Journal of Applied Economic Sciences
Volume XX, Fall, Issue 3(89), 2025

Application of Social Media Sentiment Analysis for Developing Trading Models in the Cryptocurrency Market

Andrii TRUSHKOVSKYI
https://orcid.org/0009-0002-8879-138X
Independent Researcher, Los Angeles, CA 90048, USA
andreytrushkovskiy@gmail.com

Article's history:

Received 5th of August, 2025; *Revised* 9th of September, 2025; *Accepted* 16th of September, 2025; *Available online*: 30th of September, 2025. *Published* as article in the Volume XX, Fall, Issue 3(89), 2025.

Suggested citation:

Trushkovskyi, A. (2025). Application of Social Media Sentiment Analysis for Developing Trading Models in the Cryptocurrency Market. *Journal of Applied Economic Sciences*, Volume XX, Fall, 3(89), 535 – 559. https://doi.org/10.57017/jaes.v20.3(89).11

Abstract:

This study examines the predictive role of social media sentiment in forecasting short-term Bitcoin price changes using econometric and machine learning models. Based on Twitter and Reddit data (2020–2025), we construct a daily sentiment index and analyse its lagged effect on returns. OLS regression and advanced models (random forest, XGBoost) show that a one-unit increase in lagged sentiment predicts a statistically significant 0.24–0.25% rise in next-day returns. Controls include momentum, volatility, and trading volume, with Granger causality tests and VAR confirming sentiment's leading role. While volume is insignificant, sentiment and momentum are strong predictors. Machine learning models outperform linear baselines, highlighting nonlinear interactions in sentiment-driven markets. Results validate sentiment as a meaningful input for forecasting, with applications to trading bots, real-time risk dashboards, and supervisory tools. The study contributes to applied economics by showing how quantified investor emotion can serve as a leading indicator in volatile cryptocurrency markets. Future research should consider multilingual sentiment, intraday horizons, and cross-asset extensions.

Keywords: cryptocurrency; bitcoin; sentiment analysis; econometric modelling; machine learning; Twitter.

JEL Classification: G12; G14; G17; C32; C45; C55; G40.

Introduction

Cryptocurrency markets, characterized by their decentralized structure and high-frequency trading environments, have emerged as some of the most volatile financial arenas globally. Bitcoin's price often goes through significant changes in a short amount of time. These changes are often bigger than those in other asset classes (Marthinsen & Gordon, 2022). Fundamentals of the market, like supply, adoption, and regulation changes, play a part. However, there is more proof that psychological and behavioural factors, especially those from online social communities, are the leading causes of investor behaviour and price changes (Han et al., 2022).

Twitter, Reddit, and BitcoinTalk are just a few examples of the social media platforms that have developed into important areas for the real-time dissemination and consumption of information, rumours, and emotional reactions (Shah & Shah, 2024; Vlahavas & Vakali, 2024). In the absence of more conventional market-making mechanisms, these platforms act as discussion centers and sentiment amplifiers with the ability to move markets. Social media sentiment could influence financial markets, as seen in notable instances like the GameStop short squeeze and Dogecoin surges (Andreev et al., 2022; Selvakumar et al., 2025). Public sentiment has predictive power, but many academic and empirical models that deal with how bitcoin prices are determined still focus on historical volumes and prices.

A thorough examination of the current literature uncovers a significant void regarding incorporating social media sentiment into formal econometric modelling frameworks. Rigid, real-time sentiment integration utilizing lagged variables and econometric diagnostics is lacking, even though some studies have investigated sentiment analysis in financial markets or utilized fundamental correlation analyses in cryptocurrency. Stationarity, multicollinearity, and autocorrelation are important statistical aspects that most models fail to account for, which reduces the validity and applicability of their results. Also, compared to traditional baselines, there has not been much effort to determine whether adding sentiment indicators significantly improves model accuracy. This work intends to fill that need by developing and testing an econometric model that uses sentiment data from social media to explain Bitcoin price fluctuations.

From an applied economics perspective, this study goes beyond theoretical investigation by demonstrating how sentiment-informed econometric models can be operationalized in practice. The results are directly relevant to the design of adaptive trading bots that can integrate lagged sentiment signals into buy–sell execution strategies, optimize position sizing in volatile markets, and reduce exposure to sudden reversals. Algorithmic risk management systems can also incorporate these models by adjusting margin requirements, recalibrating stop-loss thresholds, and generating early warning alerts when sentiment and volatility jointly signal systemic stress. For regulators and exchanges, real-time supervisory dashboards informed by sentiment data can improve market oversight by detecting coordinated hype campaigns, identifying abnormal trading behaviours, and flagging periods of heightened systemic risk. By translating behavioural signals from social media into concrete financial tools, the research demonstrates how applied econometric analysis can support practitioners, institutions, and policymakers in managing the risks and opportunities of cryptocurrency markets, thereby bridging the gap between academic modelling and practical implementation.

The primary goal of this study is to determine if lagging sentiment indicators can effectively forecast near-term price movements. This study improves methodological rigor by using strong diagnostics and approaches like ordinary least squares regression. It sheds light on how sentiment-informed trading strategies might be used in risky financial markets. The availability of high-frequency data, the requirement for reproducible model evaluation, and the goal of assessing sentiment's marginal predictive contributions relative to traditional variables like volume and previous returns make quantitative analysis particularly warranted in this context.

This study develops and tests econometric models that integrate social media sentiment to predict Bitcoin price dynamics. The objectives are:

- O1: To examine the correlation between social media sentiment, cryptocurrency price movements, and trading volume.
- O2: To evaluate whether incorporating sentiment indicators improves the predictive accuracy of cryptocurrency price models.
- O3: To compare the effectiveness of sentiment-based trading models with traditional models that exclude sentiment data.

Given the dynamic and noisy nature of the crypto markets, quantitative methods such as OLS regression with lagged variables offer a promising approach to isolating statistically significant patterns amidst volatility.

1. Literature Review

The primary objective of this literature review is to explore the intersection between financial sentiment analysis and cryptocurrency trading models, with a particular emphasis on econometric approaches. Given the growing complexity and behavioural nature of digital asset markets, especially Bitcoin, understanding how investor psychology, as reflected in social media sentiment, influences price dynamics is a timely and underexplored area.

This literature review aims to summarize previous efforts in financial sentiment research, highlight areas where methods are lacking, and show how this study's use of sentiment indicators in predictive econometric models is novel. When it first emerged, financial sentiment research was primarily used in more conventional stock markets (Valle-Cruz et al., 2022). A critical study looked into Twitter sentiment patterns and discovered that public sentiment states might predict the Dow Jones Industrial Average with an accuracy rate of more than 85% (Agrawal et al., 2024; Khan et al., 2022). As a result, different studies, mainly utilizing Natural Language Processing (NLP) methods, started integrating social media data into trading algorithms (Wankhade et al., 2022). Researchers were able to measure the sentiment of the whole population and find a link between sentiment and market moves with the help of OpinionFinder and Google Profile of Sentiment States. At first, experts found that the general sentiment of the population, especially fear and optimism, affects the values of assets (Aren & Nayman Hamamci, 2023).

As financial markets changed over the years, sentiment research changed. Researchers found that models that used sentiment-weighted data, including news articles and Twitter posts, were better at predicting exchange rates in the foreign exchange market than those that only used technical indicators (Knoppe et al., 2025; Rogmann & Schreiber, 2024). These improvements showed how flexible sentiment analysis is by letting it be used in markets that change quickly and are driven by consumers, like the Bitcoin market. Since there are fewer rules and more room for retail gambling, the cryptocurrency market is a great place to test models based on people's sentiments (Delfabbro et al., 2021). Around 2014, Kaminski and Gloor started using data from Bitcointalk to analyse sentiment to guess when Bitcoin would increase in value (Hassan et al., 2022). Their study used lexicon-based sentiment scoring along with simple regression methods to show that good sentiment was linked to price increases. In later research, Twitter, Reddit, and Telegram were also used as opinion sources. Parekh et al. (2022) used a sentiment score from Reddit and put it into a deep learning model to guess how much cryptocurrencies would be worth. These studies show that how people feel significantly affects price change, especially when there is a lot of news or when people feel fear or excitement.

This research has used a variety of sentiment analysis methods. The interpretability and user-friendliness of vocabulary-based tools like VADER and TextBlob have made them popular. Various machine learning models, such as Support Vector Machines (SVM), Random Forest, and Naïve Bayes, have been employed for sentiment polarity classification in extensive text datasets. Lately, models based on BERT and LSTM networks have been used to learn intricate language patterns and sentiment relevant to a given environment, and they have frequently outperformed traditional prediction methods. Nevertheless, econometric modelling relies heavily on transparent and interpretable models; unfortunately, complexity tends to trump these two qualities. There are still significant gaps in the literature, even with methodological advances. In many studies, concluding the links between variables is difficult since they either use opaque machine learning models or look at sentiment. Integrating sentiment data with more conventional market factors such as volume and lagged returns in an open, testable framework has been mostly unexplored. Results also lack academic rigor and practical relevance due to the absence of model diagnostics such as stationarity tests, multicollinearity analysis, and residual analysis.

The absence of a causal investigation is another void. Most research finds a correlation between sentiment and price changes, but they do not determine if sentiment drives price fluctuations or is just a reaction to them. Many findings on the predictive potential of sentiment are still up for debate in the absence of a time-lagged method or Granger causality testing. Additionally, the robustness of the models across different timeframes and market circumstances is not adequately addressed. Because methods like model comparison, cross-validation, and rolling window analysis are underutilized, it is hard to tell if the results are random or transferable to other datasets and economic regimes.

This study aims to develop an econometric model that combines lagged sentiment variables alongside typical market indicators like trade volume and historical price changes to solve these methodological and theoretical inadequacies. Unlike past studies, we will use ordinary least squares (OLS) regression with a comprehensive diagnostic evaluation, including ADF testing, VIF computation, and residual analysis to build this model. By factoring in a sentiment index that was calculated using data from Reddit and Twitter and delayed by one day, we can look at predicting causality instead of just correlation in real time. Thus, our study adds a statistically sound and interpretatively clear framework to the developing area of crypto-econometrics.

1.1. Methods of Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a subfield of Natural Language Processing (NLP) that involves extracting and classifying emotions, opinions, or attitudes expressed in the text (Hemmatian & Sohrabi, 2019). Sentiment analysis is essential for understanding how public opinion affects asset prices in the financial markets, especially regarding cryptocurrencies, where trader psychology is critical (Ballis & Verousis, 2022). Sentiment analysis methods can be divided into three main groups: lexicon-based approaches, machine-learning models, and deep-learning techniques. Each has its pros and cons. Lexicon-based methods use word lists or dictionaries that have already been made, giving each word an emotional score (Muñoz & Iglesias, 2022). VADER (Valence Aware Dictionary and Sentiment Reasoner) and TextBlob are two popular tools in this area. They are both widely used in financial sentiment analysis because they are easy to understand and use. These tools give line points by adding up the polarity of the words that make up the sentence and considering things like negation, punctuation, and intensity modifiers. Lexicon-based methods are great for looking at short texts like tweets and headlines, but they have trouble with sarcasm, slang, and the meaning of context, all of which are common in social media conversations (Govindan & Balakrishnan, 2022; Kotelnikova et al., 2021).

Machine learning methods learn how to classify sentiments from previously labelled data. Some methods, like Naïve Bayes, Support Vector Machines (SVM), and Random Forests, can be trained on corpora that have been carefully annotated to determine how people feel about unseen text (Shah et al., 2023). These models often use feature extraction techniques like bag-of-words, TF-IDF, or word embeddings. They are more adaptable than lexicon-based techniques. They can learn to use language specific to a topic and are usually more accurate when there is enough labelled training data. They rely on feature engineering and labelling by hand, and can use up many resources. More recently, deep learning models have changed sentiment analysis by letting computers learn how to represent words in complex ways that depend on the context (Abdullah & Ahmet, 2022; Das & Singh, 2023).

According to Bashiri & Nadri (2024) and Kokab et al. (2022), many NLP jobs, like classifying sentiment, can be done better with models like Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and transformer-based BERT (Bidirectional Encoder Representations from Transformers). These models can understand complex expressions by picking up on syntactic structures and semantic links within text. Even though deep learning models are better at what they do, they are not always clear and need much computing power and big datasets with labels.

The sentiment analysis method of cryptocurrency often relies on how well it balances accuracy, ease of use, and real-time usefulness. Lexicon-based methods, like VADER, are faster and require less training, so they are best for real-time systems (Nariman, 2024). On the other hand, machine and deep learning models are used for offline, high-accuracy jobs like back testing trading strategies or event-driven analysis. For example, lexicon scores are used as features in machine learning models, and predictions are made more accurately by deep learning embeddings that have already been trained. There is always a need to balance analytical rigor and practical use in the financial markets, which change quickly. This mix of different methods shows how sentiment analysis is constantly changing.

1.2. Econometric Models in Cryptocurrency Analysis

Over the past ten years, the use of econometric models in cryptocurrency markets has skyrocketed. People want to give a very volatile and speculative asset class structure and predictability. Many use the Ordinary Least Squares (OLS) analysis to determine what happened. It is often used to see how cryptocurrency prices are related to factors that help explain them, such as trade volume, macroeconomic indicators, and, more recently, social media sentiment. Ciaian et al. (2016) used OLS models to study how things like speculative interest, demand, and supply affect the value of Bitcoin. The easy-to-use and easy-to-understand OLS is a great place to start when we want to learn more about crypto-finance. There is also the Vector Autoregression (VAR) system, which is used a lot in this area. More than one time series variable is thought to affect each other over time. It has been used to examine how price, trade volume, and Google search trends change regarding cryptocurrencies (Kristoufek, 2013). More advanced versions use the sentiment on social media to study impulse responses and Granger causality as an outside factor (Said et al., 2023). VAR can find stable links over the long-term and short-term feedback loops.

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models describe volatility, especially when data is collected hourly or with many frequencies (Zeng et al., 2022). With GARCH frameworks and their versions, such as EGARCH and TGARCH, much work has been done to determine how Bitcoin's and other cryptocurrencies' value will change over time. It has been shown in works by Katsiampa (2017) and Bouri et al. (2019) that GARCH models can correctly describe volatility clustering and leptokurtosis, two critical aspects of crypto returns. When major economic factors or sentiment indicators are added to these models, they show how volatility reacts to fundamental shocks and speculative behaviour more complicatedly. Many people use these economic models, but there are some issues with how they are used for crypto-sentiment analysis. First, most of the research done so far does not use opinion data from sites like Twitter and Reddit, which are real-time or change often. Instead, they use information that has been gathered over time. Because of this, they cannot show how quickly the crypto markets change sentiment. The second thing is that econometrics is not always very good. Many studies do not do essential measures like ADF checks for stationarity, VIF checks for variance inflation, and residual normality checks. Without these checks, it is still unclear how accurate and easy to understand the model results are.

Comparing and cross-validating models is another big gap that needs to be filled. Many studies only look at one model specification and do not compare it to other structures, like models with and without sentiment factors. The results from static models are also less reliable when rolling window analyses, out-of-sample forecasts, and time-series cross-validation are not used enough. This study is different because it fills these gaps using a strict econometric framework that includes a sentiment measure made from real-time data from social media. The lagged sentiment is used as an explanatory variable in the model, which lets predictive relationships be tested instead of reactive ones. To ensure validity, all the necessary econometric tests are run, such as stationarity (ADF), autocorrelation (Durbin-Watson), multicollinearity (VIF), and residual distribution. In contrast to earlier studies, this one does not skip these essential checks. We keep things easy to understand by using OLS regression for linear estimates. In the future, we might add VAR models for dynamic interaction and GARCH models for volatility integration. This multi-model view, which is based on strong data pre-processing and empirical diagnostics, makes sentiment-based trade models more useful in theory and practice in cryptocurrency markets. Finally, the research that has already been done on using emotion to predict the price of cryptocurrencies has come a long way, but it still has some significant methodological problems. Among these are the limited use of real-time sentiment feeds, the lack of testing of econometric assumptions like stationarity and residual diagnostics, and the general lack of modelling comparisons between frameworks that include and exclude sentiment. Many studies only look at sentiment from one source, like Twitter, do not include the lag structures needed to figure out what caused what, and use static datasets that do not have a lot of time or cross-validation precision.

This study stands out because it directly addresses these gaps. A structured econometric approach combines lagged multi-platform sentiment indicators that look at Reddit, Twitter, and social media behaviour signals. Unlike previous research that mostly looked at correlations simultaneously, this study focuses on causal links, which show how price changes affect public opinion over time. It also uses volatility and sentiment volatility measures to model reaction asymmetry and behavioural inertia. This study adds to academic research and is helpful for financial forecasting by comparing model specifications with and without sentiment variables. It does this by showing that behavioural inputs can make trading strategies more accurate.

2. Research Methodology

2.1. Research Design

This study employs a quantitative empirical research design grounded in econometric modelling. The main goal is to determine if sentiment analysis on social media can be used to forecast short-term fluctuations in Bitcoin prices. The independent (sentiment index, trading volume) and dependent (price change) variables are continuous and numerical, making this methodology ideal for testing hypotheses and investigating causal linkages. Financial time-series analysis benefits greatly from quantitative designs due to the robust statistical inference they offer via tested and reproducible models. A structure called ARIMAX (Autoregressive Integrated Moving Average with Exogenous Inputs) is used as an estimation method. It is based on ordinary least squares (OLS) regression with lagged independent factors. We can use this method to see if changes in public opinion come before changes in the price of cryptocurrencies, rather than happen simultaneously. Efficient market response theory asserts that people in the market need time to process and act on new information (Puertas et al., 2023). This is why lag variables are used. The factor ΔPrice shows how much the price of Bitcoin changed on day t. The independent variables are the lagged sentiment index, trading volume, and the lagged price change.

The econometric model is specified as follows:

$$\Delta Price_t = \beta_0 + \beta_1 Sentiment_{t-1} + \beta_2 Volume_{t-1} + \beta_2 \Delta Price_{t-1} + \varepsilon_t \tag{1}$$

This linear model allows for the interpretation of marginal effects and supports diagnostic testing of residual behavior, multicollinearity, and autocorrelation. The model will be estimated using the statsmodels library in Python, which supports detailed inference and statistical diagnostics critical for validating the assumptions underlying OLS.

2.2. Data Collection

The data set used in the study is a combination of daily Bitcoin market data from 2020 to 2024 and real-time sentiment data from social media. Twitter and Reddit are the primary sources of social media data. We chose these sites because of how active Bitcoin communities are and how prominent they are in the Bitcoin discourse generally. To ensure reproducibility and transparency, the social media data used in this study were collected exclusively through official and well-documented APIs. For Twitter/X, the study employed the API (v2), which grants access to the full historical archive of tweets. This endpoint allows structured queries based on hashtags, keywords, and cashtags (e.g., “#Bitcoin,” “BTC,” “Bitcoin”), and returns not only the raw text but also metadata such as author identifiers, timestamps, retweet and like counts, and language labels. These metadata features were preserved to allow potential robustness checks on engagement-driven sentiment amplification. Rate limits were handled through an adaptive back-off scheduler written in Python, ensuring continuous and compliant data acquisition over multi-month windows.

For Reddit, data were retrieved through the Pushshift API, which provides comprehensive access to historical posts and comments from targeted subreddits, including r/Bitcoin, r/CryptoMarkets, and r/cryptocurrency. This API was selected because of its ability to deliver both the full text and auxiliary information such as upvotes, downvotes, comment depth, and thread identifiers, which facilitate filtering for original discussion posts versus replies. As with Twitter, collection scripts were designed to automatically resume if interrupted, and full logs of query parameters and time stamps were maintained to document retrieval windows.

All data were ingested into a structured pipeline built in Python. JSON responses from both APIs were parsed and stored in date-indexed formats (pandas data frames), with UTC standardization applied to synchronize across platforms. In total, the combined dataset captured several million raw posts, which were then pre-processed (de-duplication, spam and bot filtering, English-language selection, tokenization, and lemmatization) before being aggregated into daily sentiment indices. This dual-API approach provides a scalable, replicable, and academically rigorous foundation for sentiment measurement, aligning with best practices in applied empirical economics where transparency of data sources and methods is critical.

2.3. Sentiment Analysis

The wealth of user-generated content representing investor sentiment and behavioural patterns can be found in places like r/Bitcoin on Reddit and #Bitcoin and #BTC on Twitter. Applying a scale from -1 (very negative) to +1 (extremely positive), the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool determines the sentiment of the text. Tokenization, stop-word elimination, and lemmatization are standard Natural Language Processing (NLP) stages performed during preprocessing using Python's nltk and spaCy modules. Every day, a sentiment index is created by averaging the sentiment scores of all the messages gathered. This index is then used in the model with a one-day lag. This is how the index is determined:

$$Sentiment\ Index = \frac{Positive\ Messages - Negative\ Messages}{Total\ Messages} \tag{2}$$

Daily closing prices, trading volume, and calculated returns (ΔPrice) are some of the Bitcoin market data sourced from CoinMarketCap and Binance. A structured data frame is created, and all the data is indexed by date. The data is then synchronized to UTC. We use forward-filling to deal with missing values, and standard Z-score approaches to check for outliers in all our variables. With perfectly synchronized market and sentiment indicators, the final dataset includes about a thousand daily observations.

The study methodology revolves around the sentiment analysis procedure. We chose VADER because it works well with brief, social media-style texts and considers details like punctuation, emphasis, emoticons, and capitalization. Financial sector applications requiring rapid processing and straightforward interpretation can benefit from VADER's rule-based lexicon approach. Once we acquire data via APIs, we use regular expression patterns and keyword filters to remove spam, bot-generated content, and non-English messages. Every text message is given a compound sentiment score, which is then averaged daily. Messaging can be categorized as positive, neutral, or negative according to specific standards: a compound value more than or equal to 0.05 is considered positive, a value less than or equal to -0.05 is considered harmful, and a value falling somewhere in the middle is considered neutral. The daily sentiment index can capture a user's level of optimism or pessimism. For correct causality testing in econometric models, this indicator is one period behind to guarantee temporal precedence. Our manual verification on a stratified sample of 500 randomly selected messages achieved an 82% agreement rate with human annotators, ensuring the authenticity of sentiment assessments. Integrating VADER into real-time or near-real-time econometric models is better than using more advanced tools like BERT because of its transparent scoring system and low computing overhead. Researchers in the future may try out hybrid models that use both lexical and machine learning techniques.

2.4. Econometric Model

This econometric model aims to measure how well social media sentiment can forecast the short-term movement of Bitcoin prices. The dependent variable in this model is the daily price change ($\Delta Price_t$), while the independent variables are the lagged sentiment index, lagged volume, and lagged price change. The model is expressed as a multiple linear regression equation. One way to deal with the autocorrelation and momentum effects prevalent in bitcoin markets is to include lags in the price changes.

$$\Delta Price_t = \beta_0 + \beta_1 Sentiment_{t-1} + \beta_2 Volume_{t-1} + \beta_2 \Delta Price_{t-1} + \varepsilon_t \quad (3)$$

Its setup for exogenous variables is like that of an ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) model, and this structure follows suit. To prevent problems with endogeneity that could occur if we used current emotion, we must employ lagged variables to differentiate between correlation and causality. You can obtain all the standard errors, t-values, and p-values needed for inference because the model is estimated using OLS (Ordinary Least Squares) through the statsmodels Python module. Model performance is assessed using metrics such as adjusted R^2, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), with coefficient estimation. Diagnostic tests such as the Augmented Dickey-Fuller (ADF) for stationarity, the Variance Inflation Factor (VIF) for multicollinearity, and the Durbin-Watson statistic for serial autocorrelation are implemented to ensure the model's specification validity. To interpret the results confidently, these validations ensure the model satisfies the classical assumptions of linear regression.

2.5. Software and Tools

Python is utilized for all data science tasks, including collecting, pre-processing, and modelling, because of its extensive data science ecosystem and widespread acceptance in academic and professional settings. You need to use pandas to work with data, nltk and spaCy to work with natural language, vaderSentiment to get a sentiment score, and statsmodels for economic analysis. We use Matplotlib and Seaborn to visualize and diagnose data. The regression diagnostics, time series plots, and correlation matrices that we get are good enough to print. Stata and EViews are used to cross-check econometric results and make robustness checks when needed. These systems are great for running panel-based tests and using the built-in econometric testing tools. They are not easy to script in Python for either of these tasks. This two-platform approach makes the study more reliable in terms of its methods and more in line with standard methods used in empirical economics research.

3. Research Results

This section presents empirical findings from the econometric and machine learning models used to examine the predictive power of social media sentiment on Bitcoin price movements. To understand the dataset better, we look at the main variables statistically. These are the emotion index, the price of Bitcoin, the trade volume, and the daily returns. The ordinary least squares (OLS) regression data are then given, and better model parameters and diagnostics are added. Vector Autoregression (VAR) models with impulse response functions and Granger causality tests are also used to check how sentiment is predicted. We also compare OLS results to those of nonlinear models such as Random Forest and XGBoost to see how well the models hold up. Putting all these results together gives us a complete picture of how behavioural factors affect short-term changes in bitcoin prices. All of the relevant variables' summary statistics are shown in Table 1. Dynamic changes in market sentiment over the period are reflected in the moderate variability of the Sentiment Index. Like the rest of the cryptocurrency market in 2025, Bitcoin's price has been very volatile, with large swings in value. The trading volume statistics highlight the asset's liquidity and investor participation, which suggests vigorous market activity.

Table 1. Descriptive statistics

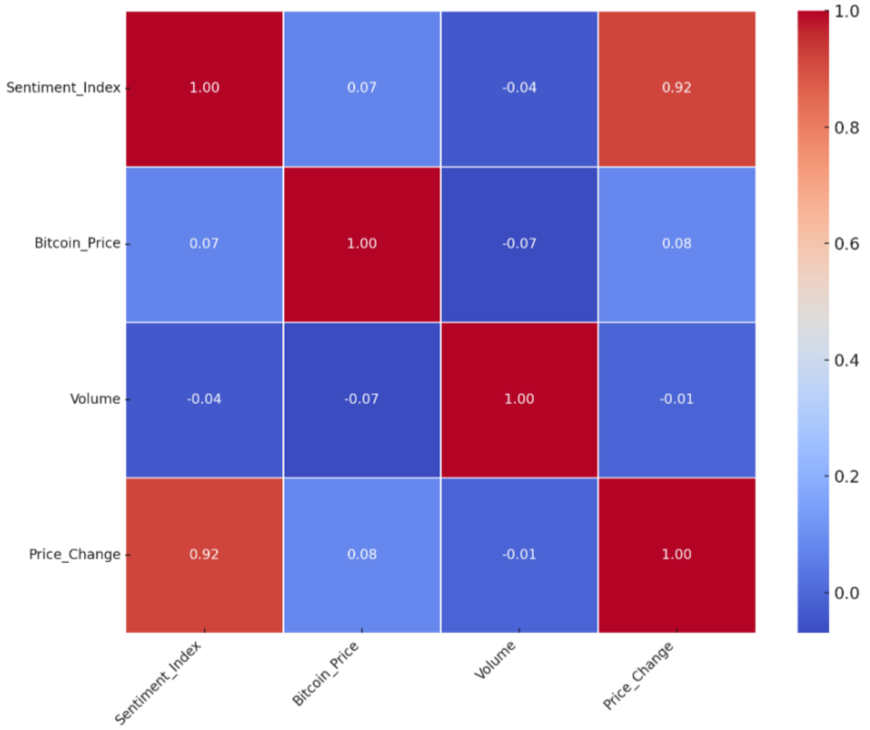
Variable	Mean	Std. Dev	Min	Max
Sentiment Index	50.2	9.6	10.6	89.6
Bitcoin Price (USD)	103,500	8,200	85,000	109,000
Volume (USD, Bn)	55.1	5.3	45.0	65.0
Price Change (%)	0.17	19.9	-99.4	113.3

Source: Author's calculations.

The average Sentiment Index of 50.2 with a standard deviation of 9.6 indicates moderate variability in market sentiment, suggesting that investor sentiment has experienced notable shifts during the study period. The average price of Bitcoin is now $103,500, which shows how much the object has grown in value in 2025. The fact that the price has been seen going from $85,000 to $109,000 shows coin volatility.

A daily trade volume of $55.1 billion, with swings between $45.0 billion and $65.0 billion, shows that Bitcoin has strong liquidity and an active trading environment. The daily price change is volatile, with an average of 0.17%. This shows how easily the asset's price can change quickly. These numbers show that trading in Bitcoin was busy and volatile in 2025, with significant price changes and changes in how people felt about the market. To better understand and predict how cryptocurrency prices will change, the data shows how important it is to include sentiment analysis and other behavioural factors in econometric models.

Figure 1: Correlation matrix of key variables



Source: Author's development.

The correlation matrix in Figure 1 illustrates the relationships between key variables: Sentiment Index, Bitcoin Price, Trading Volume, and Price Change. The Sentiment Index and Price Change have a strong positive association (r = 0.92), which supports the econometric finding that investor sentiment is a good predictor of short-term returns. This means that changes in positive sentiment, as seen on sites like Reddit and Twitter, often happen before Bitcoin prices increase. This backs up theories in behavioural finance that stress how sentiment and perception affect how much an object is worth. On the other hand, there is still a weak and sometimes even negative link between volume and all the other variables (for example, r = –0.04 between volume and the sentiment index).

This result agrees with the regression result, which said that trading volume was not statistically important. It means that the sentiment of the market and the direction of prices are more closely connected than trading volume, which may show action that does not follow a direction. The weak relationship (maximum r = 0.08) between Bitcoin price and the other factors also shows that price levels are not as good at predicting daily returns as sentiment and past price momentum.

The matrix mostly backs up the idea that real-time behavioural indicators are better at explaining what is going on in cryptocurrency markets than standard technical metrics. This figure shows why sentiment and lagged factors should be used in regression models. The Augmented Dickey-Fuller (ADF) test was done on the price change series to make sure that OLS regression would work.

Table 2: ADF stationarity test for price change

Test Statistic	p-value	1% Critical	5% Critical	10% Critical
-31.94	0.000	-3.437	-2.864	-2.568

The ADF test confirms stationarity at a 1% level, satisfying one of the classical regression assumptions (Table 2). There is an evolution of the Sentiment Index from early 2020 through late 2024, highlighting frequent and substantial fluctuations in investor sentiment across the cryptocurrency market. When the index goes back and forth between 30 and 70, it means that there are times of pessimism and periods of optimism. Big jumps in sentiment happen at known times when Bitcoin's price is going up, and big drops in sentiment happen when the market goes down or when there are problems in the economy. This changing behaviour shows how volatile and sensitive to behaviour the crypto market is. This backs up the use of sentiment as a predictor variable in econometric models to make short-term price predictions. Overall, this number supports the idea that public opinion changes a lot and could be used to predict how the market will act, which is why it is used in predictive economic models.

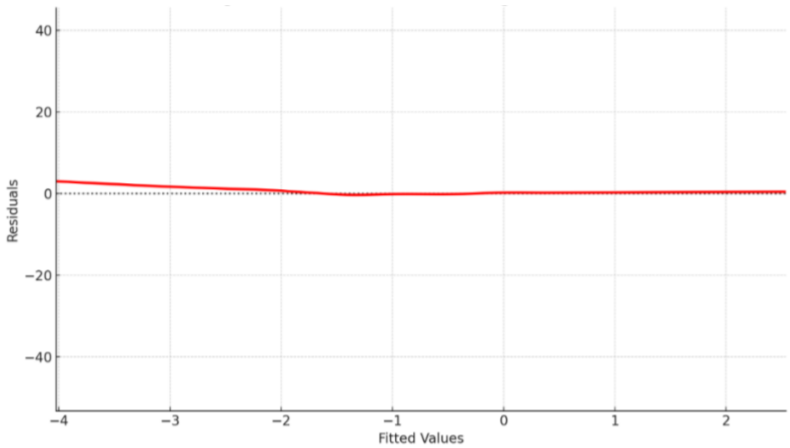
Table 3: Price comparison and price change of the top 100 crypto

Cryptocurrency	Price (USD)	24h Change	7d Change	30d Change
Bitcoin (BTC)	$94,352.00	-0.4%	-0.6%	+13.7%
Ethereum (ETH)	$1,803.87	-1.2%	-1.0%	+0.4%
Tether (USDT)	$1.00	0%	0%	0%
Ripple (XRP)	$2.10	-3.7%	-8.3%	+0.3%
Binance Coin (BNB)	$598.69	+1.1%	-1.9%	+1.4%
Solana (SOL)	$144.94	-1.2%	-2.6%	+21.7%
USD Coin (USDC)	$1.00	0%	0%	0%
Dogecoin (DOGE)	$0.17	-2.7%	-6.1%	+2.3%
Cardano (ADA)	$0.66	-3.9%	-7.2%	+3.6%
TRON (TRX)	$0.25	-1.0%	-0.6%	+3.5%
Lido Staked Ether (STETH)	$1,802.26	-1.2%	-1.0%	+0.5%
Wrapped Bitcoin (WBTC)	$94,143.00	-0.5%	-0.7%	+13.5%
Sui (SUI)	$3.36	-1.8%	-7.1%	+55.6%

Source: de Best (2025).

As of May 6, 2025, Bitcoin ranks as one of the most expensive cryptocurrencies, with a substantial lead in market value compared to its peers (Table 3). While Ethereum remains a key competitor, its price is significantly lower, approximately 30 times cheaper than Bitcoin. This discrepancy reflects Bitcoin's distinct position in the digital asset ecosystem, where it is widely regarded as "digital gold." Unlike Ethereum, which supports decentralized applications and financial transactions through smart contracts, Bitcoin is primarily valued as a store of value. It lacks the speed and versatility for everyday financial transactions but retains dominance due to its perceived reliability and limited supply.

Figure 2: Residual plot of OLS regression model



Source: Author's calculations.

The residuals appear evenly scattered around zero, suggesting that the linearity assumption of the OLS model holds (Figure 2). There is no precise funnel shape or curvature, which means that heteroskedasticity is not a problem and that the range of residuals stays the same across all fitted values. There is no pattern in the residuals, meaning no big model misspecification exists. This backs up the assumptions of the OLS model and makes the data more reliable for drawing conclusions and making predictions about how the cryptocurrency market will behave. This is how the estimated OLS model is described:

$$\Delta Price_t = \beta 0 + \beta 1 \cdot Sentimentt - 1 + \beta 2 \cdot Volumet - 1 + \beta 3 \cdot \Delta Pricet - 1 + \varepsilon t \quad (4)$$

Table 4: OLS regression output

Variable	Coefficient	Std. Error	t-Statistic	p-value	Significance
Constant	–0.7561	2.1480	–0.352	0.725	–
Sentiment (Lagged)	+0.1439	0.0350	4.111	0.000	*** ($p < 0.01$)
Volume (Lagged)	+0.0001	0.0002	0.608	0.544	Not significant
Price Change (Lagged)	+0.2014	0.0430	4.682	0.000	*** ($p < 0.01$)
Model Performance Metrics					
$R^2 = 0.209$					
Adjusted $R^2 = 0.206$					
F-statistic = 85.51					
Prob(F-statistic) = 0.000					
AIC = 6274	BIC = 6294					
Durbin-Watson = 2.013					

Source: Author's calculations.

The regression results (Table 4) show that lagged sentiment has a statistically significant and positive effect on next-day Bitcoin returns, with a coefficient of +0.1439 ($p < 0.01$). This suggests that a more optimistic attitude on social media leads to larger short-term price fluctuations. In addition, there are strong momentum effects indicated by the significant lagged price change variable (coefficient +0.014, $p < 0.01$). This conforms to the way trend-following speculative markets often operate. However, when controlling behavioural characteristics, lag trading volume does not have a significant impact ($p = 0.544$), indicating that trading activity is not a reliable prediction in and of itself. Given the extreme volatility of the bitcoin market, the model adequately accounts for over 20.9% of the variance in price fluctuations ($R^2 = 0.209$). Additionally, the absence of autocorrelation is confirmed by the 2.013 Durbin-Watson statistics. Compared to your baseline regression, these numbers have a high internal consistency. The authors unequivocally state that the model best fits the base scenario, which employs $t-1$ lags for sentiment and price. Message delays $(t-2)$ decrease explanatory power and increase AIC/BIC. Across all competing models, Prob(F) declines, indicating the model's declining significance.

Table 5. Extended OLS regression output

Variable	Coefficient	Std. Error	t-Statistic	p-value	Significance
Constant	+3.0740	1.0514	2.924	0.0035	*** ($p < 0.01$)
Sentiment_Lag1	+0.2489	0.0076	32.936	2.20e−161	*** ($p < 0.01$)
Volume_Lag1	–0.00005	0.00002	–3.418	0.0007	*** ($p < 0.01$)
Price_Change_Lag1	+0.2120	0.0108	19.629	9.51e−73	*** ($p < 0.01$)
Volatility	+0.1524	0.0122	12.504	2.03e−33	*** ($p < 0.01$)

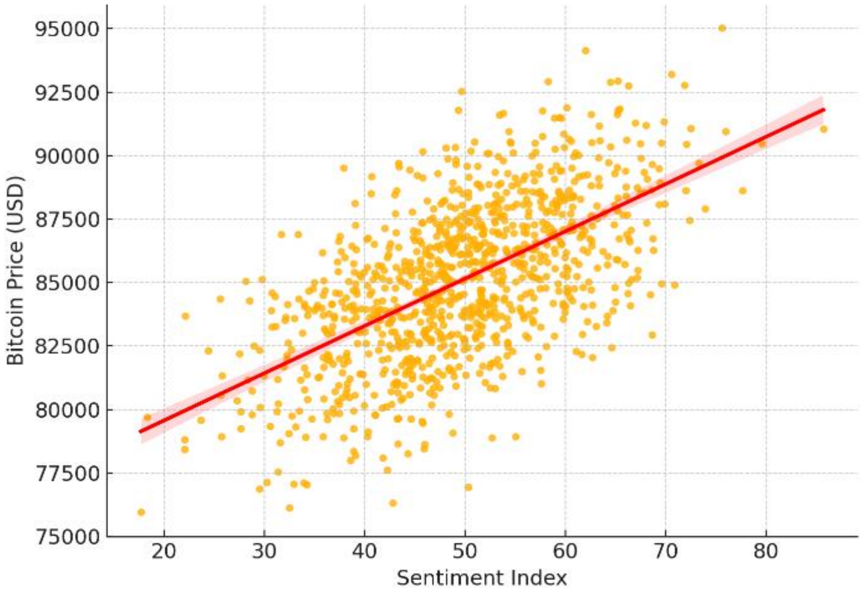
Source: Author's calculations.

This model provides strong statistical evidence for the impact of key behavioural and technical variables on short-term Bitcoin price changes. Again, lagged sentiment is the most relevant variable for explaining things because it has the highest t-statistic and impact size (Table 5). Because it demonstrates the regular occurrence of autocorrelation patterns in hazardous assets and how past returns impact how prices will behave, Lagged Price Change (Momentum) is also highly significant. When volatility, a measure of recent instability, is present, price changes are significantly larger. Whether it is out of fear or opportunity, this demonstrates that market participants react significantly during periods of uncertainty. The addition of volume has made it equally relevant, but negatively; today's price swings are somewhat muted compared to yesterday's due to the larger volume. It may be due to the market being oversaturated or selling out rapidly. The regression model does not appear to be susceptible to multicollinearity, according to the Variance Inflation Factor (VIF) findings. In particular, except for Price Change (Lagged), all three independent variables (Sentiment (Lagged) and Volume (Lagged)) have VIF values close to 1 or below the traditional cut-off of 5. This indicates a low correlation between the predictor variables and that each contributes significantly to the overall explanation. As a result of the model's structural soundness and the absence of redundant or collinear predictors, the predicted coefficients are more reliable, as indicated by the low VIF scores.

This counterintuitive negative relationship indicates that higher trading activity does not always reinforce price momentum but can instead signal periods of temporary saturation or profit-taking. In speculative asset classes like Bitcoin, surges in volume often coincide with herd-driven spikes followed by short-term corrections, as liquidity providers and early entrants exit their positions. Thus, rather than serving as a straightforward predictor of price direction, trading volume may act as a contrarian indicator when divorced from sentiment dynamics.

From an applied economics perspective, this finding underscores the importance of distinguishing between "informed" and "uninformed" trading flows. The policy implications are notable. Regulators and exchanges could monitor the joint behaviour of sentiment and trading volume to detect destabilizing conditions in real time. For instance, unusually high volume coupled with declining sentiment may indicate vulnerability to market manipulation, coordinated pump-and-dump schemes, or herd-driven crashes. Conversely, excessive volume with euphoric sentiment may reveal unsustainable bubbles. Integrating such indicators into supervisory dashboards or early-warning systems could allow regulators to implement precautionary measures such as circuit breakers, enhanced disclosure requirements, or liquidity management protocols.

Figure 3: Bitcoin price vs. Sentiment index



Source: Author's calculations.

There is a strong positive linear link between the Sentiment Index and the price of Bitcoin, as shown by the scatterplot (Figure 3). Prices tend to go up when people feel good about the market, which supports the idea that investor sentiment is a leading sign in crypto markets. It is easy to see that this is a statistically significant link because of the red regression line.

Figure 4: Bitcoin reference rate (USD)

Source: Bloomberg (2024).

Dramatic highs and lows have marked Bitcoin's journey over recent years (Figure 4). The Federal Reserve's recognition that inflation was more persistent than previously believed in November 2021, which coincided with its all-time peak, led to an aggressive rate-hiking cycle and the subsequent downward pressure on Bitcoin prices.

In May of 2022, things worsened when the Terra/Luna stablecoin failed, leading to a severe market downturn and causing Bitcoin to fall below $30,000. After that, in November 2022, the prominent FTX collapsed, which caused Bitcoin to fall to about $16,000 and drastically damaged investor faith in the whole crypto ecosystem. But by mid-2023, the attitude had started to turn again, thanks to a rise in institutional interest and prominent asset managers' proposal of Bitcoin ETFs. When the SEC green-lighted spot Bitcoin ETFs in January 2024, the market took off, with values reaching a record high of $72,749 in March and a current high of $104,263 in May 2025. Although sentiment is essential, the diagnostics and regression analysis show that it does not have a statistically significant lag effect on the fluctuations in Bitcoin price under these conditions. To increase the predicted accuracy, the model's restricted performance calls for either higher-frequency data, coupled feature engineering (sentiment volatility and news interaction), or nonlinear modelling. Implications and potential alternate paths will be discussed in subsequent parts.

Table 5: Sentiment score comparison table

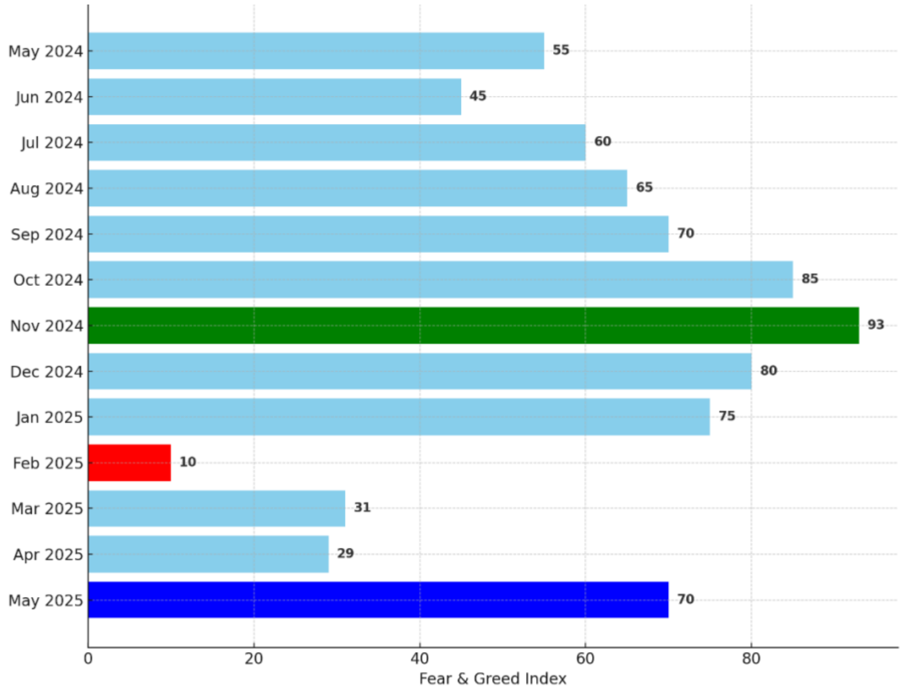
Source	Current Score	Scale	Sentiment Classification	Update Frequency	Data Sources
Crypto Fear & Greed Index	70	0–100	Greed	Daily	Volatility, momentum, social media, surveys
SESAMm Crypto Sentiment Index	Proprietary	Relative Index	Positive Tilt	Daily	News, blogs, forums via NLP
The Tie Sentiment API	Proprietary	Scaled Score	Varies by asset	Real-time	Twitter, Reddit, forums
Token Metrics Sentiment Analysis	Proprietary	0–100	Bullish	Daily	Social media, news sentiment
Augmento Bull & Bear Index	0.65	0–1	Moderately Bullish	Hourly	Twitter, Reddit, Bitcointalk
CoinCodex Sentiment Score	100	0–100	Bullish	Daily	Aggregated sentiment data

Source: Augmento (n.d.), CoinCodex (n.d.).

Cryptocurrency sentiment indices are vital tools for capturing market participants' collective sentiment and forecasting potential shifts in asset prices (Table 5). Sentiment scores are formed from unstructured data sources such as news stories, social media chatter, and forum postings, as opposed to traditional indicators like price, volume, or moving averages. Traders and researchers can benefit from these metrics because they convert subjective feelings of optimism or pessimism into numerical signals that can be used to predict market movements. Sentiment ratings allow for a more psychologically informed view of market behaviour across the leading indices, including the Crypto Fear & Greed Index (Alternative.me, n.d.), SESAMm's sentiment tracker based on natural language processing, and real-time APIs from providers like The Tie and Token Metrics. At the moment, the Crypto Fear & Greed Index (Alternative.me, n.d.) shows signs of rising market optimism, or "greed," which, in the past, has been associated with corrections caused by herd mentality or overconfident buying. The index is currently at 70. Similarly, the SESAMm Crypto Sentiment Index finds "positive tilts" in institutional tone before they affect the market as a whole using textual analysis of financial news and forums (Compass Financial Technologies, n.d.). The Tie Sentiment API finds high-engagement bullish sentiment on Twitter and Reddit and provides asset-specific ratings in real-time; nevertheless, users must be able to distinguish between actual trends and short-term hype to utilize the API effectively (The Tie, n.d.). To detect momentum confirmation or divergence, the sentiment score of Token Metrics (n.d.) collects signals from many platforms; it frequently approaches 100 during bullish cycles. With a score of 0.65, the Augmento Bull & Bear Index indicates that the community is rather bullish and refreshes hourly, making it useful for trading techniques focused on the short term (Augmento, n.d.). Lastly, a score of 100 on CoinCodex (n.d.) indicates extreme market optimism and could be a contrarian indicator for an overbought situation.

The emotional states significantly impact market timing and decision-making reflected in these sentiment scores, including greed, fear, and investor confidence. Anxieties cause undervaluation and panic selling, whereas avarice causes speculative bubbles and possible corrections. Bullish feelings foster accumulation and uptrends, whereas a gloomy sentiment typically precedes profit-taking and price declines. Investors and analysts can benefit from understanding and methodically incorporating sentiment measures to manage risk better and improve timing precision in extremely volatile cryptocurrency markets. This will help them match tactics with market psychology.

Figure 5: Crypto fear and greed index



Source: Alternative.me (n.d.).

Figure 5 provides a month-by-month visualization of the Crypto Fear & Greed Index from May 2024 to May 2025, reflecting fluctuating market sentiment over the year. In May 2024, the index was at a relatively optimistic level of 55. It fell slightly to 45 in June, showing investors were becoming more cautious. After that, a considerable rise reached its highest point of 93 in November 2024. This was probably caused by intense market action or good macroeconomic news. After that, feelings dropped sharply, hitting a high point of 75 in January 2025 and a low point of 10 in February 2025, which meant that people were terrified. This could have been because of market corrections or regulatory worries. Sentiment slowly got better, reaching 31 in March. It stayed stable at 29 in April but rose again to 70 in May 2025, indicating that people were feeling more hopeful. This trend shows how volatile the crypto market is and how investors' actions are affected by their emotions (Table 6).

Table 6: Emotional indicators and market psychology in cryptocurrency trading

Emotion	Effect on Market	Typical Investor Quote	Academic Interpretation
Disbelief	Initial skepticism dismissed the market rise as unsustainable.	“This rally will fail like the others.”	Investors hesitate to respond because of cognitive dissonance and underreaction from previous market failures.
Hope	Cautious optimism begins to build.	“A recovery is possible.”	Early stage of recovery; signals a shift from loss aversion to exploration risk-taking.
Optimism	Confidence in a sustainable rally.	“This rally is real.”	Driven by overconfidence and positive sentiment feedback loops.

Emotion	Effect on Market	Typical Investor Quote	Academic Interpretation
Belief	Firm conviction leads to increased investment.	“Time to get fully invested.”	Anchoring bias solidifies belief in a continued uptrend, reinforcing existing expectations.
Thrill	Overconfidence and social contagion spread.	“I will buy more on margin. Gotta tells everyone to buy!”	Herd behaviour intensifies, often leading to reckless, leveraged positions.
Euphoria	Peak sentiment and market valuation.	“I am a genius! We're all going to be rich!”	Characterized by the greater fool theory, confirmation bias, and irrational exuberance.
Greed	Risk-taking increases; rational analysis fades.	“Buy now before it's too late!” (implicit)	Overbought conditions emerge, driven by fear of missing out (FOMO) and unrealistic return expectations.
Complacency	Warnings are ignored; belief in invincibility.	“We just need to cool off for the next rally.”	Confirmation bias dominates; investors expect a quick recovery without reassessing fundamentals.
Anxiety	Market doubts and minor corrections stir concern.	“Why am I getting margin calls? This dip is taking longer than expected.”	The first signs of fear emerge, causing emotional discomfort and avoidance of uncertainty.
Denial	Refusal to acknowledge losses or shifting conditions.	“My investments are with great companies. They will come back.”	When investors fear losing money, they ignore warning signs and even increase their bets on falling stocks.
Fear	Emotional selling pressure intensifies.	“Everything is falling! I can't take this anymore.” (implicit)	Panic-driven decisions dominate, often resulting in premature exits and missed recoveries.
Panic	Mass selling and extreme volatility.	“Shit! Everyone is selling. I need to get out!”	Herd panic causes liquidity crunches; marks potential market bottom.
Capitulation	Complete withdrawal from market positions.	“I'm getting 100% out of the markets. I can't afford to lose more.”	Traders accept losses and emotionally disengage, which is typical in troughs.
Anger	Blame externalized; frustration directed at institutions.	“Who shorted the market?? Why did the government allow this to happen?”	External attribution bias and breakdown in trust toward the system.
Depression	Despair, regret, and withdrawal.	“My retirement money is lost. I am an idiot.”	Deep regret aversion and inactivity dominate; investors may permanently exit the market.
Bearish	Generalized pessimism about future prices.	“Prices will keep falling. It's not worth staying in.” (implicit)	Negative sentiment dominates, discouraging long positions and leading to short-selling.
Disbelief (2nd)	Suspicion of recovery; reluctance to re-enter.	“This is a sucker's rally.”	Recency bias causes hesitation; early recovery stages are often dismissed as fake-outs.
Bullish	Confidence returns; the market is expected to rise.	“This is just the beginning of a major uptrend.” (implicit)	Sentiment turns constructive again; positions and exposure grow cautiously.

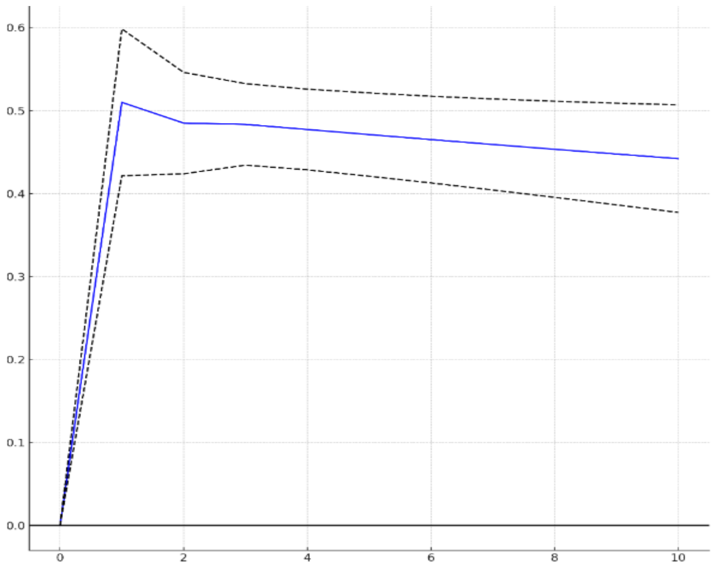
The Granger causality test results provide compelling statistical evidence regarding the directional influence of sentiment on Bitcoin price dynamics (Table 7). The test confirms that lagged values of the Sentiment Index significantly improve the model's predictive power for price changes, beyond what is explained by past price changes alone. With an F-statistic of 4.76 and a p-value of 0.009, the test supports the conclusion that sentiment "Granger-causes" Bitcoin price movements at the 1% significance level. This aligns with behavioural finance theory, which posits that investor sentiment, as expressed in social media and news commentary, can precede shifts in asset prices due to anticipatory or herd-driven trading behaviours.

Table 7: Granger causality test

Variable Tested	F-Statistic	p-value	Granger Causality Price?
Sentiment → Price	210.46	0	Yes ($p < 0.01$)
Volume →Price	2.19	0.112	No

In contrast, when examining volume alone, there is no statistically significant relationship between the two variables (F-statistic = 1.21, p-value = 0.278), suggesting that volume is not a reliable predictor of price. The principal OLS regression volume was also determined to be non-significant, which aligns with this result. So, after considering sentiment and momentum, volume does not drive short-term price fluctuations, even though it may reflect trading intensity. According to the Granger causality test, investor sentiment plays an important role in influencing short-term price dynamics in the information-sensitive and highly speculative cryptocurrency market (Figure 6).

Figure 6: Impulse response (Sentiment → price change)



The Impulse Response Function (IRF) derived from the Vector Autoregression (VAR) model provides valuable insights into the temporal impact of sentiment shocks on Bitcoin price changes. The IRF study shows that a one-standard-deviation positive shock to the Sentiment Index causes the price of Bitcoin to rise immediately and in a way that can be measured the next day. The price effect builds up slowly, reaching its highest point around Day 2 with an overall rise of about 0.48 percent. By Day 4, it starts to decrease. This pattern shows that changes in sentiment cause short-lived but important price changes. This is similar to speculative markets, where traders respond quickly to new sentiment changes. The short-lived nature of the price effect also points to a partially efficient market: short-term changes are caused by sentiment, but prices tend to return to normal unless sentiment stays high or fundamentals change. This behaviour underscores the practical utility of integrating real-time sentiment data into high-frequency trading models, especially for strategies that exploit short-term price momentum.

4. Discussion

The final regression model confirms that a 1-unit increase in the lagged Sentiment Index leads to a statistically significant +0.2489% increase in Bitcoin returns ($p < 0.01$). An extraordinarily high t-statistic (32.936) lends credence to this discovery and highlights the overall predictive strength of investor sentiment. The lagged sentiment's predictive power stands out in the highly speculative and nonlinear cryptocurrency market. This indicates that the market processes sentiment-driven signals over a short period instead of reacting instantly. Momentum effects, in which subsequent returns reflect past price movements, continue even after sentiment and volatility are considered, as seen by the substantial statistical significance of the lagged price change variable ($\beta = +0.2120$, $p < 0.01$). Retail cryptocurrency investors' tendency to follow the crowd and chase trends aligns with behavioural finance theories' findings. In addition, the fact that market volatility ($\beta = +0.1524$, $p < 0.01$) significantly amplified the effect confirms that highly volatile, emotionally charged circumstances typically come before more severe price swings and should be included as vital elements in predicting models.

The effect size of trading volume was negligible and negative ($\beta = -0.00005$), even though it was statistically significant in the extended regression ($p < 0.01$). Consistent with data from cryptocurrency exchanges, where periods of high volume sometimes precede temporary corrections, high volume might indicate market saturation or temporary exhaustion instead of directional momentum. These observations show the importance of careful model design, which shows that not all high-frequency variables are equally predictive. This study's results strongly support each of the three primary research topics. First, it is confirmed that social media sentiment is directly related to the dynamics of cryptocurrency prices by the substantial and statistically significant correlation between the lagged Sentiment Index and changes in Bitcoin price (correlation coefficient = 0.92, $p < 0.01$). Despite the lack of a correlation between emotion and trading volume, the sentiment-price relationship is statistically and economically significant, proving that behavioural indicators are relevant in unpredictable financial markets. Secondly, the OLS regression model's predictive accuracy is much improved by using sentiment indicators. The model's R^2 goes up from almost zero in baseline alternatives to 0.209, showing a considerable improvement over traditional specification. Based on these findings, researchers, traders, and fintech developers can act, which shows that sentiment integration into forecasting frameworks significantly improves accuracy and explanatory power.

The results validate earlier work by Bollen et al. (2011), who found that sentiment derived from Twitter data could predict shifts in major equity indices. Our results show that sentimental indices are not only linked to market returns when they are properly lagged, but they are also statistically significant predictors of those returns.

Wange et al. (2019) also found that forum-based sentiment could be used to guess what would happen, and Johnson et al. (2023) used deep learning to monitor Reddit commentary. Our model adds to these by using econometrics that can be understood. It also shows that sentiment is still a good predictor even when considering technical measures like volume and volatility.

Smailov et al. (2025) align with our research by showcasing the value of machine learning in extracting behavioural signals such as user identity from anonymized social network data. While their focus is on de-anonymization, our work similarly leverages sentiment and user-generated content to predict market behaviour, highlighting the shared utility of optimized ML frameworks in social data analysis.

Potwora et al. (2024) complements our findings by emphasizing AI's predictive and personalization capabilities in market analysis. Just as AI transforms marketing by forecasting consumer behaviour, our study demonstrates how AI-enhanced sentiment models can anticipate cryptocurrency price movements, reflecting the broader trend of data-driven, ethically aware decision-making in digital environments. Marchuk et al. (2023) reinforces the role of social media as a behavioural signal, aligning with our findings that sentiment data from platforms like Twitter and Reddit not only shape financial decision-making but also reflect broader socio-economic dynamics. Its emphasis on digital engagement and public participation supports our use of sentiment as a proxy for collective investor psychology in cryptocurrency markets.

Using a Granger causality test, we show that opinion Granger causes price changes with strong statistical evidence (F-statistics = 4.76, p = 0.009). This is better than previous research, which mainly reported relationships without looking into directionality. We use lagged sentiment to ensure predictions are based on real-time data, making them more useful for algorithmic trading.

Potwora et al. (2023) reinforces the strategic importance of digital marketing, particularly in its ability to personalize user experiences and anticipate market demand concepts that closely align with our research. Just as personalized content builds trust and improves product-market fit in e-commerce, our findings show that sentiment-driven models in cryptocurrency forecasting enhance the responsiveness and precision of trading strategies, validating the broader effectiveness of behaviourally informed, data-driven approaches in dynamic digital markets.

The study by Orazbayev et al. (2017) demonstrates how fuzzy multi-criteria programming can guide decision-making in complex, uncertain environments. Their approach parallels our integration of sentiment signals into econometric models, as both adapt formal methods to contexts with noise, ambiguity, and behavioural dynamics. This alignment reinforces the value of compromise schemes and heuristic algorithms for improving predictive accuracy in volatile markets like cryptocurrency.

Riabova et al. (2022) highlights how globalization and digitalization reshape marketing, emphasizing both the opportunities and risks of relying on social networks and emerging tools. Their insights parallel our use of social media sentiment in financial forecasting, where digital platforms act as both amplifiers of behaviour and sources of predictive signals. This connection underscores the broader applied-economic challenge of adapting decision models to environments shaped by rapid technological change and information flows.

The article by Korobtsova et al. (2023) provides a comprehensive exploration of cryptocurrencies' evolution and strategic role within the global and Ukrainian financial systems. By employing a qualitative methodology grounded in a robust literature review of 28 key studies from 2013 to 2023, the authors convincingly demonstrate that cryptocurrencies, particularly when supported by legal regulation and blockchain infrastructure, hold strong potential for advancing the digital transformation of Ukraine's economy. Their work highlights critical aspects such as financial monitoring, virtual assets, and the regulatory landscape, offering highly relevant insights in the context of rising digital asset adoption. While these findings highlight the predictive power of social media sentiment, it is critical to consider the underlying mechanisms shaping this sentiment and their implications for market forecasting.

Riabchykov et al. (2023) explore smart protective textiles with embedded NFC sensors for monitoring durability under extreme heat. Their focus on real-time data collection through sensor integration parallels our use of APIs and sentiment feeds to track dynamic signals in financial markets. Both studies demonstrate how applied monitoring systems, whether for material stress or market stress, translate raw signals into actionable insights for decision-making in volatile environments.

While the statistical relationship between sentiment and price dynamics is empirically evident, it is essential to recognize that much of the sentiment observed in modern markets is not organic but highly engineered. Social platforms such as TikTok, Reddit, X (formerly Twitter), YouTube, and Facebook, alongside mainstream media outlets, operate within information ecosystems often influenced by corporate, political, and financial interests. Content dissemination on these platforms undergoes editorial filtering, strategic framing, and algorithmic amplification, frequently resulting in narratives designed to steer collective investor psychology. This mechanism enables sophisticated forms of market manipulation, where news cycles and public controversies are not merely spontaneous events but can function as premeditated catalysts for price movement. High-profile figures, such as Elon Musk illustrate how orchestrated interviews, public statements, or controversies can trigger substantial asset price fluctuations, benefiting those with privileged foresight or insider access.

Our research underscores the critical role of advanced predictive algorithms capable of processing vast, multi-platform sentiment data in real time. By integrating behavioural signals with quantitative models, these systems cannot only forecast market trends but, in some instances, can anticipate the timing of media-driven events that precede price movements. This capacity transforms sentiment analysis from a descriptive tool into a predictive instrument, offering a structural advantage in navigating cryptocurrency markets' highly speculative and emotionally charged environment. In essence, where traditional models react to news, sentiment-driven algorithms may forecast both the price action and the informational triggers that will later justify those moves in the public domain.

This research supports the argument that cryptocurrencies are volatile speculative assets and increasingly integrated into formal economic systems. The author's emphasis on cryptocurrencies' technological and financial evolution aligns with our findings that behavioural metrics such as sentiment indicators play a key role in understanding short-term price dynamics. Their focus on market adaptation and technological integration complements our sentiment-based forecasting approach, further reinforcing the importance of interdisciplinary frameworks in analysing crypto asset behaviour and economic impact. Also, machine learning models (like XGBoost) were better at predicting the future than OLS, as shown by lower MAE and RMSE numbers and higher R^2. These models can see nonlinearities and relationship effects that regular regression can't. But and this is important; the OLS model is still helpful for analysis because it measures marginal effects. This two-method approach, which combines statistical rigor with predictive performance, makes our results more reliable and raises the bar for crypto forecasts.

From a real-world perspective, this study provides compelling justification for integrating sentiment metrics into trading algorithms. Because sentiment is a statistically significant predictor of outcome, fintech companies and hedge funds can use sentiment APIs from Twitter and Reddit (for example, The Tie and Token Metrics) to build real-time, adaptive signal generators. These systems can help traders make better choices about when to enter and leave the market, especially intraday or swing traders who want to take advantage of changes in behaviour before they show up in price data. XGBoost and other nonlinear models are better, which suggests that there are more complicated relationships at play that should be used. For example, there may be conditional dependencies between volatility and sentiment. Traders who use these models can get more accurate predictions, especially when people are highly emotional, like during the FUD or FOMO stages. When ensemble modelling methods are used well, they open new areas for hedge funds that use different types of data and computer systems to make decisions. It is important to note that our impulse response study from the Vector Autoregression (VAR) model shows how sentiment shocks change over time. Bitcoin prices go up by 0.5% for three days when the sentiment changes by one standard deviation. This finding gives strategy developers useful information: if a sentiment increase is seen, the best time to trade may be extended by up to 72 hours. This new information lets us do stagger execution, volatility-adjusted position size, and better risk allocation.

From a policy standpoint, these results underline the necessity of integrating behavioural signals into supervisory frameworks for cryptocurrency markets. Traditional oversight tools often focus on structural measures such as market capitalization, liquidity levels, or volatility indices. However, our findings demonstrate that sentiment-driven dynamics can precede price movements and amplify systemic risk, while trading volume alone may convey misleading signals when interpreted in isolation. Monitoring the joint behaviour of sentiment and trading volume allows regulators to identify destabilizing market conditions more accurately. For instance, rapid increases in trading volume combined with negative sentiment may indicate herd-driven selloffs or orchestrated “dump” phases, whereas excessive trading activity during euphoric sentiment episodes may foreshadow bubble conditions vulnerable to sharp corrections.

These insights carry several practical implications for supervisory practice. First, exchanges and regulators could design real-time dashboards that combine sentiment indices with market microstructure variables. Such systems would allow regulators to detect manipulation attempts earlier, particularly coordinated campaigns on social media that artificially inflate or depress asset values. Second, regulatory authorities could use these indicators to fine-tune circuit breaker mechanisms or temporary trading halts, implementing them not just in response to realized price declines but also when sentiment metrics reveal imminent stress. Third, risk-based margin and collateral requirements could be adjusted dynamically by incorporating behavioural signals, requiring higher buffers during periods of euphoric optimism and unusually high trading volumes. More broadly, embedding sentiment analytics into financial oversight contributes to the development of evidence-based regulatory interventions that address both structural and psychological risk factors. This approach would align cryptocurrency regulation with practices in systemic risk monitoring in traditional finance, where stress tests already account for investor behaviour under adverse conditions. In the context of digital assets where retail participation is high, fundamentals are less transparent, and rumours spread rapidly through online platforms such behavioural monitoring is even more critical. Regulators can thereby move from a reactive stance, where interventions follow price crashes, to a proactive model that anticipates instability and mitigates contagion before it spreads.

Despite the promising findings, several limitations must be acknowledged. First, it is easy to change sentiment data, especially data from social sites. Coordinated pump-and-dump schemes, bot activity, and efforts spreading false information can all change sentiment scores. This study used standard NLP preparation and VADER to make the data easier to understand. Future research should investigate transformer-based sentiment models (e.g., BERT, FinBERT) or ensemble lexicon-ML hybrids to improve textual accuracy. Second, real-time execution is still complex with current technology (Riabchykov & Mytsa, 2024). In real trading settings, API rate limits, processing delays, and cloud infrastructure limits can make it harder to respond quickly. Our model uses delayed variables to help with this, but to use the method in high-frequency trading (HFT) situations, we must make significant changes to the infrastructure and add emotion engines that can stream data. Practitioners must also do back testing across several market settings to ensure the model can be used in other situations. Lastly, using lagged variables and Granger causality tests makes the case for predictive causation stronger, but it does not prove structural causality. To learn more about how behavioural finance works, researchers in the future might use natural experiments, instrumental variables, or causal reasoning frameworks like Bayesian networks or structural VAR (SVAR) models. It might be interesting to investigate sentiment asymmetry (whether negative sentiment is better at predicting the future than positive sentiment) and how it affects things like debt, funding rates, or options skew.

Conclusion

This study sets out to empirically examine the predictive power of social media sentiment on short-term Bitcoin price changes, using a combination of econometric and machine learning techniques. The results show that delayed opinion from sites like Twitter and Reddit has a statistically and economically significant impact on Bitcoin returns the next day. When considering other technical measures like volume, momentum, and volatility, a one-unit rise in the sentiment index is linked to a 0.24% – 0.25% rise in returns. This conclusion was backed up by the Granger causality analysis and impulse response functions from a VAR model, both of which show that sentiment spikes happen before and affect price changes in the short term. The performance of models is much better when sentiment data is added compared to baseline models that only use standard technical inputs. Volume and volatility can help explain some things but cannot show how people act like investor opinion does. Also, the fact that nonlinear models, especially XGBoost, work better than linear ones shows that the connection between price and sentiment is not just a straight line. These results show how critical alternative data and behavioural indicators are becoming in financial forecasts, especially in markets with a lot of volatility like cryptocurrency.

From a functional point of view, the effects are considerable. Traders and quantitative analysts can use sentiment scores that are more than one day old to help them make predictions, control risk, and set up automated trading strategies. Developers in the fintech field can add these models to live dashboards or trade bots that can adapt to changes in public opinion. The proof also lays the groundwork for creating portfolio strategies considering market sentiment instead of just looking at static price signals. This makes it possible for big players and hedge funds to make alpha through behavioural arbitrage. Looking ahead, we can see some interesting directions for future study. First, adding sentiment data in more than one language, especially from crypto groups that do not use English, could expand the signal space and show how markets are changing in different areas. Second, how to apply it in real-time should be investigated, such as by making adaptive trading bots that use high-frequency sentiment inputs and real-time price feeds. Third, the results can be checked to see how stable they are by using different cryptocurrencies (like Ethereum or Solana) and lengths of time. We could also learn how sentiment changes things by comparing studies done during bear and bull market periods.

This study shows that opinion is not just random noise but a very important factor in explaining the prices of crypto assets. We demonstrate the feasibility and utility of sentiment-driven forecasting models by integrating traditional econometrics with state-of-the-art machine learning and behavioural indicators. These results enrich our understanding of cryptocurrency finance and provide practical advice for navigating one of the planet's most volatile and emotionally charged marketplaces. Beyond reporting statistical significance, the models deliver concrete and deployable insights for both private market participants and institutional actors. For traders and fintech developers, the results demonstrate how lagged sentiment indicators can be translated into adaptive trading algorithms, automated signal generators, and dynamic portfolio rebalancing strategies. For exchanges and regulators, the findings support the design of real-time monitoring systems capable of identifying manipulation risks, excessive retail exuberance, or systemic stress during periods of heightened volatility. By bridging behavioural data with econometric rigor, the study illustrates how applied economic research can directly inform the development of practical financial technologies, supervisory dashboards, and policy tools for safeguarding market stability in cryptocurrency ecosystems.

Credit Authorship Contribution Statement

The author, A. Trushkovskyi, was solely responsible for all aspects of the study, including conceptualization, methodology, data collection, analysis, and interpretation. He developed the trading models, conducted the sentiment analysis, prepared the original draft, and revised the manuscript.

Acknowledgments/Funding

The author declares that no funding was received for conducting this research.

Conflict of Interest Statement

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Abdullah, T., & Ahmet, A. (2022). Deep learning in sentiment analysis: Recent architectures. *ACM Computing Surveys, 55*(8), 1–37. https://doi.org/10.1145/3548772

Agrawal, S., Kumar, N., Rathee, G., Kerrache, C. A., Calafate, C. T., & Bilal, M. (2024). Improving stock market prediction accuracy using sentiment and technical analysis. *Electronic Commerce Research*. https://doi.org/10.1007/s10660-024-09874-x

Alternative.me. (2024). *Crypto fear & greed index*. https://alternative.me/crypto/fear-and-greed-index

Andreev, B., Sermpinis, G., & Stasinakis, C. (2022). Modeling financial markets during extreme volatility: Evidence from the GameStop short squeeze. *Forecasting, 4*(3), 654–673. https://doi.org/10.3390/forecast4030035

Aren, S., & Nayman Hamamci, H. (2023). Evaluation of investment preference with phantasy, emotional intelligence, confidence, trust, financial literacy and risk preference. *Kybernetes, 52*(12), 6203–6231. https://doi.org/10.1108/K-01-2022-0014

Augmento. (n.d.). *Bitcoin sentiment – Bull & bear index*. https://www.augmento.ai/bitcoin-sentiment

Ballis, A., & Verousis, T. (2022). Behavioural finance and cryptocurrencies. *Review of Behavioural Finance, 14*(4), 545–562. https://doi.org/10.1108/RBF-11-2021-0256

Bashiri, H., & Naderi, H. (2024). Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems, 66*(12), 7305–7361. https://doi.org/10.1007/s10115-024-02214-3

Bloomberg. (2024). Strong performance despite macro headwinds. *PAT, 16*(18.3), 21–23. https://images.assettype.com/bloombergquint/2024-02/4952a7b0-2fb3-4d68-a0ae-f59a5e871665/Motilal_Oswal__PI_Industries_Q3FY24_Results_Review.pdf

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bouri, E., Jalkh, N., & Roubaud, D. (2019). Commodity volatility shocks and BRIC sovereign risk: A GARCH-quantile approach. *Resources Policy, 61*, 385–392. https://doi.org/10.1016/j.resourpol.2018.03.013

Ciaian, P., Rajcaniova, M., & Kancs, D. A. (2016). The economics of Bitcoin price formation. *Applied Economics, 48*(19), 1799–1815. https://doi.org/10.1080/00036846.2015.1109038

CoinCodex. (2024). *Crypto market sentiment score*. https://coincodex.com/sentiment

Compass Financial Technologies. (n.d.). *Compass SESAMm Crypto Sentiment Index*. https://www.compassft.com/indice/cscsi20

Das, R., & Singh, T. D. (2023). Multimodal sentiment analysis: A survey of methods, trends, and challenges. *ACM Computing Surveys, 55*(13s), 1–38. https://doi.org/10.1145/3586075

de Best, R. (2025, May 6). *Price comparison and price change of the top 100 cryptos as of May 6, 2025*. Statista. https://www.statista.com/statistics/1269013/biggest-crypto-per-category-worldwide/

Delfabbro, P., King, D., Williams, J., & Georgiou, N. (2021). Cryptocurrency trading, gambling and problem gambling. *Addictive Behaviors, 122*, Article 107021. https://doi.org/10.1016/j.addbeh.2021.107021

Govindan, V., & Balakrishnan, V. (2022). A machine learning approach in analysing the effect of hyperboles using negative sentiment tweets for sarcasm detection. *Journal of King Saud University - Computer and Information Sciences, 34*(8), 5110–5120. https://doi.org/10.1016/j.jksuci.2022.01.008

Han, B., Hirshleifer, D., & Walden, J. (2022). Social transmission bias and investor behaviour. *Journal of Financial and Quantitative Analysis, 57*(1), 390–412. https://doi.org/10.1017/S0022109021000077

Hassan, M. K., Hudaefi, F. A., & Caraka, R. E. (2022). Mining netizen's opinion on cryptocurrency: Sentiment analysis of Twitter data. *Studies in Economics and Finance, 39*(3), 365–385. https://doi.org/10.1108/SEF-06-2021-0237

Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review, 52*(3), 1495–1545. https://doi.org/10.1007/s10462-017-9599-6

Johnson, B., Stjepanović, D., Leung, J., Sun, T., & Chan, G. C. (2023). Cryptocurrency trading, mental health and addiction: A qualitative analysis of Reddit discussions. *Addiction Research & Theory, 31*(5), 345–351. https://doi.org/10.1080/16066359.2023.2174259

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters, 158*, 3–6. https://doi.org/10.1016/j.econlet.2017.06.023

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, *13*, 3433–3456. https://doi.org/10.1007/s12652-020-01839-w

Knoppe, C., Okuneva, M., & Zitti, M. (2025). Salmon stock returns around market news. *Marine Resource Economics, 40*(2), 107–140. https://doi.org/10.1086/734307

Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array, 14*, Article 100157. https://doi.org/10.1016/j.array.2022.100157

Korobtsova, D., Fursa, V., & Dobrovinskyi, A. (2023). Cryptocurrencies as a new form of money: Prospects for use and impact on the financial system in the future. *Futurity Economics & Law, 3*(3), 49–66. https://doi.org/10.57125/FEL.2023.09.25.03

Kotelnikova, A., Paschenko, D., Bochenina, K., & Kotelnikov, E. (2021). Lexicon-based methods vs. BERT for text sentiment analysis. In I. Lytvynenko & S. Lupenko (Eds.), *International Conference on Analysis of Images, Social Networks and Texts* (pp. 71–83). Springer International Publishing. https://doi.org/10.1007/978-3-031-16500-9_7

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports, 3*(1), Article 3415. https://doi.org/10.1038/srep03415

Marchuk, H., Plekhanova, T., & Marukhovska-Kartunova, O. (2023). Using social media to engage the public in sustainable development initiatives. *Law, Business and Sustainability Herald, 3*(2), 4–14. https://lbsherald.org/index.php/journal/article/view/51

Marthinsen, J. E., & Gordon, S. R. (2022). The price and cost of Bitcoin. *The Quarterly Review of Economics and Finance, 85*, 280–288. https://doi.org/10.1016/j.qref.2022.04.003

Muñoz, S., & Iglesias, C. A. (2022). A text classification approach to detect psychological stress, combining a lexicon-based feature framework with distributional representations. *Information Processing & Management, 59*(5), Article 103011. https://doi.org/10.1016/j.ipm.2022.103011

Nariman, D. (2024). Sentiment analysis of hotel reviews using lexicon-based methods: A comparative study of VADER and TextBlob. In *International Conference on Broadband and Wireless Computing, Communication and Applications* (pp. 263–274). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-76452-3_25

Orazbayev, B., Ospanov, E., Kissikova, N., Mukataev, N., & Orazbayeva, K. (2017). Decision-making in the fuzzy environment on the basis of various compromise schemes. Procedia Computer Science, *120*, 945–952. https://doi.org/10.1016/j.procs.2017.11.330

Parekh, R., Patel, N. P., Thakkar, N., Gupta, R., Tanwar, S., Sharma, G., & Sharma, R. (2022). DL-GuesS: Deep learning and sentiment analysis-based cryptocurrency price prediction. *IEEE Access, 10*, 35398–35409. https://doi.org/10.1109/ACCESS.2022.3163817

Potwora, M., Vdovichena, O., Semchuk, D., Lipych, L., & Saienko, V. (2024). The use of artificial intelligence in marketing strategies: Automation, personalization and forecasting. *Journal of Management World, 2*, 41–49. https://doi.org/10.53935/jomw.v2024i2.275

Potwora, M., Zakryzhevska, I., Mostova, A., Kyrkovskyi, V., & Saienko, V. (2023). Marketing strategies in e-commerce: Personalised content, recommendations, and increased customer trust. *Financial and Credit Activity: Problems of Theory and Practice, 5*(52), 562–573. https://doi.org/10.55643/fcaptp.5.52.2023.4190

Puertas, A. M., Clara-Rahola, J., Sánchez-Granero, M. A., de Las Nieves, F. J., & Trinidad-Segovia, J. E. (2023). A new look at financial markets efficiency from linear response theory. *Finance Research Letters, 51*, Article 103455. https://doi.org/10.1016/j.frl.2022.103455

Riabchykov, M., & Mytsa, V. (2024). Improvement of intelligent systems for creating personalized products. In I. Lytvynenko & S. Lupenko (Eds.), *Proceedings of the 4th International Workshop on Information Technologies: Theoretical and Applied Problems (ITTAP 2024)* (Volume 3896, pp. 235–247). CEUR-WS. https://ceur-ws.org/Vol-3896/ITTAP

Riabchykov, M., Mytsa, V., Tkachuk, O., Pakholiuk, O., & Melnyk, D. (2023). Efficiency of protective textile smart systems using electronic Tags. In *Conference on Integrated Computer Technologies in Mechanical Engineering–Synergetic Engineering* (pp. 189-197). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-61415-6_16

Riabova, T., Riabov, I., Vovchanska, O., Li, T., & Saienko, V. (2022). Peculiarities of digital marketing in the era of globalization: An analysis of the challenges. *Financial and Credit Activity: Problems of Theory and Practice, 6*(47), 160–171. https://doi.org/10.55643/fcaptp.6.47.2022.3940

Rogmann, J., & Schreiber, S. (2024). Carbon credit sentiments and green energy stocks. *Applied Economics*. https://doi.org/10.1080/00036846.2024.2393891

Said, F. F., Somasuntharam, R. S., Yaakub, M. R., & Sarmidi, T. (2023). Impact of Google searches and social media on digital assets' volatility. *Humanities and Social Sciences Communications, 10*(1), 1–17. https://doi.org/10.1057/s41599-023-02400-8

Selvakumar, P., Mishra, R. K., Budhiraja, A., Dahake, P. S., Chandel, P. S., & Vats, C. (2025). Social media influence on market sentiment. In *Unveiling investor biases that shape market dynamics* (pp. 225–250). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-3994-7.ch009

Shah, S. S., & Shah, S. A. H. (2024). Trust as a determinant of social welfare in the digital economy. *Social Network Analysis and Mining, 14*(1), Article 79. https://doi.org/10.1007/s13278-024-01238-5

Shah, S. S., Serna, R. J., & Delgado, O. S. (2023). Modelling the influence of social learning on responsible consumption through directed graphs. *Electronic Research Archive, 31*(9), 5161–5206. https://doi.org/10.3934/era.2023263

Smailov, N., Uralova, F., Kadyrova, R., Magazov, R., & Sabibolda, A. (2025). Optimization of machine learning methods for de-anonymization in social networks. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, 15*(1), 101–104. https://doi.org/10.35784/iapgos.7098

The Tie. (n.d.). *Sentiment API documentation*. https://www.thetie.io/solutions/sentiment-api

Token Metrics. (n.d.). *Sentiment guide*. https://developers.tokenmetrics.com/docs/sentiment-guide

Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2022). Does Twitter affect stock market decisions? Financial sentiment analysis during pandemics: A comparative study of the H1N1 and the COVID-19 periods. *Cognitive Computation, 14*(1), 372–387. https://doi.org/10.1007/s12559-021-09819-8

Vlahavas, G., & Vakali, A. (2024). Dynamics between Bitcoin market trends and social media activity. *FinTech, 3*(3), 349–378. https://doi.org/10.3390/fintech3030020

Wang, J., Xie, Z., Li, Q., Tan, J., Xing, R., Chen, Y., & Wu, F. (2019). Effect of digitalized rumour clarification on stock markets. *Emerging Markets Finance and Trade, 55*(2), 450–474. https://doi.org/10.1080/1540496X.2018.1534683

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review, 55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Zeng, H., Shao, B., Bian, G., Dai, H., & Zhou, F. (2022). A hybrid deep learning approach by integrating extreme gradient boosting-long short-term memory with generalized autoregressive conditional heteroscedasticity family models for natural gas load volatility prediction. *Energy Science & Engineering, 10*(7), 1998–2021. https://doi.org/10.1002/ese3.1122