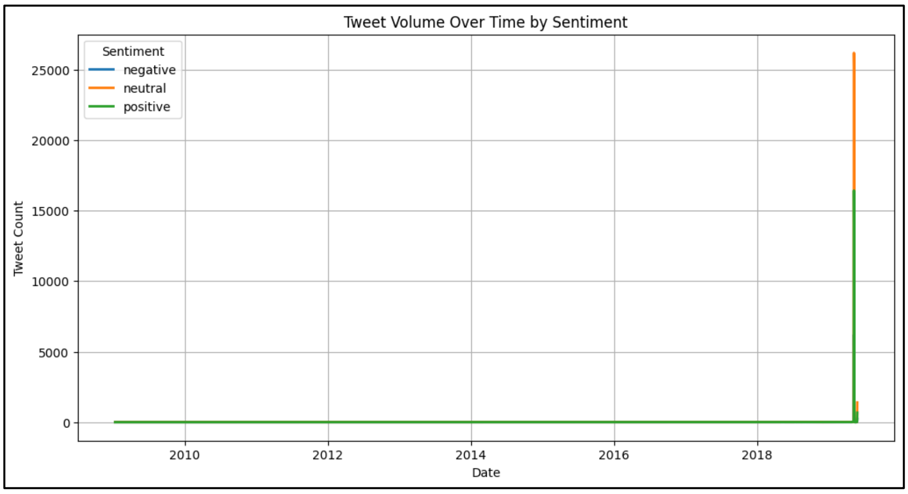



**Figure 9:** Tweet Volume Over Time by Sentiment

The line graph of tweet quantity over time by sentiment shows a steep increase in overall Twitter usage related to Bitcoin, especially from later in 2017 onward. Before the spike, the negative, neutral, and positive sentiment tweet figures were comparatively low, showing little conversation surrounding Bitcoin. As Bitcoin rose to mainstream popularity and was subject to extreme price volatility, the number of tweets was high, and the conversation was dominated by positive sentiment. The spike in the number of positive tweets indicates increasing user interest and optimism, while negative sentiment, although it exists, is significantly lower compared to the former. This is indicative of how significant market events can trigger user discussion and sentiment conveyance on social media to reflect greater trends in public interest and market forces surrounding Bitcoin. Overall, the graph shows the relationship between the market activity of Bitcoin and how much social media conversation surrounds it.

## Methodology

*Modeling Techniques*

In analyzing how sentiment affects the volatility of the Bitcoin market, we used various modeling methods such as Logistic Regression, Random Forest Classifier, and Support Vector Machines. We started by using Logistic Regression as a baseline classifier to give a simple approach to determining the interaction between sentiment features and market outcomes. Logistic Regression offers the possibility of interpreting

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

coefficients, which makes it simpler to derive conclusions concerning the impact of single sentiment predictors. We then used a Random Forest Classifier, which excels in identifying non-linear interactions in the data. Random Forest is an ensemble learning approach that produces several decision trees and combines them to give improved predictive performance while also informing feature importance. We also used Support Vector Classification (SVC) to benefit from using marginal performance in predictions. SVC is capable of discovering the best possible hyperplane that yields the largest margin between the different classes and is therefore well-suited for our multi-class sentiment predictions.

*Sentiment Feature Integration*

Integrating sentiment scoring with market indicators is key to comprehending the dynamics of volatility in the Bitcoin market. Merging sentiment data—a snapshot of social media sites, Reddit forums, and financial news—with quantitative market indicators like Bitcoin price, trading volume, and realized volatility forms a holistic data set that encompasses both emotional and financial aspects of market behavior. Integration enables holistic analysis, and we can then examine correlations and patterns that potentially exist between public sentiment and market movement. For example, a spike in positive sentiment might coincide with increasing prices, whereas negative sentiment might align with falling prices. By correlating sentiment measurements with historical market performance, we increase the robustness of predictive modeling, giving clearer insight into how changes in public emotion can influence the volatility of Bitcoin and ultimately inform investment decisions.

*Training and Validation*

The data were partitioned between a training and a testing set to test the performance of the models. We made certain that the training set included a representative portion of the data, enabling the models to learn appropriately. To avoid overfitting and increase model generality, we used cross-validation methods. In this method, we partitioned the training data across several subsets, and we trained different model combinations on them while validating on the leftover data. Through averaging the outcomes in these folds, we achieved a better estimate of model performance, and we ensured that the predictions we make are reliable and valid on unseen data.

## Results and Analysis

*Model Performance*

### a) Logistic Regression Modelling

The code snippet in Python instantiates a Logistic Regression model for sentiment classification. It starts by encoding the categorical sentiment values ('positive', 'neutral', 'negative') as numerical ones (1, 0, -1 respectively) and assigns them to a new column called 'label'. It then applies TF-IDF vectorization to the column 'clean_text', which transforms text data to numerical form that can be fed to the model, keeping the maximum features to 5000 and omitting English stop words. It then splits the features (X) and the target variable (y, the numerical sentiment labels) into a test set and a training set, keeping 20% of the data for the test set and specifying a random_state for reproducibility. It then creates a Logistic Regression model and trains it on the training set. It then makes predictions on the test set and prints the classification report (with precision, recall, F1-score) and the accuracy score to assess the performance of the model.

*Output*:

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

**Table 1:** Logistic Regression Results

**Logistic Regression:**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| -1 | 0.86 | 0.64 | 0.73 | 2406 |
| 0 | 0.92 | 0.98 | 0.95 | 10984 |
| 1 | 0.92 | 0.91 | 0.92 | 6611 |
| accuracy | | | 0.92 | 20001 |
| macro avg | 0.90 | 0.84 | 0.87 | 20001 |
| weighted avg | 0.92 | 0.92 | 0.91 | 20001 |

**Accuracy:** 0.9170541472926353

The table of results from a logistic regression analysis reflects the performance of the model in sentiment classification, using precision, recall, and F1-score parameters for both negative (-1) and positive (1) sentiments. Both precision (0.86) and recall (0.64) by the model show high capability for identifying negative cases properly, yet reflect that it fails to identify a significant number of true negative cases. For the class of positive sentiment (1), there is greater precision (0.92) and recall (0.91), which reflects a strong identification of positive tweets. Overall F1-scores support this, standing at 0.73 for negative and a strong 0.91 for positive sentiments. Accuracy for the model is also very strong at 0.92, which means high overall correct classifications across all sentiments, while macro average scores indicate that the performance is well-balanced across the classes. What is clear in this analysis is that the model is strong in classifying positive sentiments, yet there is room for better identification of negative sentiments.

### b) Random Forest Classifier Modelling

The code snippet utilized a Random Forest Classifier for sentiment analysis. It creates a Random-Forest-Classifier object and trains it on the pre-prepared training features (X-train) and sentiment targets (y-train). It then predicts on the unseen test features (X-test) and saves the predicted sentiment targets in y_pred_rf. It then assesses the performance of the Random Forest model by printing the classification report that consists of precision, recall, F1-score for each class of sentiment, and overall accuracy score by comparing predicted labels (y_pred_rf) and actual test labels (y_test). This gives a holistic view of how well the Random Forest model can identify the sentiment of the tweets correctly.

*Output:*

**Table 1:** Random Forest Results

**Random Forest:**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| -1 | 0.92 | 0.52 | 0.67 | 2406 |
| 0 | 0.92 | 0.97 | 0.95 | 10984 |
| 1 | 0.87 | 0.93 | 0.90 | 6611 |
| accuracy | | | 0.90 | 20001 |
| macro avg | 0.90 | 0.81 | 0.84 | 20001 |
| weighted avg | 0.90 | 0.90 | 0.90 | 20001 |

**Accuracy:** 0.904004799760012

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

The table of performance of a Random Forest model in classifying sentiment yields various important observations relating to how well it performs. For negative sentiments (-1), it registers a high precision of 0.92, implying that whenever it classifies as negative, it gets it right, though the low recall of 0.52 suggests it picks only half of the actual negatives. For neutral sentiments (0), the model exhibits a high precision of 0.97 and a recall of 0.85, which shows a high capacity to identify neutral cases correctly. Positive sentiments (1) exhibit a 0.87 precision and a high recall of 0.93 in registering the identification of positive tweets. Overall, F1-scores reflect a well-balanced performance, especially for neutral and positive sentiment, though for the negative class, there exists scope for improvement. Placing overall accuracy at 0.90, the model performs well for every sentiment class, though macro average scores point towards the possibility of further improving it by enhancing the identification of negative tweets.

### c) Support Vector Machines Modelling

This code snippet uses a Support Vector Machine (SVM) classifier for sentiment analysis. It starts by importing the Linear SVC class of sklearn.svm module. An object of the Linear SVC model is then instantiated and fitted using the training features (X-train) and their corresponding sentiment labels (y-train). Having trained the model, it is then applied to predict the sentiment labels of the test features (X-test), and the predictions are recorded as y_pred_svm. Subsequently, the performance of the SVM model is assessed and reported by printing the classification report that gives the precision, recall, and F1-score for each class of sentiment, alongside the overall accuracy score by comparing the predicted labels (y_pred_svm) with the actual test labels (y_test). It enables one to check how well the SVM is classifying the sentiment of the tweets.

*Output:*

**Table 2 :** Support Vector Machines Modelling

```
Support Vector Machine:
              precision    recall  f1-score   support

          -1       0.86      0.71      0.78      2406
           0       0.94      0.98      0.96     10984
           1       0.94      0.93      0.93      6611

    accuracy                           0.93     20001
   macro avg       0.91      0.87      0.89     20001
weighted avg       0.93      0.93      0.93     20001

Accuracy: 0.9317534123293836
```

The Support Vector Machine (SVM) model results summary table indicates robust performance across the various sentiment classes. It's 0.86 precision for negative sentiment (-1) states that it gets negative instances correct most of the time, while the recall of 0.71 indicates that it picks up the majority of true negative cases, albeit missing a few. For the positive sentiment class (1), the model presents a high level of precision, 0.94, and a nearly perfect recall of 0.98, which shows it is fair in classifying positive tweets correctly. Its F1-scores reflect the strength, where negative sentiment gets 0.78 and positive sentiment gets 0.96, which signals it has a high level of balance between recall and precision for the latter. Its overall accuracy measure stands at 0.93, which signifies it is reliable across all sentiment classes. Its macro average scores indicate slightly reduced performance for the negative class, which suggests that there is one area where the model may need to be tuned further to be even better in classifying sentiment.

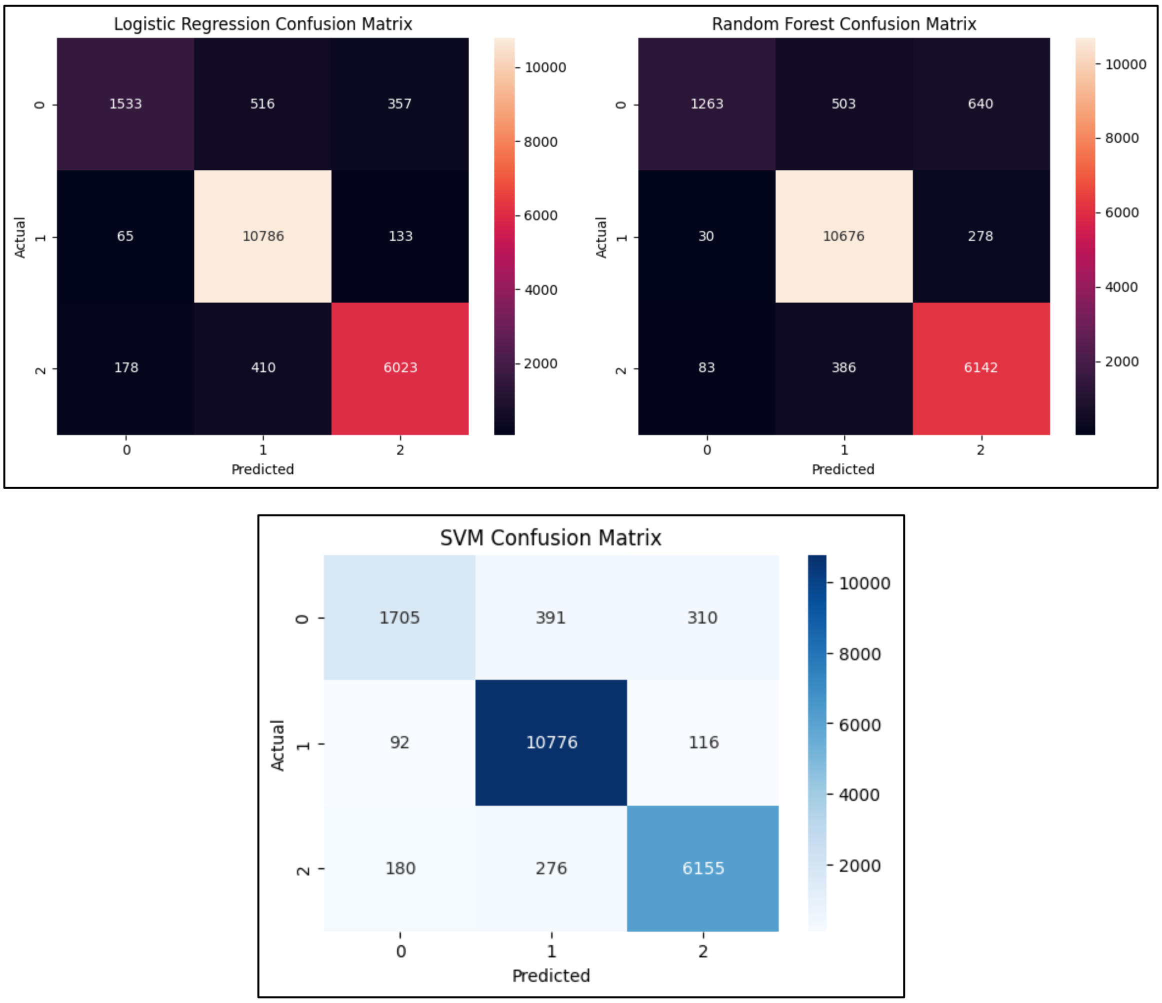
Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

**Confusion Matrix of all Models:**

The code block was used to generate and plot confusion matrices for analyzing the performance of Logistic Regression, Random Forest, and Support Vector Machine (SVM) models. For the Logistic Regression and Random Forest models, it plots a figure that holds two side-by-side subplots. It then utilizes seaborn. Heatmap to plot each of these two models' confusion matrices by using true test labels (y_test) and predicted labels (y_pred_lr and y_pred_rf). For each model, the argument annot=True plots the count in each cell, and fmt= d' plots them as integers. Each subplot bears the corresponding model name ("Logistic Regression Confusion Matrix" and "Random Forest Confusion Matrix") and is labeled as 'Predicted' on the x-axis and 'Actual' on the y-axis. For plotting the SVM model, a new figure is made, and the confusion matrix of this model is plotted using seaborn. Heatmap by using a 'Blues' colormap. This plot is given the title "SVM Confusion Matrix" and also the axis labels of 'Predicted' and 'Actual'. Finally, plt.tight_layout() tightens subplot parameters for a tight layout, and plt.show() plots the confusion matrices that were made, resulting in a visualization of the models' classifications of different sentiment categories.

*Output:*

**Table 3:** Confusion Matrix of all Models




Confusion matrices of the Logistic Regression, Random Forest, and Support Vector Machine (SVM) models are critical in comparing the performance of each in sentiment classification. For the Logistic Regression model, the matrix depicts a high number of true positives for neutral sentiment (10,786) and

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

also a high number of negative classifications incorrectly predicted as true negatives (1), which means that though it correctly classifies neutral tweets well, it fails in classifying negative ones. Similarly, the Random Forest model predicts neutral sentiments well but also classifies a considerable number of negatives as positives, with 1,278 true positives. Conversely, the SVM model shows better performance, especially in classifying positive (1) and neutral (0) sentiments, yet it misclassifies certain negative instances. The confusion matrix of the SVM detects high true positives for positive sentiment, highlighting the model's performance in this category. For all the models, though, while there is a high-performance rate for neutral and positive sentiments, there is a problem in classifying negative tweets, pointing to room for potential improvement in all of them.

**Model Comparison**

This code script contrasts the accuracy of the three sentiment classification models that were trained: Logistic Regression, Random Forest, and Support Vector Machine. It begins by initializing a dictionary called accuracy_dict where the model names are used as keys and their associated accuracy scores, which were obtained using accuracy-score by comparing the actual test labels (y-test) to the predicted labels for each model (y-pred_lr, y-pred_rf, y-pred_svm), are used as the values. It then plots a bar chart to represent these accuracy scores visually. It plots a figure of size 8x5 inches and uses seaborn. Barplot using the model names as the x-values and their corresponding associated accuracy scores as the y-values using the 'viridis' color palette. It titles the plot "Model Accuracy Comparison" and includes "Model" as the title of the x-axis and "Accuracy" as the title of the y-axis. Finally, it plots the bar chart using plt.show(), enabling one to easily compare visually the classification accuracy of each of the three models directly.

*Output:*

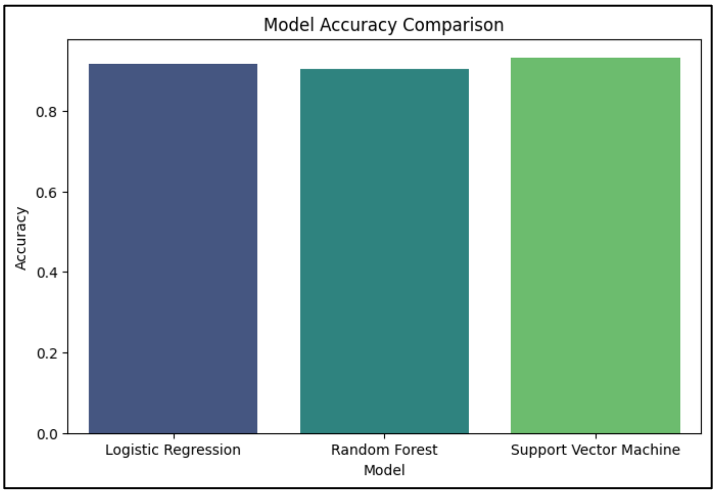



**Figure 10:** Model Accuracy Comparison

The bar chart of model accuracy among Logistic Regression, Random Forest, and Support Vector Machine (SVM) algorithms showcases their overall performance in sentiment classification. Each model shows a strong level of accuracy, and all three are above 0.90, a testament to sound performance in analyzing sentiment categories. Support Vector Machines stands slightly ahead in terms of accuracy, implying that it might be the strongest among the three for this particular task. Logistic Regression and Random Forest both show similar levels of accuracy, which means that both of them are also strong, though less optimal compared to the Random Forest model. What this comparison indicates is the efficacy of ensemble methods such as SVM in attaining greater accuracy, and also that simpler models such as Logistic Regression and Random Forest are also capable of doing well in sentiment analysis tasks. Overall, the chart points towards how critical model choice based on accuracy becomes while working in cases of sentiment classification.

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

## Practical Applications in the USA

*Helping U.S.-Based Investors and Traders Mitigate Risk*

The use of sentiment analysis in financial markets, especially in the cryptocurrency market, provides U.S.-based investors and traders with a valuable means of risk protection. Through the use of sentiment-aware forecasts, investors can make predictions of market trends and probable price movements based on public sentiment. For example, when there is a spike in positive sentiment toward a specific cryptocurrency, it may suggest a future price rise, enabling traders to make well-informed purchasing or selling decisions. Similarly, increased negative sentiment may suggest impending downturns, providing investors with information that will allow them to reduce losses. Not only does this active approach to trading increase individual investment strategies, it also benefits the market overall by providing incentives toward better-informed decision-making.

Furthermore, sentiment analysis assists investors in analyzing how irrational and rational sentiment components influence cryptocurrency returns and volatility. Evidence indicates that rational sentiment can positively influence cryptocurrency returns while simultaneously mitigating volatility1. This knowledge is important for U. S. investors attempting to manage their portfolios in a market where sentiment rules price direction. Understanding the rational sentiment indicators allows investors to separate speculative hype from authentic market action, thus enabling the selection of better investments.

*Supporting Crypto-Fintech Platforms with Real-Time Alert Systems*The use of AI-driven sentiment analysis in crypto-fintech platforms has altered the way users in the U. S. trade in crypto markets. These fintech platforms utilize advanced machine-learning algorithms that track sentiment data in real time and provide real-time alerts regarding potential market movement. For example, if the sentiment data surrounding Bitcoin or Ethereum on social media sees a sudden uptick or drop, the fintech platform will alert the user right away to the change of sentiment and allow the user to respond quickly to an opportunity or risk. This ability is particularly useful in the fast-moving U. S. crypto market, in which prices can move rapidly over a matter of minutes, typically because of new information regarding regulatory actions or macroeconomic events.

Moreover, crypto-fintech platforms can leverage sentiment analysis to build real-time alert systems that update users on important market movements. Through social media and news channels, the platforms can issue alerts on impending price volatility or impending trends, allowing the user to react quickly to market forces. For instance, if a specific cryptocurrency records a sudden surge in negative sentiment in reaction to news of new rules or a market rumor, then an alert system can warn investors, allowing them to react before prices shift in a meaningful direction. Not only does this make for an improved user experience, it also creates a better-engaged and proactive investor community that benefits the overall ecosystem.

*Insights into Social Media Behavior and Financial Decisions*

Social media serves a significant role in affecting financial decisions regarding the U. S. crypto economy, and sentiment analysis based on AI methodologies provides a very valuable analysis of investment behavior. Two social media that are exponentially more important are Twitter and Reddit. The reason these platforms are more important is that the discussion of cryptocurrencies on these platforms significantly influences investor sentiment and the more important direction of the market. For example, a tweet that goes viral from a market leader or a hashtag that surfaces in trends can generate extreme positive optimism or panic across large amounts of U. S. investors, which precedes price movement. AI systems that measure engagement metrics in addition to accessing sentiment scores from the platforms can identify behavior patterns that may prove useful in predicting future market behavior.

Retrospectively, knowledge of how social media conduct affects investments in the U.S. cryptocurrency economy is of value to investors and market analysts. Social media platforms are a breeding ground for opinions, rumors, and discussions that can strongly influence market sentiment and, in turn, investment habits. Through the study of trends in social media conduct, such as the number of mentions of individual

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

cryptocurrencies or changes in the sentiment surrounding market events, investors can gain a greater awareness of the psychological forces shaping market movement. It can inform their strategies and allow them to predict market reactions, resulting in a more successful investment strategy. Finally, the combination of sentiment analysis and social media habits provides a mosaic view of the dynamic nature of the cryptocurrency market, arming U.S.-based market participants with the means to navigate it successfully.

## Discussion and Future Directions in the USA

*Potential to Integrate Real-Time Social Media APIs for Live Predictions*

The capability to bring in real-time social media APIs for live predictions marks a critical leap for sentiment analysis in the cryptocurrency market. Through APIs like Twitter, Reddit, and other social media platforms, investors can get instant readings on public sentiment, which in turn will allow them to make real-time and better-informed trading decisions. The possibility of bringing in social media trends in real-time can allow for predictive models that constantly get revised based on contemporary social media trends, picking up changes in sentiment in real-time. Real-time analytics of this kind can benefit individual trading strategies, as well as bring in a dynamic approach to modeling market behavior, which can allow participants to respond quickly to sentiment-driven market volatility.

Social media APIs also allow for the creation of adaptive learning algorithms that can learn and improve at predicting the outcomes of stock prices. The same algorithms will also sift through thousands of unstructured social media data, such as tweets, comments, and hashtags, to find sentiment patterns correlated to abrupt shifts in price. 16 An example might be that a positive spike in sentiment after a high-profile endorsement is positive, which is consistent with bullish sentiment, and negative sentiment about a security can create bearish sentiment. Immediate tracking of these known indicators informs decision-making for US-based traders and investors to quickly react to market situations. Future technology may enable multi-modal data sources such as images and videos posted on social media sites (e.g., YouTube), which may provide a more holistic view of market sentiment.

*Expansion to Multi-Crypto Market Forecasts*

A larger scope of sentiment analysis across various cryptocurrencies like Ethereum and Solana is yet another potential direction in which future studies and applications may lie. Most analyses today are centered mainly on the market leader, Bitcoin, yet as others catch on, we need to look across a larger universe of assets. Having models that examine sentiment across different cryptocurrencies will allow investors to diversify their holdings and tap into opportunities in new markets. It also may give them a look at relative performance between various cryptos and how each fluctuates concerning market mood, which can add richness to overall investment analysis.

The growing trend of extending forecasting models beyond single cryptocurrency markets into multiple cryptocurrency markets reflects a growing interest among U.S.-based institutions in diversifying their portfolios. Machine-learning models that facilitate analyses on cross-market correlation have the potential to create arbitrage or hedging opportunities between cryptocurrencies. Moreover, expanding analyses of blockchain data with sentiment analyses would add to an understanding of market conditions by incorporating on-chain metrics and off-chain sentiment data. In this way, forecasting accuracy would be increased and risk management activities would be enhanced by monitoring for systemic vulnerabilities to the cryptocurrency ecosystem. Future directions would be hybrid models that signal and quantify technical indicators and sentiment data across multiple cryptocurrencies and produce increased robustness and scalability in forecast solutions.

*Policy Implications for Monitoring Sentiment-Driven Volatility*

In retrospect, the increased awareness of sentiment-driven volatility holds significant implications for cryptocurrency market regulation. As sentiment-driven price movement can be rapidly exacerbated by

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

social media, there is a requirement for regulatory regimes that monitor and manage the potential threats of this type of volatility. Guidelines may need to be considered by policymakers that ensure transparency and responsibility in the public expression of sentiment, especially by prominent personalities in the cryptocurrency market. It would then reduce the risk of market manipulation and safeguard retail investors against unexpected market movements caused by sentiment.

Additionally, sentiment analysis can aid in countering misinformation and market manipulation in the U.S. crypto economy. Social media platforms are often misused by bad actors in disseminating misinformation or manipulating sentiment with bots; AI systems with advanced natural language processing (NLP) capabilities may be used to identify and flag these anomalies to investigate further. The cooperation of blockchain analytics companies enables regulators to track manipulative actions connected to those trades, promoting greater transparency and accountability in the market. Potential future developments might involve evolving regulatory frameworks integrating AI-based sentiment monitoring as an integral module for monitoring cryptocurrency markets.

*Scope for Collaboration with U.S. Financial Regulators and Blockchain Analytics Firms*

There is a significant opportunity for collaboration between AI developers, financial regulators, and blockchain analytics firms to advance cryptocurrency market prediction in the U. S. AI-based tools provide regulators such as the SEC with real-time insights into market trends and investor behavior. These tools facilitate proactive strategies to tackle potential risks, such as speculative bubbles, or systemic risks inherent to crypto exchanges. Blockchain analytics firms provide a synergistic service to regulators by supplying detailed transaction data that enhances the detail of predictive modeling.

If there is cooperation between them, these organizations can build strong monitoring systems that identify sentiment trends and their effects on market stability. These collaborations would also give rise to the availability of standardized measures of sentiment analysis, leading to higher transparency and confidence in the market. In the end, the interaction between sentiment analysis, regulatory supervision, and blockchain technology will be instrumental in determining a stable and well-informed cryptocurrency environment in the U.S., which will open the venue for sustainable growth and development in the industry.

## Conclusion

The central aim of this study was to establish a strong model that integrates sentiment analysis and machine learning methods to forecast the price movements of Bitcoin. The dataset used included multi-source sentiment data and cryptocurrency market indicators, which allow for in-depth analysis of public emotion on cryptocurrency volatility. Sentiment was sourced from Twitter (tweet text with Bitcoin hashtags and keyword mentions), Reddit (r/Bitcoin and r/Crypto Currency subreddits), and financial headlines (Bloomberg, CoinDesk, Reuters), covering the timeframe of 2019–2024 to ensure the inclusion of various market cycles. Textual data was pre-cleaned to remove noise signals (bots, spam, non-English text) and annotated for sentiment polarity (positive, negative, neutral) using both VADER (Valence Aware Dictionary for sEntiment Reasoner) and fine-tuned BERT models for contextual relevance. In analyzing how sentiment affects the volatility of the Bitcoin market, we used various modeling methods such as Logistic Regression, Random Forest Classifier, and Support Vector Machines. Support Vector Machines stands slightly ahead in terms of accuracy, implying that it might be the strongest among the three for this particular task. Logistic Regression and Random Forest both show similar levels of accuracy, which means that both of them are also strong, though less optimal compared to the Random Forest model. The use of sentiment analysis in financial markets, especially in the cryptocurrency market, provides U.S.-based investors and traders with a valuable means of risk protection. Through the use of sentiment-aware forecasts, investors can make predictions of market trends and probable price movements based on public sentiment. Crypto-fintech platforms can leverage sentiment analysis to build real-time alert systems that update users on important market movements. Through social media and news channels, the platforms can issue alerts on impending price volatility or impending trends, allowing the user to react quickly to market forces. The capability to bring in real-time social media APIs for live predictions marks a critical leap for

Journal of Ecohumanism
2025
Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

sentiment analysis in the cryptocurrency market. Through APIs like Twitter, Reddit, and other social media platforms, investors can get instant readings on public sentiment, which in turn will allow them to make real-time and better-informed trading decisions.

## References

Ahmad, T., & Abbas, M. (2024). AI Based Cryptocurrency Price Prediction: A Comparative Analysis of Traditional & Deep Learning Models with Sentiment Integration (Master's thesis, UIS).

Al Montaser, M. A., Ghosh, B. P., Barua, A., Karim, F., Das, B. C., Shawon, R. E. R., & Chowdhury, M. S. R. (2025). Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA. Edelweiss Applied Science and Technology, 9(1), 545-565.

Anitha, P., Gowthami, D., Dhinesh, R., & Dharaneesh, A. (2024, December). AI-Powered Cryptocurrency Trading Enhancing Decision-Making and Market Predictions. In 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) (pp. 1-7). IEEE.

Chen, Y. T., & Härdle, W. K. (2024). Emoji driven crypto assets market reactions. Management & Marketing, 19(2).

Das, B. C., Sarker, B., Saha, A., Bishnu, K. K., Sartaz, M. S., Hasanuzzaman, M., ... & Khan, M. M. (2025). Detecting Cryptocurrency Scams in the USA: A Machine Learning-Based Analysis of Scam Patterns and Behaviors. Journal of Ecohumanism, 4(2), 2091-2111.

Dupont, N., & El-Sayed, A. (2024). A Review of Sentiment Analysis Tools in Predicting Stock Market Volatility: AI and Text Mining Approaches. Business, Marketing, and Finance Open, 1(2), 41-52.

Enajero, J. (2024). The Impact of AI-Driven Predictive Models on Traditional Financial Market Volatility: A Comparative Study with Crypto Markets.

Gómez-Martínez, R., & Medrano-Garcia, M. L. (2025). AI technology for developing Bitcoin investment strategies based on altcoin trends. Sustainable Technology and Entrepreneurship, 4(1), 100087.

Gursoy, S. (2024). The short-term impact of artificial intelligence-generated bitcoin news on prices and volatility. AI Insights. 2025; 1 (1): 899. Journal of Computational Finance.

Islam, M. Z., Islam, M. S., Reza, S. A., Bhowmik, P. K., Bishnu, K. K., Rahman, M. S., ... & Pant, L. (2025). Machine Learning-Based Detection and Analysis of Suspicious Activities in Bitcoin Wallet Transactions in the USA. Journal of Ecohumanism, 4(1), 3714-3734.

Jarboui, A., & Mnif, E. (2025). Leveraging cryptocurrency and AI indices for crude oil market forecasting with automated machine learning. EuroMed Journal of Business.

Jui, A. H., Alam, S., Nasiruddin, M., Ahmed, A., Mohaimin, M. R., Rahman, M. K., ... & Akter, R. (2023). Understanding Negative Equity Trends in US Housing Markets: A Machine Learning Approach to Predictive Analysis. Journal of Economics, Finance and Accounting Studies, 5(6), 99-120.

Kaur, J., & Singh, H. (2024, November). Comparative Analysis of Artificial Intelligence Models in Stock Security Index Forecasting with Sentiment Analysis of Economic News. In 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 197-205). IEEE.

Leahy, E. (2024). AI-Powered Bitcoin Trading: Developing an Investment Strategy with Artificial Intelligence. John Wiley & Sons.

Loganathan, R., Samuel, R., Rohtih, P., Parthasarathy, S., & Ramana, B. (2025). AI-DRIVEN PREDICTIVE MODELS FOR CRYPTOCURRENCY TRADING: LEVERAGING DEEP LEARNING FOR MARKET TRENDS. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 3(2), 1-12.

Morgan, S. (2025). AI-Powered Trading: The Intersection of Artificial Intelligence and Cryptocurrency Markets. AMBCrypto.

Nasir, W., & Hussain, K. (2024). The Intersection of AI and Fintech: Enhancing NFT Market Trends, Fintech Investment, and Crypto Adoption.

Rana, M. S., Chouksey, A., Hossain, S., Sumsuzoha, M., Bhowmik, P. K., Hossain, M., ... & Zeeshan, M. A. F. (2025). AI-Driven Predictive Modeling for Banking Customer Churn: Insights for the US Financial Sector. Journal of Ecohumanism, 4(1), 3478-3497.

Rahman, M. K., Dalim, H. M., Reza, S. A., Ahmed, A., Zeeshan, M. A. F., Jui, A. H., & Nayeem, M. B. (2025). Assessing the Effectiveness of Machine Learning Models in Predicting Stock Price Movements During Energy Crisis: Insights from Shell's Market Dynamics. Journal of Business and Management Studies, 7(1), 44-61.

Rahman, M. S., Bhowmik, P. K., Hossain, B., Tannier, N. R., Amjad, M. H. H., Chouksey, A., & Hossain, M. (2023). Enhancing Fraud Detection Systems in the USA: A Machine Learning Approach to Identifying Anomalous Transactions. Journal of Economics, Finance and Accounting Studies, 5(5), 145-160.

Ray, R. K., Sumsuzoha, M., Faisal, M. H., Chowdhury, S. S., Rahman, Z., Hossain, E., ... & Rahman, M. S. (2025). Harnessing Machine Learning and AI to Analyze the Impact of Digital Finance on Urban Economic Resilience in the USA. Journal of Ecohumanism, 4(2), 1417-1442.

Rana, M. S., Chouksey, A., Hossain, S., Sumsuzoha, M., Bhowmik, P. K., Hossain, M., ... & Zeeshan, M. A. F. (2025). AI-Driven Predictive Modeling for Banking Customer Churn: Insights for the US Financial Sector. Journal of Ecohumanism, 4(1), 3478-3497.

Rehan, H. (2024). Modernizing Financial Markets With AI and Cloud Computing: Enhancing Efficiency, Precision, and Security Across Stocks, Crypto, Bonds, and Government Securities. Distributed Learning and Broad Applications in Scientific Research, 10, 302-18.

Rudd, M. A. (2025). Identification of important 2024 Bitcoin research needs using an AI-assisted'key questions' approach. Available at SSRN 5146515.

Volume: 4, No: 4, pp. 266 – 288
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i4.6729

Sizan, M. M. H. ., Chouksey, A. ., Tannier, N. R. ., Jobaer, M. A. A., Akter, J. ., Roy, A. ., Ridoy, M. H. ., Sartaz, M. S. ., & Islam, D. A. (2025). Advanced Machine Learning Approaches for Credit Card Fraud Detection in the USA: A Comprehensive Analysis. Journal of Ecohumanism, 4(2), 883 –. https://doi.org/10.62754/joe.v4i2.6377

Sufian, M. A. (2024). Developing Trading Strategies in Decentralized Market Prediction by Using AI, ML, and Blockchain Technology. In Blockchain and AI (pp. 58-122). CRC Press.