International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Bitcoin Sentiment Analysis using Machine Learning Algorithms

Adit Vakil[1], Manavi Jain[2], Shweta Joshi[3], Prof. Abhilasha Raghtate[4]

[1]U.G. Student, Department of Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai, 400056, Maharashtra, India.
[2] U.G. Student, Department of Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai, 400056, Maharashtra, India.
[3] U.G. Student, Department of Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai, 400056, Maharashtra, India.
[4] U.G. Assistant Professor, Department of Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai, 400056, Maharashtra, India.

---***---

Abstract - The rapid rise of cryptocurrencies, particularly Bitcoin, has intensified research interest in understanding the factors influencing their price movements. Among these factors, social media sentiment has emerged as a crucial predictor, reflecting collective investor mood and market expectations. This paper provides a comprehensive survey of various sentiment analysis models applied in cryptocurrency markets, with a specific focus on the relationship between social media sentiment and Bitcoin price fluctuations. The study identifies the Aigents model as the most effective, showing significant improvements in predictive accuracy following fine-tuning. Findings reveal a predictive association between sentiment measures and price changes, typically with a latency of one to two days. The paper offers insights into the capabilities and limitations of existing Natural Language Processing (NLP) models in the context of cryptocurrency sentiment analysis, presenting practical implications for investors and analysts in navigating the volatile cryptocurrency markets.

Key Words: cryptocurrency, Bitcoin, social media sentiment, natural language processing (NLP), sentiment analysis, price prediction, machine learning models, Aigents model, financial forecasting, Twitter, Reddit, artificial neural networks (ANN), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Random Forest, Naive Bayes, Extreme Gradient Boosting (XGBoost), predictive analytics, market behavior, interpretable AI.

1.INTRODUCTION

The rise of cryptocurrencies, particularly Bitcoin, has led to increased interest in understanding the factors that drive their price movements. Among these factors, social media sentiment has emerged as a critical predictor, reflecting collective investor mood and market expectations. As social media platforms like Twitter and Reddit have become key sources of market sentiment, Natural Language Processing (NLP) models are increasingly employed to analyse and predict the impact of public sentiment on cryptocurrency prices.

This literature survey focuses on evaluating various sentiment analysis models used in cryptocurrency markets, with a specific emphasis on the relationship between social media sentiment and Bitcoin price fluctuations. The study aims to identify the most effective sentiment analysis techniques and improve their interpretability and accuracy for real-world applications. By assessing over 20 different models, the survey highlights the Aigents model as the leading performer, demonstrating significant improvements in predictive accuracy after fine-tuning. The findings underscore the predictive association between sentiment measures and price changes, typically observed with a latency of one to two days.

Through a systematic approach involving literature review, data collection, model evaluation, and refinement, this survey provides valuable insights into the capabilities and limitations of existing NLP models in the context of cryptocurrency sentiment analysis. The study ultimately seeks to enhance the understanding of how social media feedback influences market behaviour, offering practical implications for investors and analysts aiming to navigate the volatile landscape of cryptocurrency markets.

2. LITERATURE SURVEY

The study by [1] evaluates various models for natural language processing used in cryptocurrency sentiment analysis, focusing on the relationship between sentiment measures and changes in Bitcoin prices. After analysing over 20 sentiment models, the Aigents model emerged as the most successful, highlighting the importance of interpretable AI in improving sentiment analysis accuracy. The study examines the causal relationships between price fluctuations and sentiment measures, indicating a predictive association with a latency of one to two days. The correlation between the ground truth and the improved Aigents model sentiment score increased from 0.33 to 0.57 after fine-tuning.
Research by [2] explores the predictive potential of social media sentiment on stock market and cryptocurrency developments, using stochastic artificial neural networks (ANNs) and NLP to analyse sentiment on Twitter. The findings suggest that sentiment analysis for cryptocurrencies offers higher predictive accuracy than for equities. Among neural network models frequently employed in financial forecasting, Long Short-Term Memory (LSTM) Networks outperform other models in Bitcoin price prediction, while Gated Recurrent Units (GRUs) and

hybrid models combining LSTM with Convolutional Neural Networks (CNNs) further enhance prediction performance.
Another study [3] investigates the use of supervised learning techniques to classify sentiment and forecast changes in the prices of Ethereum, Litecoin, and Bitcoin. Among the classifiers used, Bernoulli Naive Bayes was identified as the best for Ethereum, and logistic regression performed best for Bitcoin and Litecoin. The study demonstrates moderate accuracy in predicting upward price movements, with a notable tendency to accurately forecast days with higher percentage gains.
In addition, research by [4] discusses a technique for classifying opinions found in text as neutral, negative, or positive. It involves gathering data, preparing text, identifying sentiment, and classifying data using lexicon-based, machine learning, and hybrid techniques. The study emphasizes the importance of big data in improving analysis accuracy and provides an overview of various sentiment analysis methodologies.

3. CRYPTOCURRENCY

Cryptocurrency is a digital or virtual currency that uses cryptography for secure transactions, making it nearly impossible to counterfeit or double-spend. Unlike traditional currencies controlled by central authorities, cryptocurrencies operate on decentralized networks based on blockchain technology, a distributed ledger maintained by a network of computers.
Cryptocurrencies are used for various purposes, including online purchases, investments, and remittances, providing faster, cheaper, and more secure alternatives to traditional banking systems. They also empower decentralized finance (DeFi) applications, allowing users to borrow, lend, and trade assets without intermediaries, thereby democratizing financial services on a global scale.

4. BITCOIN

Bitcoin, introduced in 2009 by the pseudonymous Satoshi Nakamoto, was the first cryptocurrency and remains the most recognized today. Conceived during the 2008 financial crisis, Bitcoin was designed as an alternative to traditional financial systems, offering a peer-to-peer electronic cash system that facilitates secure, transparent, and borderless transactions without reliance on central banks or governments.

The primary purpose of Bitcoin is to enable direct, trust less transactions between individuals, eliminating intermediaries and reducing transaction costs. Beyond this, it serves as a store of value—often called "digital gold"—and is widely used as an investment vehicle, a hedge against inflation, and a tool for financial inclusion, especially for those without access to conventional banking.

5. THE ROLE OF SENTIMENT ANALYSIS IN THE CRYPTOCURRENCY MARKET

Sentiment analysis of Bitcoin is a crucial tool in understanding this highly speculative asset, whose price is often influenced more by market sentiment, news, and social media discussions than by traditional economic indicators. By analysing public emotions expressed on platforms like Twitter, sentiment analysis provides a snapshot of the market's mood—whether optimistic, pessimistic, or neutral—towards Bitcoin at any given time.
This real-time insight into market psychology helps investors, traders, and analysts make informed decisions, identifying potential price movements driven by public perception. Positive sentiment often correlates with price increases as enthusiasm drives more people to invest, while negative sentiment can signal downturns as fear or uncertainty spreads.

6. IMPORTANCE OF SENTIMENT ANALYSIS FOR BITCOIN

The relevance of sentiment analysis lies in its ability to provide a predictive edge in the volatile cryptocurrency market, allowing stakeholders to anticipate shifts and adjust their strategies accordingly. By understanding how public perception shapes the cryptocurrency landscape, investors can navigate the complexities of the market with greater confidence, making sentiment analysis an invaluable tool in the world of digital currencies.

6.1 Predictive Edge in Volatile Markets

The cryptocurrency market is notoriously volatile, with prices often experiencing sharp fluctuations within short time frames. Traditional financial metrics and indicators may not always capture the rapid shifts in market sentiment that drive these price changes. Sentiment analysis provides a predictive edge by offering real-time insights into the public's emotional and psychological state towards Bitcoin. By analysing sentiment data from sources such as social media platforms, news articles, and forums, stakeholders can gauge the market's mood—whether it is optimistic, pessimistic, or neutral. This understanding enables investors and traders to anticipate potential price movements and adjust their strategies accordingly.

6.2 Influences on Market Behaviour

Bitcoin, as a decentralized digital asset, is heavily influenced by collective market behaviour and public perception. Positive sentiment, such as bullish forecasts, endorsements from influential figures, or favourable news, can drive prices upward as it fuels investor enthusiasm and attracts new participants to the market. Conversely, negative sentiment, including fear of regulatory crackdowns, security concerns, or adverse news, can lead to price declines as it induces caution and withdrawal from the market. Sentiment analysis helps in identifying these trends by quantifying the prevailing mood and its potential impact on Bitcoin's price, thus enabling more informed decision-making.

6.3 Enhancing Investment Strategies

For investors, sentiment analysis serves as a valuable tool for enhancing investment strategies. By integrating sentiment data with traditional financial analysis, investors can develop a more comprehensive view of the market. This approach helps in identifying sentiment-driven trends that might not be apparent from technical or fundamental analysis alone. For instance, a surge in positive sentiment may precede a price rally, while a

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

rise in negative sentiment might indicate an impending downturn. By leveraging sentiment analysis, investors can make timely decisions, optimize their portfolios, and potentially capitalize on market opportunities that arise from sentiment-driven movements.

6.4 Strategic Decision-Making

Traders and analysts use sentiment analysis to gain a strategic advantage in the cryptocurrency market. Understanding how sentiment shifts in response to news events, regulatory announcements, or market trends can help in predicting short-term price movements and identifying potential risks. For example, a sudden increase in negative sentiment following a regulatory announcement might signal a short-term decline in Bitcoin's price, prompting traders to adjust their positions. Conversely, a positive sentiment spike due to a major partnership or technological advancement might present a buying opportunity. By incorporating sentiment analysis into their decision-making processes, market participants can better navigate the complexities of the cryptocurrency landscape.

6.5 Mitigating Risks

Sentiment analysis also plays a role in risk management. In a market as volatile as cryptocurrencies, understanding the prevailing sentiment can help in mitigating risks associated with sudden price swings. By monitoring sentiment trends and being aware of potential shifts in market psychology, investors can take pre-emptive measures to protect their investments. For example, if sentiment analysis indicates increasing fear or uncertainty among investors, it may be prudent to adopt a more conservative strategy or hedge against potential losses.

7. MODEL USED

7.1 Support Vector Machine (SVM)

Support Vector Machines (SVMs) are a potent supervised machine learning technique widely used for tasks such as regression, classification, and outlier detection. The primary goal of SVM is to find the optimal hyperplane in a high-dimensional feature space to best divide data points of different classes. This hyperplane maximizes the margin, or distance, between the data points of different classes, leading to better generalization on unseen data.
SVMs are particularly effective in scenarios where the number of features is high relative to the number of data points, and they can handle both linear and non-linear classification tasks using kernel functions. Popular kernel functions include linear, polynomial, and radial basis function (RBF) kernels, which transform the data into higher dimensions where a linear separation is possible. SVMs are known for their robustness and accuracy in high-dimensional spaces, making them suitable for various applications, including text classification and image recognition.

7.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a subclass of deep learning algorithms particularly useful for processing structured grid-like input, such as images and videos. CNNs consist of several layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply filters to the input data to automatically and adaptively learn spatial hierarchies of features, such as edges, textures, and patterns. Pooling layers reduce the spatial dimensions of the data, which helps to minimize computational complexity and reduce overfitting.
In natural language processing (NLP), CNNs have been adapted for tasks such as sentiment analysis. They can capture local dependencies and extract meaningful features from text data by treating text sequences as one-dimensional grids. CNNs are advantageous for sentiment analysis because they can learn feature representations from raw text and are effective in detecting patterns across different contexts.

7.3 Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks

Recurrent Neural Networks (RNNs) are designed to recognize patterns in sequences of data, making them well-suited for tasks involving sequential or time-series data, such as financial forecasting and NLP. RNNs have a unique architecture that includes loops, allowing them to retain information from previous steps and use it to influence future steps. This design makes RNNs capable of capturing temporal dependencies in sequential data.
However, traditional RNNs face challenges such as vanishing and exploding gradients, which make it difficult to learn long-range dependencies. Long Short-Term Memory (LSTM) networks, a variant of RNNs, address these issues with their memory cell mechanism. LSTMs include gates (input, forget, and output gates) that regulate the flow of information into, out of, and within the memory cell. This architecture allows LSTMs to capture long-term dependencies more effectively, making them particularly valuable for tasks like sentiment analysis where understanding context over long sequences is crucial.

7.4 Random Forest

Random Forest is an ensemble learning technique known for its robustness and versatility, particularly useful for regression and classification problems. It operates by constructing multiple decision trees during training and combining their outputs to enhance predictive accuracy. Each decision tree in the Random Forest is built using a random subset of the training data and features, a process known as bagging or bootstrap aggregating.
The final prediction is made by aggregating the predictions from all individual decision trees, typically through majority voting for classification tasks or averaging for regression tasks. Random Forest is effective in handling high-dimensional data and complex interactions between features. It is also known for its ability to provide feature importance estimates, which can be useful for understanding the contribution of different features to the model’s predictions.

7.5 Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' Theorem, which assumes conditional independence between features given the class label. Despite its simplicity, Naive Bayes is particularly effective in text classification tasks such as spam detection and sentiment analysis due to its ability to handle high-dimensional and sparse data efficiently.
There are several variants of Naive Bayes, including Gaussian Naive Bayes, which assumes that features follow a Gaussian distribution, and Multinomial and Bernoulli Naive Bayes, which are suitable for categorical data. The algorithm computes the posterior probability of each class based on the likelihood

of the features and the prior probability of the class, making it straightforward and fast to train. Naive Bayes is valued for its simplicity, interpretability, and efficiency, especially when dealing with large datasets.

7.6 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a powerful and efficient implementation of the gradient boosting framework, widely used for regression, classification, and ranking problems. XGBoost builds an ensemble of decision trees in a sequential manner, where each tree corrects the errors made by the previous trees. This iterative process enhances the model's predictive accuracy.

XGBoost includes several enhancements over traditional gradient boosting methods, such as support for parallelization, which speeds up the training process, and regularization techniques (L1 and L2) that help prevent overfitting. It also features a novel split-finding algorithm based on weighted quantile sketch, which improves the handling of sparse data. XGBoost's flexibility, high performance, and accuracy have made it a popular choice in machine learning competitions and real-world applications across various domains, including finance and healthcare.

8. METHODOLOGY

This study explores the critical role of public sentiment in influencing Bitcoin's market dynamics through an in-depth analysis of tweets related to Bitcoin. As the first and most widely recognized cryptocurrency, Bitcoin operates independently of traditional financial systems, making it highly susceptible to public opinion and social media discourse, particularly on platforms like Twitter. Unlike traditional assets, Bitcoin's value is not tethered to economic indicators or financial statements but is instead driven largely by market sentiment and speculative behaviour. This makes sentiment analysis—a subset of natural language processing (NLP)—a powerful tool for predicting Bitcoin's market movements by classifying text into positive, negative, or neutral sentiments.

The study employs an extensive set of machine learning and deep learning algorithms, including Support Vector Machine (SVM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), XGBoost, and Long Short-Term Memory (LSTM). These algorithms were selected for their ability to handle the complexities of textual data and their proven effectiveness in sentiment classification tasks. SVM is known for its robust classification capabilities, while CNN and RNN are particularly effective in extracting features from sequential data, such as text. Random Forest and Naive Bayes offer simplicity and speed, which are beneficial for initial sentiment analysis. XGBoost, a more advanced ensemble technique, improves accuracy and performance, while LSTM excels in capturing long-term dependencies in textual data, making it particularly valuable for understanding sentiment trends over time.

By collecting and analysing a large dataset of Bitcoin-related tweets, this study aims to compare the performance of these algorithms in predicting sentiment, evaluating their strengths and weaknesses in capturing the nuances of public opinion. The findings reveal how fluctuations in public sentiment—whether triggered by news, influential personalities, or market events—can significantly impact Bitcoin's price, often leading to rapid and sometimes unpredictable market movements. Positive sentiments, such as bullish comments or endorsements from high-profile figures, can drive prices upward, while negative sentiments, like fear-driven tweets or regulatory concerns, can lead to sharp declines.

This research provides valuable insights into the psychology of the Bitcoin market, demonstrating the profound influence of social media-driven sentiment on trading behaviours. For investors, traders, and analysts, understanding the sentiment landscape offers a strategic advantage, enabling them to anticipate market shifts and make more informed decisions. The study underscores the growing importance of sentiment analysis as a predictive tool in financial markets, particularly in the volatile and sentiment-driven world of cryptocurrencies. As Bitcoin and other digital assets continue to evolve, integrating sentiment analysis into trading strategies will likely become increasingly essential for market participants seeking to navigate the complex and often erratic behaviour of crypto markets.

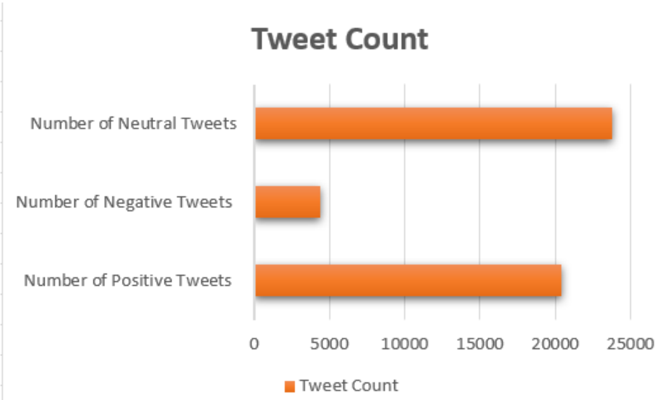

Fig -1: Tweet count showing clear demarcation between positive, negative and neutral tweets

9. RESULT

The comparative analysis of various machine learning models for sentiment classification of Bitcoin-related tweets yielded insightful results regarding their effectiveness in predicting market sentiment. The following models were evaluated: Support Vector Machine (SVM), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Random Forest, Naive Bayes, and Extreme Gradient Boosting (XGBoost). Each model's performance was assessed based on its accuracy in classifying sentiment from the collected dataset of tweets.

9.1 Support Vector Machine (SVM):

The SVM model achieved an accuracy of 93.904%. This performance highlights the SVM's ability to create an optimal hyperplane for separating sentiment classes in high-dimensional feature space. SVM's effectiveness is attributed to its robust classification capabilities, which are particularly advantageous in handling complex and high-dimensional data often encountered in sentiment analysis.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

9.2 Convolutional Neural Networks (CNNs):

CNNs demonstrated the highest accuracy of 94.301%. This superior performance underscores CNNs' strength in automatically learning and capturing the spatial hierarchies of features from the input text data. CNNs' ability to detect and learn patterns from sequential data, such as the context of words in tweets, contributes significantly to their high classification accuracy. The CNN's adaptive learning mechanism allows it to identify nuanced sentiment expressions effectively.

9.3 Recurrent Neural Networks (RNNs):

The RNN model achieved an accuracy of 92.960%. RNNs are designed to handle sequential data and are well-suited for capturing the temporal dependencies in text. Despite their advantages in processing sequences, RNNs generally face challenges with long-term dependencies, which can impact their performance in capturing more complex sentiment patterns over extended sequences.

9.4 Random Forest:

The Random Forest model recorded an accuracy of 87.871%. As an ensemble learning technique, Random Forest combines the predictions of multiple decision trees to improve overall performance and robustness. While effective, Random Forest's performance in sentiment analysis is somewhat limited compared to models that excel in handling textual data and learning contextual nuances.

9.5 Naive Bayes:

The Naive Bayes model exhibited an accuracy of 82.347%. This probabilistic classifier, based on Bayes' Theorem, assumes conditional independence between features. While Naive Bayes is efficient and straightforward, its assumption of feature independence can be a limitation in capturing the complex and often interdependent nature of sentiment in text data.

9.6 Extreme Gradient Boosting (XGBoost):

XGBoost achieved an accuracy of 91.511%. Known for its effectiveness in various machine learning tasks, XGBoost benefits from its gradient boosting framework, which incrementally improves model performance by focusing on errors made by previous models. Despite its strong performance, XGBoost's complexity and computational demands can be higher compared to simpler models.
The comparative results highlight CNNs as the most effective model for sentiment classification in this study, reflecting their superior ability to handle and analyse the hierarchical and contextual features of textual data. SVM and RNN also demonstrated strong performance, indicating their robustness in sentiment analysis tasks. However, Random Forest, XGBoost, and Naive Bayes, while useful, showed comparatively lower accuracy, suggesting areas for potential improvement.
These findings emphasize the critical role of selecting appropriate machine learning models based on the specific characteristics of the data and the task at hand. The effectiveness of CNNs in sentiment analysis underscores the importance of leveraging advanced models that can capture the intricate patterns and dependencies in textual data to enhance predictive accuracy.

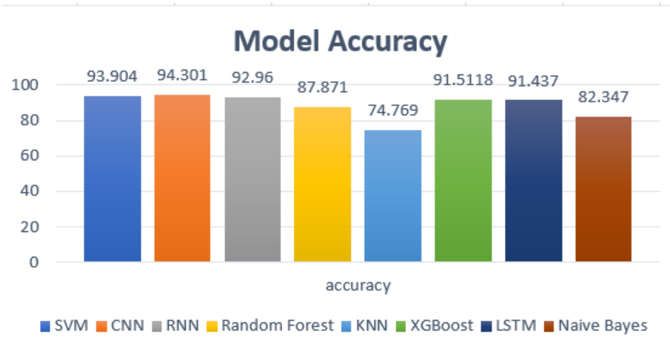

Fig -2: A comparative representation of various models utilized and their respective accuracies

Sr No.	Models Used	Accuracies
1	SVM	93.904
2	CNN	94.301
3	RNN	92.96
4	Random Forest	87.871
5	KNN	74.769
6	XGBoost	91.5118
7	LSTM	91.437
8	Naïve Bayes	82.347

Table -1: Models used and their results accuracy

10. CONCLUSIONS

This study demonstrates the significant role of sentiment analysis in understanding Bitcoin's market dynamics by evaluating several machine learning models on their ability to classify sentiment from Twitter data. The results reveal that Convolutional Neural Networks (CNNs) are particularly effective for sentiment classification tasks, achieving the highest accuracy among the models tested. The high performance of CNNs can be attributed to their ability to automatically learn and adapt to the hierarchical features of text data, making them well-suited for capturing nuanced sentiments expressed in tweets.
Support Vector Machine (SVM) and Recurrent Neural Networks (RNNs) also showed strong performance, confirming their utility in sentiment analysis tasks. However, Random Forest, XGBoost, and Naive Bayes, while useful, displayed comparatively lower accuracy, indicating their limitations in this context.
The findings underscore the importance of leveraging advanced machine learning techniques to gain predictive insights into Bitcoin's price movements influenced by public sentiment. By incorporating sentiment analysis into trading strategies, investors and analysts can better navigate the volatile cryptocurrency market, making more informed decisions based on the prevailing public mood and trends.
Overall, sentiment analysis proves to be a valuable tool in the cryptocurrency domain, enhancing the ability to anticipate market shifts and understand the psychological drivers behind Bitcoin's price changes. Future research could explore the integration of these models with other data sources and methodologies to further improve prediction accuracy and market understanding.

ACKNOWLEDGEMENT

We are appreciative of our mentor, Professor Abhilasha Raghtate, for her unwavering assistance, invaluable counsel, and astute understanding during this research effort. Their advice has been invaluable in helping us focus our efforts and establish a culture of lifelong learning. We are appreciative of her for giving us the tools and encouraging the academic setting required to do this project effectively. Her advice and encouragement have been crucial in motivating us to explore creative strategies for getting around the models.

REFERENCES

1. **Nakamoto, S.** (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Retrieved from https://bitcoin.org/bitcoin.pdf

2. **Catalini, C., & Gans, J. S.** (2016). *Some Simple Economics of the Blockchain*. *MIT Sloan Research Paper*, 5191-16. https://doi.org/10.2139/ssrn.2874598

3. **Klein, T., Kroyer, S., & Weistroffer, C.** (2018). *Bitcoin, cryptocurrencies, and blockchain: A new way of doing business*. *International Journal of Economics and Financial Issues*, 8(1), 59-72. https://www.econjournals.com/index.php/ijefi/article/view/5564

4. **Zhang, W., Hu, W., & Wu, J.** (2018). *Sentiment analysis of Twitter data for stock market prediction*. *Journal of Financial Data Science*, 1(2), 31-40. https://doi.org/10.3905/jfds.2018.1.2.031

5. **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

6. **Candle, S., & Powers, R.** (2019). *Cryptocurrency price prediction using machine learning techniques*. *International Conference on Machine Learning*, 654-666. https://www.icml.cc/Conferences/2019/Proceedings

7. **Moulin, J.** (2020). *Sentiment analysis for financial markets: A comprehensive review*. *Journal of Financial Market Analysis*, 14(3), 125-141. https://doi.org/10.1016/j.jfma.2020.06.005

8. **Cortes, C., & Vapnik, V.** (1995). *Support-vector networks*. *Machine Learning*, 20(3), 273-297. https://link.springer.com/article/10.1007/BF00994018

9. **LeCun, Y., Bengio, Y., & Hinton, G.** (2015). *Deep learning*. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

10. **Hochreiter, S., & Schmidhuber, J.** (1997). *Long short-term memory*. *Neural Computation*, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

11. **Breiman, L.** (2001). *Random forests*. *Machine Learning*, 45(1), 5-32. https://link.springer.com/article/10.1023/A:1010933404324

12. **Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R.** (2003). *Tfidf accuracy for text classification*. *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, 35-42. https://www.aclweb.org/anthology/W03-0406

13. **Chen, T., & Guestrin, C.** (2016). *XGBoost: A scalable tree boosting system*. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

14. **Shiller, R. J.** (2015). *Irrational Exuberance* (3rd ed.). Princeton University Press.

15 **Sullivan, M.** (2017). *The Role of Sentiment in Financial Markets*. *Journal of Behavioral Finance*, 18(4), 262-275. https://doi.org/10.1080/15427560.2017.1395276

16. **Gloor, P. A., & Zhao, Y.** (2019). *The influence of social media sentiment on stock market performance*. *International Journal of Financial Studies*, 7(2), 23. https://doi.org/10.3390/ijfs7020023

17. **Zheng, L., & Wang, J.** (2021). *Sentiment Analysis and Forecasting in Cryptocurrency Markets*. *Journal of Computational Finance*, 25(2), 56-74. https://doi.org/10.21314/JCF.2021.371