ISSN: 2633-4828 Vol. 5, No.1, January, 2023

International Journal of Applied Engineering & Technology

Topic Modeling on Crowd Trading Ideas for Digital Asset Price Prediction

Seksak Prabpala[1], Wirapong Chansanam[2], Kulthida Tuamsuk[3*]
[1,2,3]Department of Information Science, Khon Kaen University, Khon Kaen, Thailand.

*Corresponding author: Kulthida Tuamsuk, Email: kultua@kku.ac.th

Date of Submission: *12 November 2022* ***Revised:*** *24 December 2022* ***Accepted:*** *05 January 2023*

How to Cite: **Prabpala, S., Chansanam, W., and Tuamsuk, K. (2023). Topic Modeling on Crowd Trading Ideas for Digital Asset Price Prediction. International Journal of Applied Engineering & Technology 5(1),pp 6-12.**

Abstract **- Tradingview is the most widely recognized social network platform for stock and digital asset trading, which a number of investors access to share their thoughts on investments each day. The present study sought to analyze contents published in the "Ideas" forum on Tradingview from November 2021 – October 2022 using topic modeling through Latent Dirichlet Allocation (LDA). The results demonstrated that 9,553 texts drawn from Tradingview's Ideas were classified into six topics about digital asset trading. Based on the hypothesis and the prediction, it was found that the most shared idea, particularly 3,354 texts, concerning trading strategy and risk/profit, followed by 3,101 about trend line, high/low price pattern, and support/resistance. The number of texts on both topics accounted for 67.56%. It can be implied that investments in digital assets require good strategies, risk management, analysis of price trends, and knowledge of support and resistance. Identifying significant words and topics using LDA uncovers a larger number of words latent in the texts than hypothesiz.**

Index Terms - digital assets, crowd idea user, topic modeling; Latent Dirichlet Allocation, digital trading, trading ideas

INTRODUCTION

In general, digital assets refer to anything which is created and stored digitally, is identifiable and accessible, and has values or can be priced. Information, images, videos, and written content digitally recorded are regarded as digital assets with ownership and possession rights. Certain types of digital assets, such as corporate brands, may possess monetary or intangible values. On the other hand, the others, such as photos on a mobile phone, are created by a person. There had been no channels to generate incomes from such assets legally and safely. Later, with the technological emergence of blockchain and cryptocurrency in 2009, a novel idea about digital assets emerged. Specifically, anything in the digital format can be valued for its monetary value through tokenization on blockchain [1]. Simply put, digital assets are electronic data which hold values just as general assets and are intangible in nature. Their ownership can be traded without middle-persons or intermediaries on an electronic network commonly referred to as a public blockchain, allowing any individual to access information and conduct transactions. Each transaction on a blockchain system is recorded, thus being immutable and indelible. As a result, this system is deemed highly secure. Hence, those wishing to make investments or purchase goods or services can carry out transactions through a blockchain system with cryptocurrency or digital tokens to establish individual rights [2, 3].

Digital assets are indeed regarded as a product for investment which has been growing in popularity among the public. According to Coinmarketcap, the market capitalization is approximately USD 1.7 trillion. Currently, there are over 18,000 digital coins in the world, more than 480 markets to attract investors to trade, and upwards of 300 million investors across the globe using or trading cryptocurrency [4]. In Thailand, there are roughly 2.7 million active accounts on the stock exchange [3]. In addition, Statista Digital Economy Compass estimates that around 4 million Thai people possess digital assets while upwards of 5.65 million hold NFTs (non-fungible tokens) [5]. As reported by the world's top-ranking blockchain data analytics company Chainanalysis, Thailand ranks 12th among the countries where cryptocurrency is well-known and accepted [6]. Additionally, in Digital 2022: Global Overview Report, it is mentioned that Thai internet users holding cryptocurrency account for 20.1%, ranking the first for the highest number of cryptocurrency holders in the world, as depicted in Figure 1 [7].

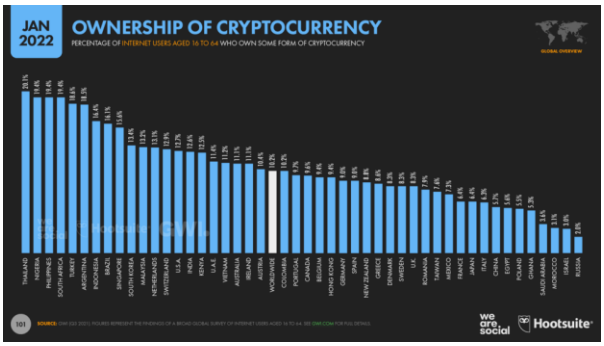

FIGURE 1
OWNERSHIP OF CRYPTOCURRENCY REPORT

Retail investors are required to learn about and stay abreast of financial news in order to make effective investments. It is undeniable that social media serve as one of the crucial channels which people use to exchange ideas and enhance knowledge about trading, contributing to the emergence of many investor communities on social media. Among many platforms, Tradingview is recognized as the social trading platform with tremendous popularity, with 10 million users per month from 150 countries worldwide. This platform provides tables of asset values, such as foreign currency exchange rates and stock price databases from stock exchanges worldwide, namely the United States of America and leading markets in Europe and Asia. Its distinguishing features lie in the fact that a graph can be displayed on a browser instantly without any application, iOS and Android-supported applications are available to allow for ease of use among traders, a variety of technical tools and indicators are provided, and users are allowed to write scripts and simulate trading. Most importantly, Tradingview is a social network where investors, as users worldwide, can post their graphs to share their trading ideas on which other users are allowed to share their opinions, and the users can follow whom they admire. The "Ideas" forum on the platform enables investors to capture the market direction, collect ideas in Idea Stream, and draw on those ideas to form an idea for intraday trading or review their trading processes and share them publicly with other traders or their followers [8]. Thus, contents in the "Ideas" forum on Tradingview contain useful information in relation to technical analysis, investment factors, investment mindsets, and price predictions as in an uptrend and a downtrend.

Topic modeling is a technique for constructing a data distribution model through natural language processing to categorize data. It can identify a pattern of words and phrases in the contents of documents and classify similar words to determine the characteristics of the data set. Topic modeling is grounded on the notion that the contents of the documents are created through a combination of various topics, each of which has a probability distribution of many words occurring in that particular topic [9]. This technique is commonly implemented through Latent Dirichlet Allocation (LDA) in order to extract and categorize words in each latent topic in respective order; taking into consideration the relationship and consistency of words in the latent topic will help specify topics and provide a description of what is contained in the contents of the documents [10, 11]. Prior research shows that LDA has been applied to text classification to develop a research paper classification system [12], to establish the facts of news through comparison of facts and news during a certain period [13], and to summarize many copies of online news documents to enable readers to recognize the intent of such documents [14]. It can be clearly seen that LDA has been extensively applied to sources of information in different domains and to text analysis.

The present study was intended to analyze the contents in the "Ideas" forum on Tradingview, a renowned investor community on a social network where sophisticated investors worldwide share their ideas about technical analysis, investment mindsets, investment factors, and directions of investments in digital assets. In particular, topic modeling was carried out using LDA. This study was conducted in the hope that the results would uncover topics and provide descriptions of those topics of the contents contained in the so-called social media about investments in digital assets. It could be drawn on to develop an ontology and forecast the direction of investments in connection with trading rates or market conditions of digital assets.

RELATED WORKS

Topic Modeling (TM) is a technique for identifying groups of words in a corpus of documents or texts constructed for text mining for complicated tasks, particularly identifying topics in a corpus of documents with a reservoir of data [15, 16]. It can provide reliable and accurate results through different methods of analysis, for example, Probabilistic Latent Analysis (PLSA), Latent Semantic Analysis (LSA), and Latent Dirichlet Allocation (LDA) [17]. In particular, LDA developed by [11] is regarded as a TM algorithm most commonly used to extract a set of topics from a large corpus of documents since it can produce accurate results. Based on previous research, it can be seen that the use of TM for investigating topics in documents or texts on social media has been growing in popularity since content on such media can reflect concepts and views of posting and sharing content in diverse dimensions, such as social aspects, behavior, politics, and lifestyles, depending on types and characteristics of sharing communities. As an example, Community Question Answering (CQA), which is becoming an important source of information, is a platform in which a considerable number of users access to post questions and answers. Tian Kochhar proposed an effective method of predicting answers for new questions on CQA by constructing a topic model based on previous questions with the answers. Specifically, learning is based on answers using a collective voting system in CQA. The study examined a set of data obtained from Stack Overflow, one of the largest CQA. It was discovered that the effective TM comprised

four parts, namely data preprocessing, user profile building, user topical interest and expertise learning, and ranking model building. It started with building user profiles on previously asked questions, including texts and voting information, used for learning about the topical interest and expertise of users in each topic. The study also proposed a ranking model to compute the probability that users were the best answerers [18].

Apart from that, there are other interesting studies on TM using LDA. One of them was the development of research paper classification systems; keywords from the abstract of each paper were captured, and the K-means algorithm was employed to classify all of the papers with similar topics on the basis of the Term frequency-inverse document frequency (TF-IDF) value of each paper [12]. The other studies were the construction of topic models for social text analysis of short texts on social media [19], analysis of research topics and research directions in information science through research articles in the national database [20], and topic classification from online news documents to allow readers to understand the intent of the documents more clearly [14].

In relation to research on digital assets and investments using TM, one study analyzed a cryptocurrency forum to understand the contents of conversations in an online community. Using LDA, it identified bags of words obtained from a popular cryptocurrency forum, user profiles, and times; additionally, the use of machine learning enabled the study to analyze comments and connections of topics as well as to compare them against incidents about cryptocurrency [21]. The other study entailed constructing a topic model by classifying "topics" from a corpus of short texts, such as posts on a social network. The study employed a set of tweet data about Bitcoin, trained the model on three topics, and evaluated the results with different scoring methods. It was found that Dirichlet Multinomial Mixture (DMM) was the most effective method, so it could be utilized to conduct an analysis of Bitcoin effectively [22].

Furthermore, with regard to research on stock market predictions, one study proposed a method for analyzing comments on stock market predictions based on information on social media using Topic Sentiment Latent Dirichlet Allocation (TSLDA). The results demonstrated that its accuracy was approximately 6.07%. Compared to other methods, TSLDA was 6.43% and 6.07% more accurate than LDA and JST. It was shown that the use of data from posts on social media could help increase the accuracy of stock market predictions [13]. The other study employed LDA to analyze the trends of the stock market based on news on Twitter to create news automatically or detect events and make predictions. However, there were certain issues arising from users' posts and sharing of inaccurate or irrelevant content. Thus, appropriate user selection is essential for randomly selecting tweets as the sample for analysis [23].

METHODOLOGY

The analysis of significant words and topic classification in the Ideas forum about investments in digital assets through topic modeling consisted of three steps as follows: 1) data collection, 2) data cleansing and preparation, and 3) topic modeling through the LDA model (Figure 2).

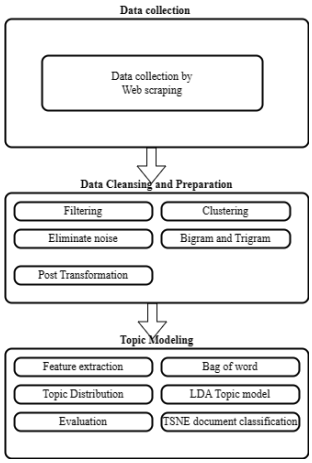

FIGURE 2
RESEARCH PROCESS ON TOPIC MODELING

As displayed in Figure 2, the initial process involves scraping data from the website Tradingview, especially the "Ideas" forum about experienced users' trading. The scraped data are saved in the Excel or Comma Separated Values (CSV) file format. Afterward, the file is used for data cleansing and preparation with three steps: 1) data cleaning, 2) word tokenization, and 3) stop words. Following that process, data are processed in the LDA model. The results are clustering of topics and performance evaluation using silhouette scores.

Regarding data collection, a set of data, namely 9,553 texts, about trading ideas and technical analysis from sophisticated traders during the period November 1st, 2021 – October 31st, 2022, was imported. The data were collected using Selenium and BeautifulSoup for web scraping data from the website Tradingview, particularly the Ideas forum (Figure 3), and were saved in the xlsx file format (Figure 4).

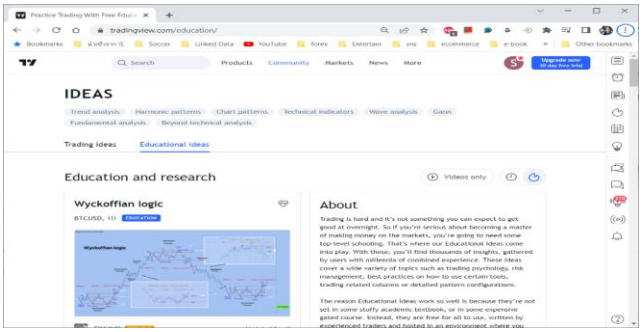

FIGURE 3
SCREEN SHOT OF TRADINGVIEW'S IDEAS

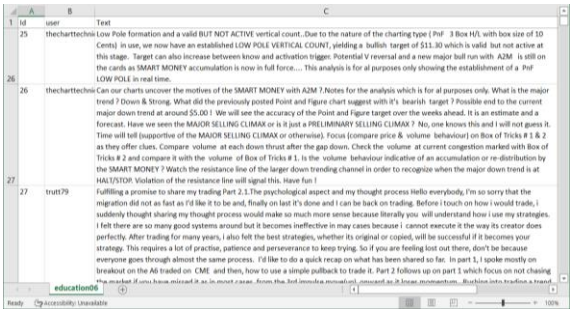

FIGURE 4
EXTRACTED DATA INTO MICROSOFT EXCEL FORMAT

Data preparation was performed using Python for text cleaning and word tokenization. Word tokenization was executed by running the command newmm, and later stop words, i.e., the insignificant or irrelevant words, were filtered out using the command PythaiNLP. The process started with importing both key modules for developing a model (Figure 5) and the file (Figure 6). The data were subsequently prepared for topic modeling (Figure 7).

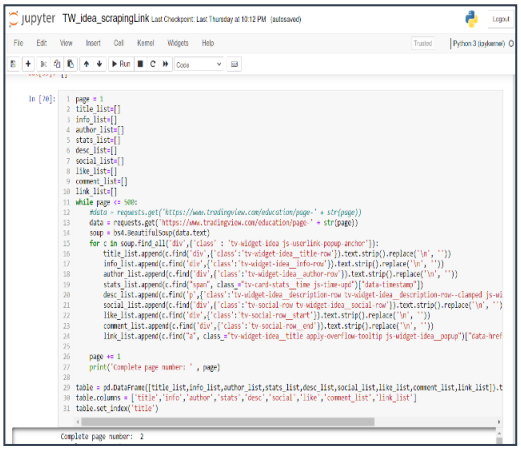

FIGURE 5
EXAMPLE OF CODES FOR DATA EXTRACTION FROM TRADINGVIEW'S IDEAS

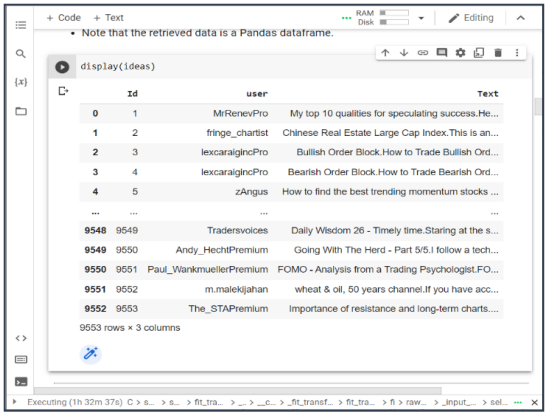

FIGURE 6
EXAMPLE OF DATA IMPORTING FILE

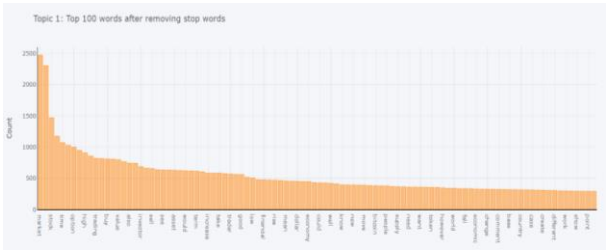

FIGURE 7
EXAMPLE OF DATA PREPARATION AND RESULTS FOR TOPIC MODELING OF TRADINGVIEW'S IDEAS

Vectorization of the texts was carried out using the Sklearn software named Tfidfvectorizer for feature extraction to calculate the frequency of unique words in each text which could help identify significant words. The results from feature extraction of 9,553 texts were drawn on to identify unique words and significant words. Particularly, the variables were defined with max_df = 0.5 to filter out words occurring in more than 50% of the documents and min_df = 10 to remove those available in fewer than 10 documents. Both the significant words in each text and unique words were counted. It was found that there were 37,572 words in total and 693 unique words.

Clustering was performed using the LDA technique. Specifically, an m-value (number of topics) was assigned to each N (documents or texts). The length vector of the m-value could be found with m substituted by the probability distribution of topics in each text. Additionally, the addition of vectors to all texts produced an N-by-m feature matrix, and topics in each document were labeled (Figure 8).

The performance of text clustering was evaluated by measuring intra-cluster similarity, i.e., members in the same cluster and inter-cluster similarity. Based on the following equation, the calculation was performed using the silhouette coefficient through the software Sklearn named silhouette_score.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

a is the average distance between each point in the same cluster.

b is the average distance between each cluster.

i is the data point.

Silhouette score S(i) is in the range of [-1, 1]. If the score is close to 1, it is good.

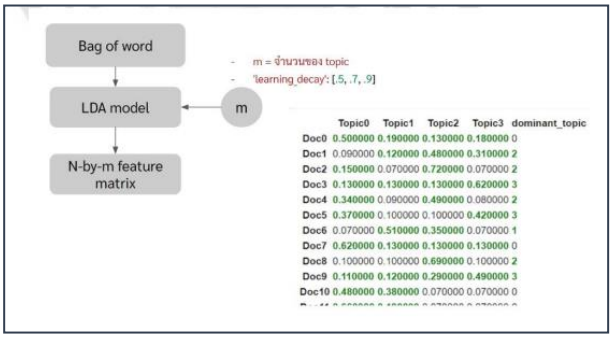

FIGURE 8
TOPIC MODELING USING THE LDA MODEL

RESULTS

The present study sought to analyze significant words and classify topics in the Ideas forum using the LDA technique developed by Python and Google Colab. 9,553 texts were compiled. Clustering was carried out using the K-means algorithm, while its performance was measured through the silhouette coefficient. The results are reported below.

In relation to clustering with the K-mean algorithm and performance evaluation through the silhouette coefficient, the results demonstrated that clustering into six clusters had a silhouette score of 0.7395, regarded as the highest average score for clustering compared to others (Table 1):

TABLE I
EVALUATION OF CLUSTER'S DISTRIBUTION USING SILHOUETTE COEFFICIENT

Number of clusters	Silhouette coefficient
2	0.6834
3	0.7197
4	0.7062
5	0.7110
6	0.7395
7	0.5793
8	0.4361
9	0.4443
10	0.3725

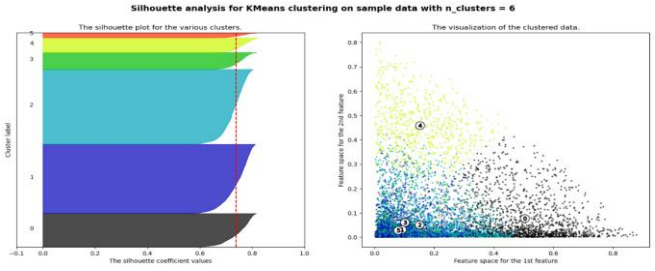

FIGURE 9
SILHOUETTE ANALYSIS OF TOPICS

Figure 9 illustrates the silhouette graph measuring the closeness between points in the same cluster and that between other clusters. The x-axis represents the silhouette values, while the y-axis refers to the number of clusters, with 6 clusters in particular. The width in the y-axis indicates the number of texts in each topic, and the red dotted line shows the average silhouette scores. Comparing the average silhouette scores, it can be seen that all of the topics had close scores. In addition, the graph points out that the clustering was good, with a small mix of texts though. LDA analysis results which presents keywords grouping and dominant topics are shown in Table 2.

TABLE II
LDA ANALYSIS RESULTS WHICH PRESENTS KEYWORDS GROUPING AND DOMINANT TOPICS

Topic No.	Topic label	Keywords	Dominant topic
1	make money from trade market	make, market, go, money, trader, think, get, know, time, people	1,491
2	stock market, gold and investment	stock, market, price, gold, option, investor, money, year, buy, investment	648
3	trading strategy and risk/profit	trade, trading, trader, strategy, risk, use, profit, time, position, take	3,354
4	candlestick, chart pattern and impulse wave	candle, wave, close, impulse, body, pattern, candlestick, correction, triangle, bar	209
5	trend, high/low price pattern and support/resistance	price, trend, level, high, pattern, low, support, line, market, resistance	3,101
6	indicator and chart analysis	indicator, market, use, price, change, period, chart, trend, analysis, time	750

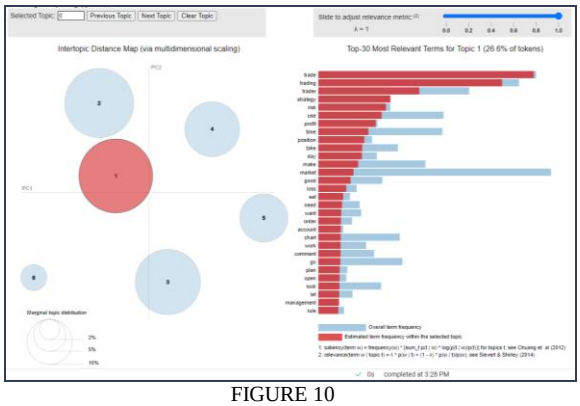

FIGURE 10
AN EXAMPLE OF VISUALIZATIOIN OF LDA TOPIC MODELING RESULTS

As shown in Figure 10, the graphs display visualization using library pyLDAvis to show the results from the LDA model (4D), and PCA was performed to visualize the figures in 6 circles. The center of each circle represents the position of the topic in the latent feature space, while the distance between topics illustrates the degree of similarities between

each topic. Additionally, the bar charts on the right-side show terms in each topic with reducing relevance; each bar specifies the frequency of a particular word, with the blue one representing its overall frequency in a corpus and the red one showing its frequency in a particular topic.

DISCUSSION

The experiment was conducted using Colab to construct a model for analysis of significant words and topics from 9,553 texts. Keywords in each topic were compared to the topics hypothesized in this study. Based on the proposed hypothesis, 6 topics were as follows: (1) make money from trade market, (2) stock market and investment, (3) trading strategy and risk/profit, (4) candlestick, chart pattern, and impulse wave, (5) trend, price pattern and support/resistance, and (6) indicator and chart analysis. The results are presented below with the comparison between the hypothesis and the prediction of topic modeling and analysis of significant words in each topic (Figure 11).

- Topic 1 concerned making money from the trade market with the following relevant terms: make, market, go, money, trader, think, get, know, time, and people.
- Topic 2 was about the stock market, gold, and investment, which related words were stock, market, price, gold, option, investor, money, year, buy, and investment.
- Topic 3 was related to trading strategy and risk/profit with related words: trade, trading, trader, strategy, risk, use, profit, time, position, and take.
- Topic 4 concerned candlestick, chart pattern, and impulse wave with the following relevant words: candle, wave, close, impulse, body, pattern, candlestick, correction, triangle, and bar.
- Topic 5 was about trend lines, high/low price patterns, and support/resistance with the related words: price, trend, level, high, pattern, low, support, line, market, and resistance.
- Topic 6 was related to indicator and chart analysis with the following relevant words: indicator, market, use, price, change, period, chart, trend, analysis, and time.

	Id	user	Text	Topic_0	Topic_1	Topic_2	Topic_3	Topic_4	Topic_5	Dominant_Topic	Perc_Dominant_Topic
0	1	MrRenevPro	top quality speculate success personal list ad...	0.748538	0.080906	0.103083	0.000135	0.046305	0.021033	Topic 0	0.75
1	2	fringe_chartist	chinese large cap index update version previou...	0.130884	0.507354	0.160889	0.005966	0.036649	0.158258	Topic 1	0.51
2	3	lexcaraigincPro	bullish trade bullish order_block look upside ...	0.038219	0.015274	0.125323	0.007634	0.689143	0.124408	Topic 4	0.69
3	4	lexcaraigincPro	bearish trade bearish order_block look downsid...	0.036227	0.014491	0.120343	0.007275	0.645356	0.176308	Topic 4	0.65
4	5	zAngus	find good trend momentum stock ask time find s...	0.165676	0.059132	0.292141	0.000773	0.229691	0.252586	Topic 2	0.29
...	...	...	...	...	...	...	...	...	...	...	...
9548	9549	Tradersvoices	timely time stare screen yield profit	0.081605	0.135355	0.461245	0.015272	0.098608	0.207915	Topic 2	0.46
9549	9550	Andy_HechtPremium	go part follow technomental analytic approach ...	0.168436	0.163307	0.065779	0.001672	0.236497	0.364110	Topic 5	0.36
9550	9551	Paul_WankmuellerPremium	trading psychologist hear phrase could pertain...	0.676356	0.045949	0.198532	0.006258	0.012570	0.061344	Topic 0	0.68
9551	9552	m.malekijahan	wheat oil year channel access see correlation ...	0.043471	0.369805	0.139237	0.005652	0.306849	0.134985	Topic 1	0.37
9552	9553	The_STAPremium	importance resistance long term chart pop char...	0.050198	0.114449	0.077874	0.001423	0.245845	0.510210	Topic 5	0.51

9553 rows × 11 columns

FIGURE 11
COMPARISON BETWEEN THE HYPOTHESIS AND THE PREDICTION OF TOPIC MODELING

The analysis of the hypothesis and the prediction showed that the users most frequently shared ideas in Topic 3, with 3,354 texts; the majority of the users considered that trading strategy and risk/profit was the most popular topic. That is followed by Topic 5, with 3,101 texts, in relation to the trend line, high/low price pattern, and support/resistance. Both of the topics accounted for 67.56%. This can be implied that investments in digital assets require good strategies, risk management, analysis of price patterns, and knowledge of support and resistance. The analysis of significant words and topics in the texts through the LDA technique showed a larger number of words latent in the texts than hypothesized

RECOMMENDATION

- The present study adopted only a single method, so other methods may analyze significant words and topics more effectively.
- During data preparation, incorrect words were not corrected, so they were filtered out in the word tokenization process. Therefore, correcting those incorrect words may contribute to identifying more significant words or analyzing topics more accurately.
- In this study, there were no experts, particularly sophisticated users, to help interpret the results to measure the accuracy of the prediction of topic modeling. Thus, the cooperation from the experts will help ensure more effective topic modeling or increase the accuracy of performance evaluation.
- This model can be improved and applied to the analysis of people's opinions to predict the prices of assets and determine a set of ideas for future research.

REFERENCES

[1] J. Frankenfield, "Digital Assets," Jun. 30, 2022. https://www.investopedia.com/terms/d/digital-asset-framework.asp

[2] N. Suepaisal, "What is digital asset?," Nov. 2021. https://thematter.co/futureverse/futureword-digital-asset/160461

[3] SEC, "Digital asset entrepreneurs.," *The Securities and Exchange Commission, Thailand*, 2022. https://www.sec.or.th/TH/pages/lawandregulations/digitalassetbusiness.aspx

[4] Coinmarketcap, "Cryptocurrency Prices, Charts And Market Capitalizations," Oct. 2022. https://coinmarketcap.com/

[5] Statista, "Digital Economy Compass 2022 Chapter 1: The ascent of the crypto economy," May 2022. [Online]. Available: https://www.statista.com/study/112911/digital-economy-compass-2022-chapter-1/

[6] Chainalysis, "The 2022 Geography of Cryptocurrency Report," Chainalysis, Oct. 2022. [Online]. Available: https://go.chainalysis.com/geography-of-crypto-2022-report.html

[7] S. Kemp, "Digital 2022: Global overview report," Jan. 2022. [Online]. Available: https://datareportal.com/reports/digital-2022-global-overview-report

[8] L. Xueyan, "Public Project Summary," Central European University Budapest, Hungary, 2020.

[9] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using topic modeling methods for short-text data: A comparative analysis," *Front. Artif. Intell.*, vol. 3, p. 42, 2020.

[10] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J Mach Learn Res*, vol. 3, no. null, pp. 993–1022, Mar. 2003.

[11] D. M. Blei, "Probabilistic Topic Models," *Commun ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012, doi: 10.1145/2133806.2133826.

[12] S.-W. Kim and J.-M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Hum.-Centric Comput. Inf. Sci.*, vol. 9, no. 1, p. 30, Aug. 2019, doi: 10.1186/s13673-019-0192-7.

[13] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 1354–1364.

[14] S. Twinandilla, S. Adhy, B. Surarso, and R. Kusumaningrum, "Multi-document summarization using k-means and latent dirichlet allocation (lda)–significance sentences," *Procedia Comput. Sci.*, vol. 135, pp. 663–670, 2018.

[15] M. Posner, "Very basic strategies for interpreting results from the Topic Modeling Tool," *MIRIAM POSNER'S BLOG Digital humanities, data, labor, and information*, Oct. 29, 2012. https://miriamposner.com/blog/very-basic-strategies-for-interpreting-results-from-the-topic-modeling-tool/

[16] P. Xie and E. P. Xing, "Integrating Document Clustering and Topic Modeling," *ArXiv*, vol. abs/1309.6874, 2013.

[17] X. Cheng, X. Yan, Y. Lan, and J. Guo, "BTM: Topic Modeling over Short Texts," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 2928–2941, 2014, doi: 10.1109/TKDE.2014.2313872.

[18] Y. Tian, P. S. Kochhar, E.-P. Lim, F. Zhu, and D. Lo, "Predicting best answerers for new questions: An approach leveraging topic modeling and collaborative voting," in *International Conference on Social Informatics*, 2013, pp. 55–68.

[19] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Toward A Real-Time Social Recommendation System," in *Proceedings of the 11th International Conference on Management of Digital EcoSystems*, New York, NY, USA, 2020, pp. 336–340. doi: 10.1145/3297662.3365789.

[20] S. Junlabuddee and K. Tuamsuk, "Analysis of Research Data in Information Science Using the Topic Modeling Method," *J. Mekong Soc.*, vol. 17, no. 1, pp. 89–109, Apr. 2021.

[21] [M. Linton, E. G. S. Teo, E. Bommes, C. Y. Chen, and W. K. Härdle, "Dynamic topic modelling for cryptocurrency community forums," in *Applied quantitative finance*, Springer, 2017, pp. 355–372.

[22] H. Schnoering, "Short Text Topic Modeling: Application to tweets about Bitcoin," *ArXiv Prepr. ArXiv220311152*, 2022.

[23] P. Siehndel and U. Gadiraju, "Unlock the Stock: User Topic Modeling for Stock Market Analysis.," in *EDBT/ICDT Workshops*, 2016.