BITCOIN PRICE FLUCTUATIONS AND GOOGLE NEWS WITH MACHINE LEARNING TECHNIQUES

Tam Phan Huy*

University of Economics and Law, Ho Chi Minh City, Vietnam

Vietnam National University, Ho Chi Minh City, Vietnam

*Email: *tamph@uel.edu.vn*

Received: 18 May 2024; Revised: 27 August 2024; Accepted: 4 November 2024

ABSTRACT

This research investigates the predictive power of news sentiment from Google News on Bitcoin price movements, leveraging a five-year dataset of news headlines (2019 to 2024). By correlating sentiment scores with historical Bitcoin prices, the study employs various machine learning algorithms to forecast price trends. The results indicate that while Decision Tree and Random Forest models offer balanced predictions, Logistic Regression and Support Vector Machines achieve high AUC scores but suffer from class imbalance. In contrast, Naïve Bayes and KNN models prove less effective. The findings suggest that sentiment analysis of news headlines can provide moderate short-term predictions for Bitcoin price fluctuations. This study introduces an innovative tool for investors and market analysts, offering insights into the influence of news sentiment on cryptocurrency prices.

Keywords: Bitcoin price, news sentiment, predictive models.

1. INTRODUCTION

Bitcoin, introduced in 2009 by an anonymous figure under the pseudonym Satoshi Nakamoto, has become the most prominent cryptocurrency in the world. Often dubbed "digital gold," Bitcoin represents a revolutionary shift in how value is stored and transferred, operating independently of any central authority, such as a government or financial institution [1]. Instead, it relies on blockchain technology—a decentralized ledger that records all transactions across a network of computers, ensuring transparency and security. Over the years, Bitcoin has gained widespread acceptance, from being a niche technological experiment to becoming a significant financial asset, attracting millions of users, institutional investors, and a large volume of global capital. Its decentralized nature, combined with its limited supply, has made it a popular hedge against inflation and traditional financial systems.

Despite its growing prominence, Bitcoin is notoriously volatile. Its price can swing dramatically within short periods, influenced by a myriad of factors such as technological advancements, regulatory news, macroeconomic trends, and shifts in investor sentiment. These fluctuations can be extreme, with Bitcoin experiencing rapid ascents to new highs followed by sharp declines. This volatility is driven not only by its speculative nature but also by the uncertainty surrounding its future adoption and regulatory landscape. Traditional financial models often struggle to account for this level of unpredictability, making Bitcoin a challenging asset to forecast. For investors, this volatility presents both opportunities for significant gains and risks of substantial losses.

In response to this challenge, sentiment analysis has emerged as a promising approach to predicting cryptocurrency price movements. Unlike traditional financial analysis, which focuses on quantifiable data like historical prices and trading volumes, sentiment analysis examines the mood and opinions expressed in text-based data, such as news articles, social media posts, and online forums. As cryptocurrencies, including Bitcoin, are heavily influenced by public perception and market sentiment, understanding these sentiments has become crucial for forecasting price trends. Recent studies have explored the use of sentiment analysis on various platforms—ranging from Twitter and Reddit to news aggregators like Google News—demonstrating that sentiment can significantly impact cryptocurrency prices. This approach bridges the gap between the qualitative nature.

2. LITERATURE REVIEW

2.1 Background theories

One of the foundational theories underpinning the study of cryptocurrency price movements is the Efficient Market Hypothesis (EMH). EMH, proposed by Fama (1970), suggests that asset prices fully reflect all available information at any given time [2]. In the context of cryptocurrencies like Bitcoin, EMH implies that prices should immediately adjust to new information, such as news articles or social media posts, rendering it impossible to consistently achieve above-average returns through market predictions. However, the extreme volatility and speculative nature of the cryptocurrency market challenge the applicability of EMH. Research has shown that the cryptocurrency market may not always be fully efficient, as price movements often appear to be influenced by irrational behaviors and market sentiment [3]. This has led to an increasing interest in understanding how sentiment analysis can uncover hidden inefficiencies in the market, enabling more accurate predictions.

Another relevant theory is Behavioral Finance, which contrasts with traditional financial theories by acknowledging that psychological factors and cognitive biases significantly influence investor behavior. Behavioral finance suggests that emotions, such as fear and greed, can lead to irrational decision-making, causing asset prices to deviate from their fundamental values [4]. In the context of cryptocurrencies, behavioral finance is particularly relevant because the market is largely driven by speculative activities rather than intrinsic value. Studies have found that sentiment-driven trading, fueled by news and social media, plays a significant role in cryptocurrency price fluctuations [5]. The application of sentiment analysis aligns with the principles of behavioral finance, as it attempts to quantify the collective emotions and opinions of market participants, providing insights into potential price movements that traditional models may overlook.

The Adaptive Market Hypothesis (AMH) offers a more flexible framework that combines elements of both EMH and behavioral finance. AMH posits that market efficiency is not static but evolves over time as market participants adapt to changing environments [6]. In rapidly evolving markets like cryptocurrencies, AMH suggests that prices may reflect a combination of rational and irrational behaviors, depending on the context and the information available. This theory supports the use of sentiment analysis in cryptocurrency price prediction, as it acknowledges that market participants continuously adapt their strategies based on new information, including sentiment signals. By incorporating both rational analysis and behavioral insights, AMH provides a comprehensive framework for understanding the dynamic nature of the cryptocurrency market and the role of sentiment in driving price movements.

2.2 Empirical studies

The application of sentiment analysis in predicting cryptocurrency prices has gained significant attention in recent years. As the cryptocurrency market is highly sensitive to public perception, researchers have leveraged sentiment analysis to capture the mood of the market and its impact on price movements. Various studies have utilized text-based data from multiple sources, such as social media platforms, news articles, and online forums, to gauge the sentiment around cryptocurrencies like Bitcoin and Ethereum. For instance, Lamon et al. (2017) explored the relationship between tweet volume and cryptocurrency prices, concluding that while the sentiment of tweets did not significantly correlate with price movements, the volume of tweets did provide predictive insights [7]. Similarly, Vo, Nguyen, and Ock (2019) employed sentiment analysis on Chinese social media posts to predict cryptocurrency price fluctuations, demonstrating that sentiment derived from these platforms could effectively forecast market trends [8]. These studies underscore the importance of sentiment in understanding the dynamics of the cryptocurrency market.

Moreover, sentiment analysis has been extended to various news aggregators and platforms, recognizing the influence of news headlines on market behavior. Huang et al. (2021) analyzed news sentiment to predict Ethereum price movements, showing that sentiment from news articles could be used to determine whether to buy, sell, or hold the cryptocurrency [9]. This research highlights that news sentiment is not only a reflection of public opinion but also a potential driver of market actions. Similarly, Gerritsen, Lugtigheid, and Walther (2022) investigated the impact of expert predictions published in news

articles on Bitcoin prices, finding that bearish and neutral predictions were followed by negative returns [10]. These studies collectively demonstrate that sentiment analysis, when applied to news data, can be a powerful tool in predicting cryptocurrency price movements, particularly in the short term.

In parallel with sentiment analysis, machine learning techniques have become increasingly popular for forecasting cryptocurrency prices. Machine learning models, with their ability to handle large datasets and identify complex patterns, are well-suited to the task of predicting volatile markets. Among the most commonly used models in this field are supervised learning algorithms, which have shown promising results in predicting price trends based on historical data and sentiment inputs. For example, Mittal, Dhiman, Singh, and Prakash (2019) applied linear regression, polynomial regression, and long short-term memory (LSTM) models to analyze the correlation between Bitcoin prices and Google search patterns, concluding that these models could effectively capture the relationship between search interest and price fluctuations. Their findings suggest that machine learning algorithms can enhance the predictive power of sentiment analysis by identifying underlying patterns in the data [11].

Support Vector Machines (SVM) and Decision Trees have also been widely used in cryptocurrency prediction, particularly in combination with sentiment analysis. In a study by Prajapati (2020), SVM and Decision Tree models were applied to Reddit posts to predict Bitcoin price movements, with the models showing varying degrees of accuracy depending on the sentiment extracted from the posts. This research highlights the versatility of machine learning algorithms in handling different types of data and their adaptability to various sources of sentiment information [12]. Additionally, Random Forest models, which are an extension of Decision Trees, have been used to enhance prediction accuracy by aggregating multiple decision trees and reducing overfitting. These models have proven effective in capturing the nonlinear relationships inherent in cryptocurrency markets.

Furthermore, more advanced machine learning models, such as deep learning techniques, have been explored for cryptocurrency price prediction. Huang et al. (2021) utilized an LSTM-based model to analyze sentiment and predict Ethereum price movements, demonstrating that deep learning models could outperform traditional machine learning algorithms in capturing the temporal dependencies in time-series data. The success of LSTM models in this field is attributed to their ability to retain information over time, making them particularly well-suited for analyzing sequential data like cryptocurrency prices. The growing use of deep learning models reflects the ongoing evolution of machine learning techniques in the field of financial forecasting [9].

Overall, the combination of sentiment analysis and machine learning has opened new avenues for predicting cryptocurrency price movements. While traditional models often struggle with the volatility and complexity of these markets, machine learning algorithms provide a more robust framework for handling large datasets and uncovering hidden patterns. As research in this field continues to evolve, the integration of sentiment analysis with advanced machine learning techniques holds significant promise for improving the accuracy and reliability of cryptocurrency price predictions.

This study contributes to the literature by focusing on the application of sentiment analysis using news headlines from Google News, a source that has been underexplored in previous studies compared to more commonly used platforms like Twitter or Reddit. While prior research has examined the impact of social media sentiment on cryptocurrency prices, the influence of news sentiment, particularly from a trusted aggregator like Google News, remains less understood. Moreover, many existing studies either focus on a limited timeframe or rely on fewer machine learning models. This study fills these gaps by utilizing a five-year dataset and applying a comprehensive suite of machine learning algorithms to provide a more robust analysis of Bitcoin price fluctuations. This approach not only enhances predictive accuracy but also offers valuable insights into the specific role that news sentiment plays in the cryptocurrency market, contributing to both academic understanding and practical investment strategies.

3. METHODOLOGY

The dataset used in this study is composed of two primary components: historical Bitcoin prices and textual data extracted from news articles related to Bitcoin. The textual data was sourced from Google News over a five-year period, from 2019 to 2024, resulting in a total of 33,847 observations. The historical price data was collected on a daily basis, allowing for a comprehensive analysis of Bitcoin's price movements over time. The reason for choosing this timeframe is because of the data availability on Google News search query. The combination of these two data sources enables the study to explore the relationship

between market sentiment, as reflected in news headlines, and Bitcoin's price fluctuations.

The data processing involved several key steps. First, the data was collected and then cleaned to address any inconsistencies or missing values. Following this, sentiment analysis was performed on the textual data to quantify the sentiment expressed in the news articles, categorizing it into positive, negative, or neutral sentiments. Finally, the processed sentiment data, along with the historical price data, was used to train and test various machine learning algorithms. This multi-step process ensures that the dataset is prepared for accurate and meaningful analysis, allowing the study to investigate the predictive power of news sentiment on Bitcoin prices.

In the machine learning phase of this research, several algorithms were employed to predict Bitcoin price movements based on the processed dataset. The models used include Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machine (SVM), Decision Tree, and Random Forest. The data was split into training and testing sets, with 70% of the data allocated for training and the remaining 30% used for testing. This approach allows the models to learn from a substantial portion of the data while still providing a separate set of data to evaluate their predictive accuracy.

The method of splitting the dataset into training and testing sets, with 70% allocated for training and 30% for testing, is a commonly used approach in machine learning to ensure robust model performance and to prevent overfitting. This approach allows the model to learn from a large portion of the data while preserving a separate dataset to evaluate its generalization ability on unseen data, ensuring that the model performs well not just on the training data but also on new, unobserved instances [13]. This method is widely accepted as it strikes a balance between providing enough data for training and maintaining sufficient data for validation, leading to more reliable and unbiased performance metrics [14].

For model evaluation, a range of performance metrics was utilized to assess the effectiveness of each algorithm. These metrics include accuracy, which measures the overall correctness of the model's predictions, and precision, which indicates the proportion of true positive predictions out of all positive predictions. Recall, or sensitivity, assesses the model's ability to correctly identify true positives, while the F1 score provides a balance between precision and recall. Additionally, the Area Under the Curve (AUC) metric was used to evaluate the model's ability to distinguish between classes, providing a comprehensive assessment of each algorithm's performance in predicting Bitcoin price movements.

4. RESULTS AND DISCUSSION

Table 1 provides an overview of the sentiment distribution in the dataset used for this study. The table categorizes the sentiment of the textual data into three distinct groups: Neutral, Positive, and Negative. Specifically, out of the total 33,847 observations, the majority of the news articles are classified as Neutral, with 19,733 observations falling into this category. This is followed by 9,466 observations classified as Positive and 4,648 observations classified as Negative:

Table 1. Sentiment analysis data count

	Neutral	Positive	Negative	Total
No. Observation	19,733	9,466	4,648	33,847

The distribution of sentiment in Table 1 highlights some key characteristics of the news coverage related to Bitcoin. The dominance of Neutral sentiment, with nearly 58% of the total observations, suggests that much of the news coverage around Bitcoin tends to be factual or balanced, focusing on objective reporting rather than overtly positive or negative perspectives. This neutrality in sentiment could be indicative of the nature of financial reporting, which often strives to present information without bias.

The Positive sentiment category, accounting for approximately 28% of the observations, reflects the optimistic outlook that is often associated with Bitcoin, especially during periods of price increases or technological advancements in the cryptocurrency space. This positive sentiment could influence investor behavior by reinforcing bullish market expectations. On the other hand, the Negative sentiment category, making up about 14% of the total observations, represents the concerns and criticisms surrounding Bitcoin. These negative sentiments could be driven by factors such as regulatory crackdowns, market downturns, or security breaches, all of which can have a significant impact on

market sentiment and, consequently, on Bitcoin's price movements.

The balance between these sentiment categories provides valuable insights into the sentiment dynamics within the cryptocurrency market. The relatively high proportion of Neutral sentiment suggests that Bitcoin-related news may not always provide clear directional signals for market participants. However, the presence of both Positive and Negative sentiment categories emphasizes the importance of sentiment analysis in capturing market sentiment shifts, which could be crucial for predicting short-term price movements.

Figure 1 is a scatter plot that illustrates the relationship between polarity and subjectivity of news articles related to Bitcoin. Polarity represents the sentiment of the news articles, with values ranging from negative (left side) to positive (right side). Subjectivity, on the vertical axis, measures the degree to which the content of the articles is opinion-based rather than fact-based, with higher values indicating more subjective content.

In the scatter plot, two distinct colors are used to represent different sentiments. The blue dots on the right side of the plot represent positive polarity values, indicating that these articles have a positive sentiment. The orange squares on the left side represent negative polarity values, indicating negative sentiment. Neutral articles with a polarity close to zero are positioned around the center of the plot. The scatter of points suggests a distribution of news articles along the polarity axis, with varying levels of subjectivity.

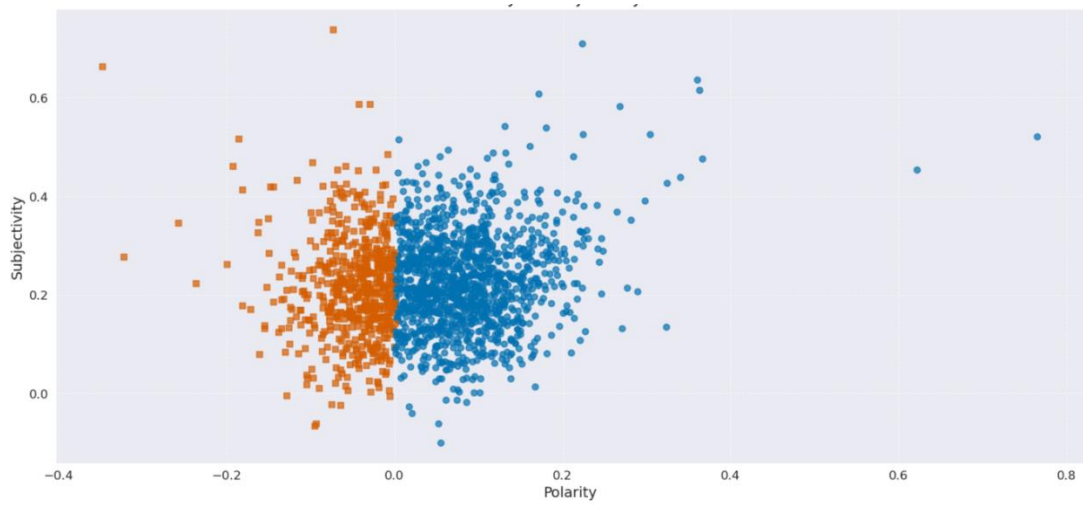

Figure 1. Polarity vs subjectivity

The distribution of points in Figure 1 highlights several important observations about the nature of the news articles in the dataset. Firstly, there is a concentration of articles near the neutral polarity value, indicating that many articles do not express strong positive or negative sentiment. This aligns with the sentiment distribution observed in the dataset, where neutral articles are the majority. The spread of subjectivity values across the plot shows that both positive and negative sentiment articles can vary significantly in their subjectivity. Some articles, despite being highly opinionated (with higher subjectivity), still express a clear positive or negative sentiment. On the other hand, articles with lower subjectivity values tend to be more fact-based and are less likely to exhibit extreme polarity.

Interestingly, there are a few outliers in the plot, particularly on the positive polarity side, where some articles exhibit very high polarity but relatively low subjectivity. This could suggest that certain articles present strong positive sentiment in a factual manner, potentially influenced by significant positive developments in the Bitcoin market. The plot also illustrates that negative sentiment articles (in orange) tend to cluster more closely around the neutral polarity line compared to positive sentiment articles, which are more spread out. This might indicate that negative news about Bitcoin is often less extreme in its sentiment, possibly reflecting cautious or balanced reporting.

Table 2 provides an overview of the performance metrics—accuracy, precision, recall, and F1 score—across six different machine learning algorithms used to predict Bitcoin price movements. The

standout performers in this analysis are Logistic Regression and Random Forest. Logistic Regression achieved the highest precision at 0.82, indicating that it is particularly effective at correctly identifying positive instances. However, its F1 score of 0.70 suggests that while it excels in precision, it may miss some positive cases, leading to a lower recall. Random Forest, on the other hand, shows a well-rounded performance with an accuracy of 0.71 and balanced precision, recall, and F1 score all at 0.71. This suggests that Random Forest is effective at handling the complexities of the dataset without sacrificing the balance between precision and recall.

Table 2. Classification report for machine learning algorithms

Classifier	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.73	0.82	0.73	0.7
KNN	0.5	0.5	0.5	0.5
Naive Bayes	0.54	0.53	0.54	0.51
SVM	0.66	0.78	0.66	0.61
Decision Tree	0.68	0.68	0.68	0.68
Random Forest	0.71	0.71	0.71	0.71

Support Vector Machine (SVM) also shows relatively strong performance, particularly in precision with a score of 0.78, though its recall and F1 score are slightly lower, indicating that it might struggle to consistently identify true positives. Meanwhile, Decision Tree offers balanced, moderate performance across all metrics, with an accuracy of 0.68 and equivalent precision, recall, and F1 scores. Although it doesn't outperform Random Forest or Logistic Regression, Decision Tree provides a stable alternative, especially when model interpretability is important.

In contrast, K-Nearest Neighbors (KNN) and Naive Bayes perform poorly in this analysis. KNN, with all metrics at 0.5, indicates that it performs no better than random guessing. Naive Bayes, while slightly better than KNN, still struggles with an accuracy of 0.54 and similarly low precision, recall, and F1 scores. These results suggest that KNN and Naive Bayes may not be well-suited for predicting Bitcoin price movements in this particular context, as they fail to effectively learn from the data. In conclusion, Random Forest and Logistic Regression stand out as the most effective models for this prediction task, with Random Forest offering a balanced approach and Logistic Regression excelling in precision. On the other hand, KNN and Naive Bayes are less effective, making them unsuitable for this specific application.

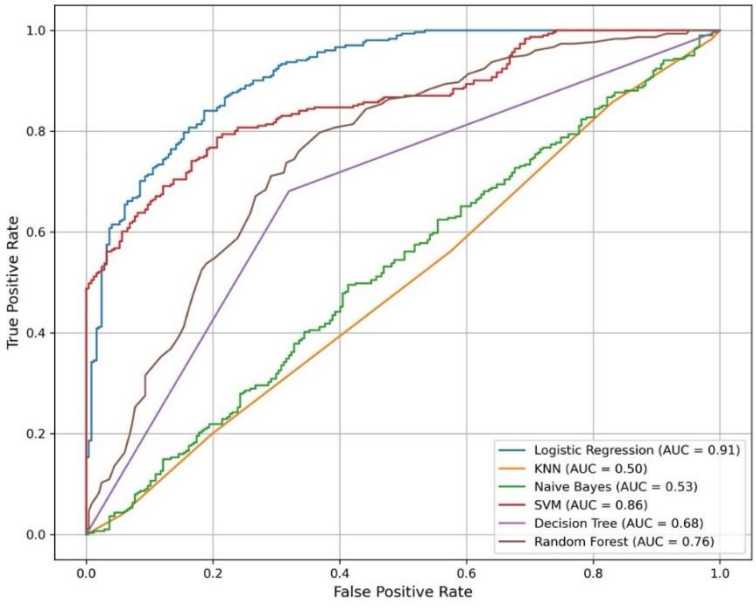

Figure 2. ROC for different machine learning algorithms

Figure 2 illustrates the Receiver Operating Characteristic (ROC) curves for six machine learning

algorithms used to predict Bitcoin price movements. The ROC curve plots the True Positive Rate (sensitivity) against the False Positive Rate, providing insight into each model's performance. The Area Under the Curve (AUC) values are also displayed, summarizing the overall ability of each model to distinguish between classes.

Logistic Regression stands out as the best-performing model, with the highest AUC of 0.91. Its ROC curve rises sharply, indicating that the model can effectively differentiate between true positives and false positives. This strong performance makes Logistic Regression a top choice for this prediction task. Following closely is the Support Vector Machine (SVM) with an AUC of 0.86, suggesting it is also highly effective, though slightly less so than Logistic Regression. Random Forest performs moderately well, with an AUC of 0.76. While it is not as powerful as Logistic Regression or SVM, it still demonstrates a reasonable ability to balance sensitivity and specificity. On the other hand, the Decision Tree model shows a noticeable drop in performance, with an AUC of 0.68. This suggests that while it is functional, it may struggle with the complexity of the data compared to more advanced algorithms.

In contrast, Naive Bayes and K-Nearest Neighbors (KNN) show poor performance, with AUCs of 0.53 and 0.50, respectively. Their ROC curves are close to the diagonal line, indicating that their predictions are only marginally better than random guessing. These models appear to be ineffective for this particular task. In summary, Figure 2 clearly highlights the superiority of Logistic Regression and SVM for predicting Bitcoin price movements, with Random Forest being a viable, albeit less effective, alternative. Conversely, KNN and Naive Bayes demonstrate inadequate performance, making them unsuitable for this prediction challenge.

The findings of this study have significant implications for both cryptocurrency market participants and institutional investors. The demonstrated effectiveness of Logistic Regression and Support Vector Machine models in predicting Bitcoin price movements based on news sentiment suggests that incorporating sentiment analysis into market strategies can provide a competitive edge in highly volatile markets like cryptocurrency. In particular, the high accuracy and AUC scores of these models highlight their practical applicability for real-time trading decisions. Investors can leverage these tools to better understand market sentiment and anticipate price fluctuations, allowing them to time their trades more effectively. Furthermore, financial institutions could integrate these machine learning algorithms into their trading platforms to offer advanced analytics and predictive capabilities, thereby enhancing the services provided to their clients. These results underscore the critical role of sentiment-driven factors in cryptocurrency pricing, offering valuable insights for developing more sophisticated risk management and investment strategies.

5. CONCLUSION & RECOMMENDATION

Based on the results, it is evident that Logistic Regression and Support Vector Machine (SVM) outperform other machine learning algorithms in predicting Bitcoin price movements. Logistic Regression, with its high AUC of 0.91, demonstrates a strong ability to distinguish between classes, making it a reliable choice for predicting price trends based on news sentiment. SVM, with an AUC of 0.86, also proves to be a robust model, though slightly less effective than Logistic Regression. These findings align with previous studies that have identified Logistic Regression and SVM as effective models for financial predictions, particularly in sentiment-driven markets [8,11]. On the other hand, models like K-Nearest Neighbors (KNN) and Naive Bayes perform poorly, with AUCs close to random guessing, indicating that they are not suitable for this type of prediction task. This outcome is consistent with other research that has found these simpler models to be less effective in complex financial environments [7].

The comparison of these models highlights the importance of selecting the right machine learning algorithm for sentiment analysis in cryptocurrency prediction. While Decision Tree and Random Forest models offer moderate performance, their results suggest they may struggle with the intricate patterns in the data, especially compared to more sophisticated models like Logistic Regression and SVM. These findings contribute to the growing body of literature that emphasizes the need for advanced machine learning techniques in accurately forecasting cryptocurrency prices, particularly in a volatile and sentiment-driven market like Bitcoin [10, 12].

For investors, the results of this study suggest that incorporating sentiment analysis, particularly using Logistic Regression or SVM models, can enhance prediction accuracy when making investment decisions in the cryptocurrency market. By leveraging these advanced models, investors can gain a more reliable understanding of potential price movements, allowing them to better time their trades and manage risks. Investors should consider integrating sentiment-based machine learning models into their trading strategies to improve their ability to anticipate market trends driven by news sentiment.

For managers of financial institutions and cryptocurrency funds, the findings highlight the value of adopting sophisticated machine learning algorithms for market analysis. As the cryptocurrency market continues to evolve, utilizing models like Logistic Regression and SVM can provide a competitive edge in developing more accurate predictive tools. Managers should invest in technology and data analytics teams to build and implement these models, enabling them to offer better financial products and services to their clients while also enhancing their own market performance.

For government agencies and regulators, the study underscores the importance of monitoring sentiment in the cryptocurrency market. As public sentiment can significantly impact price movements, regulatory bodies should consider using sentiment analysis as part of their market oversight strategies. This could help in detecting potential market manipulations and understanding the broader impact of regulatory announcements on cryptocurrency prices. Additionally, government agencies could collaborate with academic institutions to further explore the integration of sentiment analysis in policy-making processes, particularly as it relates to market stability and investor protection.

Despite the promising results, this study has several limitations that warrant further exploration. First, the dataset is limited to news sentiment from a specific time frame and source (Google News), which may not fully capture the breadth of sentiment across different media platforms. Future research could expand the dataset to include social media sentiment, such as Twitter or Reddit, to gain a more comprehensive view of market sentiment. Additionally, this study only applied a limited set of machine learning algorithms; further research could explore the use of more advanced models, such as deep learning techniques like Long Short-Term Memory (LSTM) networks, to improve prediction accuracy. Finally, the study focuses on short-term price predictions, and future studies could examine the effectiveness of sentiment analysis in predicting long-term trends in the cryptocurrency market.

Funding

The research is funded by the University of Economics and Law, Vietnam National University, Ho Chi Minh City, Vietnam.

REFERENCES

1. Lanchester, J. - When Bitcoin grows up. London Review of Books **38** (8) (2016) 3-12.
2. Fama, E. F. - Efficient capital markets: A Review of Theory and Empirical Work. Journal of Finance **25** (2) (1970) 383-417. https://doi.org/10.2307/2325486
3. Urquhart, A. - The inefficiency of Bitcoin. Economics Letters **148** (2016) 80-82. https://doi.org/10.1016/j.econlet.2016.09.019
4. Shiller, R. J. - From efficient markets theory to behavioral finance. Journal of Economic Perspectives **17** (1) (2003) 83-104. https://doi.org/10.1257/089533003321164967
5. Chen, C. Y.-H., Guo, L., Renault, T. - What makes cryptocurrencies special? Investor sentiment and return predictability. Working Paper (2019). http://dx.doi.org/10.2139/ssrn.3398423
6. Lo, A. W. - The adaptive markets hypothesis: market efficiency from an evolutionary perspective. Journal of Portfolio Management, Forthcoming. Available at SSRN: https://ssrn.com/abstract=602222
7. Lamon, C., Nielsen, E., Redondo, E. - Cryptocurrency price prediction using news and social media sentiment. SMU Data Science Review **1** (3) (2017) 1-22.
8. Vo, A. D., Nguyen, Q. P., Ock, C. Y. - Semantic and syntactic analysis in learning representation based on a sentiment analysis model. Applied Intelligence **50** (2020) 663-680. https://doi.org/10.1007/s10489-019-01540-2

9. Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., Zhang, J. - LSTM based sentiment analysis for cryptocurrency prediction. In: International Conference on Database Systems for Advanced Applications (2021) 617-621. Springer, Cham. https://doi.org/10.1007/978-3-030-73200-4_47
10. Gerritsen, D. F., Lugtigheid, R. A., Walther, T. - Can bitcoin investors profit from predictions by crypto experts? Finance Research Letters **46** (2022) 102266. https://doi.org/10.1016/j.frl.2021.102266
11. Mittal, A., Dhiman, V., Singh, A., Prakash, C. - Short-term bitcoin price fluctuation prediction using social media and web search data. In: Twelfth International Conference on Contemporary Computing (IC3) (2019) 1-6. IEEE. https://doi.org/10.1109/IC3.2019.8844899
12. Prajapati, P. - Predictive analysis of Bitcoin price considering social sentiments. arXiv preprint arXiv:2001.10343 (2020). https://doi.org/10.48550/arXiv.2001.10343
13. Kohavi, R. - A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International Joint Conference on Artificial Intelligence (1995) 1137-1145.
14. Goodfellow I., Bengio Y., Courville A. - Deep learning, MIT press, 2016.

TÓM TẮT

PHÂN TÍCH BIẾN ĐỘNG GIÁ BITCOIN VÀ GOOGLE NEWS BẰNG PHƯƠNG PHÁP MÁY HỌC

Phan Huy Tâm*

Trường Đại học Kinh tế - Luật và Đại học Quốc gia Tp. Hồ Chí Minh, Việt Nam

*Email: *tamph@uel.edu.vn*

Nghiên cứu này điều tra khả năng dự đoán của cảm xúc tin tức từ Google News đối với biến động giá Bitcoin, dựa trên tập dữ liệu tiêu đề tin tức trong 5 năm (2019 đến 2024). Bằng cách tương quan điểm số cảm xúc với giá Bitcoin lịch sử, nghiên cứu áp dụng nhiều thuật toán máy học khác nhau để dự báo xu hướng giá. Kết quả cho thấy mặc dù các mô hình Decision Tree và Random Forest đưa ra dự đoán cân bằng, nhưng Logistic Regression và Support Vector Machines đạt điểm AUC cao nhưng gặp vấn đề về mất cân bằng lớp. Ngược lại, các mô hình Naïve Bayes và KNN kém hiệu quả hơn. Những phát hiện này gợi ý rằng phân tích cảm xúc từ tiêu đề tin tức có thể cung cấp dự đoán ngắn hạn mức độ vừa phải về biến động giá Bitcoin. Nghiên cứu này giới thiệu một công cụ sáng tạo cho các nhà đầu tư và nhà phân tích thị trường, cung cấp những hiểu biết sâu sắc về ảnh hưởng của cảm xúc tin tức đối với giá tiền điện tử.

Từ khóa: Giá Bitcoin, Phân tích cảm xúc, Mô hình dự báo.