Received 30 October 2023, accepted 16 November 2023, date of publication 20 November 2023,
date of current version 27 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3334617

APPLIED RESEARCH

Unveiling Cryptocurrency Conversations: Insights From Data Mining and Unsupervised Learning Across Multiple Platforms

HAE SUN JUNG[1], **HAEIN LEE**[1,2], **AND JANG HYUN KIM**[2,3]

[1]Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul 03063, South Korea
[2]Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul 03063, South Korea
[3]Department of Interaction Science, Sungkyunkwan University, Seoul 03063, South Korea

Corresponding author: Jang Hyun Kim (alohakim@skku.edu)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government under Grant RS-2023-00208278.

ABSTRACT The rapid growth of the cryptocurrency market has led to an increasing interest in the subject. Cryptocurrency is now recognized as an asset, and laws and financial regulations have begun to emerge for supporting its practical use. As a result, it has become essential to perform data mining and attain knowledge from text data related to cryptocurrency. Previous studies have focused on analyzing data from a single source such as Twitter. However, there are unique insights to be gained from data across multiple platforms. In the present study, we utilized data mining techniques to extract insights from LexisNexis, Web of Science, and Reddit, representing the media, academia, and general public, respectively. Among unsupervised learning technologies, topic modeling was employed for the analysis. Topic modeling is a methodology that uncovers hidden meanings within the collected data. Among the diverse topic modeling techniques available, bidirectional encoder representations from transformers topic was chosen for the analysis. BERTopic considered to be state-of-the-art in the field of topic modeling. Dynamic topic modeling was employed to track changes in themes over time. Our experimental results reveal a tendency in the news to cover major events related to cryptocurrencies, such as regulatory developments and market trends. Academic papers, on the other hand, tend to focus on the technology behind cryptocurrencies and related research. Finally, social media conversations center more around information delivery from an investor's psychological perspective, such as market sentiment and investment strategies.

INDEX TERMS Bitcoin, cryptocurrency, data mining, machine learning, natural language processing, unsupervised learning, topic modeling, BERTopic.

I. INTRODUCTION

Bitcoin represents the first cryptocurrency that utilizes decentralized cryptographic technology – namely blockchain – allowing for payments or transfers between parties over a short period without relying on financial institutions [1]. The successful launch and growth of Bitcoin have triggered the creation of many other cryptocurrencies, which in turn has substantially expanded the cryptocurrency market, driven primarily by notable price increases [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Rongbo Zhu.

Despite the growing interest in cryptocurrency market, there exists a research gap marked by a deficiency in achieving a comprehensive understanding of the major themes and perceptions surrounding cryptocurrency. This gap is particularly pronounced due to the relatively short history of Bitcoin compared to other assets.

To address the existing research gap and contribute to the body of knowledge, the authors employed topic modeling, an unstructured text data mining method, incorporating data from three distinct sources: LexisNexis, Web of Science, and Reddit. For the analysis, query "BTC" and "Bitcoin" were employed based on previous research that suggests level of interest on Bitcoin can serve as an indicator of the broader interest in cryptocurrencies [3], [4], [5]. This approach enabled authors to extract valuable insights from a wide spectrum of sources, encompassing mass media, academia, and social media, providing a more comprehensive perspective on the subject.

Topic modeling is a methodology that automatically clusters words based on statistics and machine learning [6]. The method can help attain insights by extracting the main theme from a given data source [7]. Although numerous topic modeling methods, such as latent Dirichlet allocation (LDA) and non-negative matrix factorization (NMF) are available, but in this study, the state-of-the-art (SOTA) model known as bidirectional encoder representations from transformers topic (BERTopic) was applied [8]. BERTopic is a topic modeling technique that employs Bidirectional Encoder Representations from Transformers (BERT) embeddings and class-based TF-IDF (c-TF-IDF) to formulate compact clusters that are easily explicable while preserving important words in topic explanations [9]. Additionally, BERTopic is recognized for its modularity, enabling the incorporation and utilization of diverse algorithms within its framework. Considering the previously mentioned advantages, the authors have selected modified BERTopic as the most suitable algorithm for investigating perceptions related to Bitcoin.

At the beginning of the analysis on each specific platform, the authors assessed the coherence values of baseline models (LDA, NMF) and the modified BERTopic. This process confirmed the capability of BERTopic to accurately depict topics in the chosen domain. In summary, modified BERTopic demonstrated the highest coherence values in all three analyses, resulting in the subsequent findings derived from its application.

The results topic representations revealed that news sources primarily cover major events, academic journals focus on technological advancements, and social media platforms discuss the sentiments of cryptocurrency investors, affirming the distinct characteristics within each domain.

In summary, our research not only addresses the existing research problem but also endeavors to bridge the research gap by offering a comprehensive analysis of Bitcoin-related text data from diverse sources. The insights gained from this study have the potential to advance understanding of the challenges and issues faced by blockchain technology, setting the stage for future developments in this field. In addition, the utilization of modified BERTopic across various platforms constitutes a novel approach that remains notably unexplored within the existing research.

II. RELATED WORKS

The following subsections present brief reviews of literature that aim to provide insight by applying natural language processing (NLP) to Bitcoin. Subsequently, the concept of topic modeling is explained, and studies that sought to obtain knowledge from various domains using topic modeling are exemplified. Lastly, the default structure and advantages of using BERTopic are illustrated.

A. RESEARCH ON BITCOIN USING NATURAL LANGUAGE PROCESSING

NLP is utilized to extract valuable information from texts across various fields, with the objective of deriving insights and practical applications. The current section presents literature reviews on prior studies that utilized NLP to gain insights into Bitcoin.

In [10], the relation between social media topic discussion and cryptocurrency market price fluctuations was analyzed via statistical and NLP models (i.e., DMR). In [11], NLP algorithms were used to measure the relationship between investment sentiment and bitcoin price fluctuations using data from the subreddits "r/bitcoin" and "r/investing." In [12], a prediction was made on the direction and magnitude of Bitcoin price fluctuations using sentiment analysis and post volume extracted from Twitter data. A relative accuracy of 63% was achieved using a model based on recurrent neural networks (RNN) and convolutional neural networks (CNN). Satarov et al. [13] confirmed that sentiment analysis of tweets related to Bitcoin can be used to predict Bitcoin price changes. An accuracy of 62.48% was attained by a random forest regression model when applying sentiment analysis to Twitter data. Jung et al. [14] aimed to predict Bitcoin price trends by analyzing both the volume and sentiment of Reddit data, as well as technical indicators of chart analysis. The authors achieved an accuracy of 90.57% and an area under the curve (AUC) value of 97.48% using an extreme gradient boosting (XGBoost) model.

B. TOPIC MODELING

Topic modeling is a statistical model used in the field of NLP to discover abstract main themes, referred to as topics, within sets of documents [15]. In other words, it is a text mining technique that is utilized to uncover hidden semantic structures within textual data. Examples of representative topic modeling technologies are DMR and LDA.

Yin and Yuan [16] employed LDA topic modeling to analyze research subjects and progress trends related to blended learning using keyword analysis, with results showing that the ratio of element topics in blended learning has been increasing every year. Moreover, the text analysis provided theoretical and methodological reference materials to facilitate future research. Polyzos and Wang [17] conducted an LDA analysis on Twitter data to quantify energy market efficiency. The extracted topic was then applied to a classification model to measure prediction accuracy for market movements. Sharma and Sharma [18] collected research papers related to blockchain technology from various databases and attempted to create a semantic map using the LDA model. Through a metadata analysis, an abstract perspective of blockchain was attained. Avasthi et al. [19] conducted a comparison of various topical models, including LDA, correlated topic model (CTM), hierarchical Dirichlet process (HDP), and DMR, using adolescent drug use and depression as keywords. Egger and Yu [20] applied LDA, NMF, and BERTopic to Twitter posts and conducted a comparative analysis for each topic modeling algorithm.

C. DEFAULT BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS TOPIC (BERTOPIC) STRUCTURE

BERTopic is a SOTA framework for topic modeling technology that consists of five sub models, each of which can be selected and used independently [9]. In the current section, the default structure of BERTopic is presented.

1) DOCUMENT EMBEDDING

The default BERTopic approach utilizes sentence-transformers (SBERT) to convert text data into numerical representations, allowing for exclusive semantic similarity that significantly enhances clustering tasks in comparison to LDA [21].

2) DIMENSIONALITY REDUCTION

Because clustering models struggle with high-dimensional data due to the curse of dimensionality, it is essential to perform dimensionality reduction after obtaining representation [22]. By default, BERTopic utilizes uniform manifold approximation and projection (UMAP) for this task. UMAP is a dimensionality reduction procedure that preserves both local and global structures in data, thereby enabling the clustering of semantically similar documents [23].

3) CLUSTERING DOCUMENTS AND BAG OF WORDS

Default BERTopic uses hierarchical density-based spatial clustering of applications with noise (HDBSCAN), a density-based clustering technique that clusters text data [24]. This approach allows for outlier detection and the identification of different cluster shapes, preventing text data from being forcibly included in the wrong cluster.

Because HDBSCAN generates clusters with varying densities and shapes, centroid-based topic presentation techniques may not be suitable. Instead, BERTopic combines all documents within each cluster into a single document to create a cluster-level bag-of-words (BoW) that records the frequency of each word in each cluster. This holds significance since topic modeling primarily examines words at the topic (i.e., cluster) level.

4) ATTAINING TOPIC REPRESENTATION

Finally, L1 normalization is applied to the BoW representation to account for clusters of varying sizes. Term Frequency-Inverse Document Frequency (TF-IDF) for BoW must consider topics or clusters, rather than individual documents. By extracting the most significant words from each cluster, it is possible to obtain an explanation of the topic. This model is known as class-based TF-IDF (c-TF-IDF).

$$W_{x,c} = \left\| tf_{x,c} \right\| \cdot log\left(1 + \frac{A}{f}\right) \tag{1}$$

In the aforementioned equation, $tf_{x,c}$ is the frequency in class c of word x, f_x is the frequency of word x among all classes, and A is the average number of words per class. Similar to the traditional TF-IDF approach, the importance score of a word in each class is obtained by multiplying the term frequency $tf_{x,c}$ and inverse document frequency $log(1 + \frac{A}{f_x})$.

In contrast to traditional topic modeling methodologies, BERTopic stands out by leveraging pretrained language models for document and word representations, making it adept at capturing complex relationships between words and context. Furthermore, its non-linear dimensionality reduction approach and modularity enable BERTopic to enhance the quality of topic representation compared to conventional methods.

This study is meaningful in that it is the first to combine the keyword 'Bitcoin' with modified BERTopic, the SOTA in topic modeling, to obtain knowledge from three distinct sources - LexisNexis, Web of Science, and Reddit.

III. METHOD

The following subsections outline the experimental procedures followed in this study. Firstly, the sources and descriptions of collected data are represented. Secondly, the preprocessing steps undertaken for the text data are presented. Finally, the application of BERTopic for deriving topic modeling results is explained.

A. DATA COLLECTION

It is possible to gauge the general sentiment toward cryptocurrency by analyzing data on Bitcoin, which is representative of blockchain technology and cryptocurrency [3], [4], [5]. Accordingly, all data examined in this study were collected using search queries for "Bitcoin" and "BTC." To interpret the sentiment toward cryptocurrency from media, academic, and public perspectives, data were collected from LexisNexis, Web of Science, and Reddit, respectively. Data obtained from LexisNexis comprise the full body of news articles, whereas those collected from Web of Science encompass the abstracts of academic papers, and those collected from Reddit encompass both posts and comments from the r/Bitcoin and r/BTC subreddits. The data encompass a period of six years from March 1, 2017, to March 1, 2023. In total, 17,230 news articles, 9,520 academic papers, and 10,914,149 social media texts were collected.

B. DATA PREPROCESSING

First, any instances of data wherein the acronym BTC was used for extraneous contexts (e.g., Cu-BTC, Biliary tract cancer) were eliminated. All data that use English spelling to express other languages, as well as duplicates and missing values, were eliminated. A total of 6,011 academic papers,

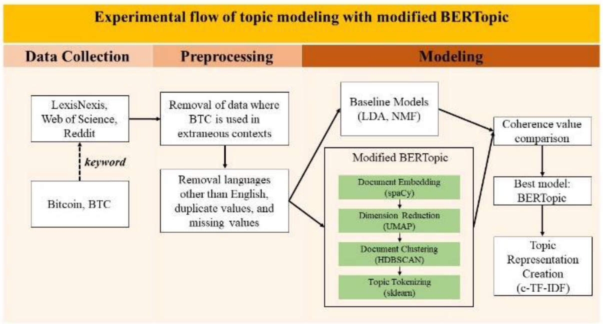

FIGURE 1. **Experimental flow diagram of topic modeling with modified BERTopic.**

10,925 news articles, and 9,830,800 social media texts were used in the analysis.

C. APPLICATION OF MODIFIED BERTOPIC

In this study, the authors partially modified the default BERTopic structure for the diversity of topic representations.

1) DOCUMENT EMBEDDING

Although SBERT is the default embedding model, the authors utilized spaCy as an alternative embedding algorithm [25]. One of the motivations for utilizing spaCy lies in its ability to deliver rapid processing speed while ensuring higher quality in capturing the textual semantics and maintaining pertinent information during the embedding generation process.

2) DIMENSIONALITY REDUCTION

The authors adopted UMAP in the dimensionality reduction phase over other algorithms as it effectively maintains the non-linear and complex structures in textual data.

3) CLUSTERING DOCUMENTS AND BAG OF WORDS

Consequently, the authors adopted HDBSCAN because it reduces noise and enhances topic representation. Additionally, Scikit-Learn's tokenizer was utilized to generate BoW at the cluster level.

4) ATTAINING TOPIC REPRESENTATION

As a result, topic representation was generated using c-TF-IDF to confirm the results. For each topic, the words with the highest c-TF-IDF score were provided. An experimental diagram has been constructed to illustrate the flow of the experiment (Fig. 1), and the implementation setup and hyperparameters can be found in Table 1.

IV. EXPERIMENTS

The following subsections describe the differences in coherence values between baseline models (LDA, NMF) and BERTopic, as well as the insights gained through BERTopic analysis on data collected from three domains: LexisNexis, Web of Science, and Reddit.

TABLE 1. **Implementation setup.**

Package version			
joblib 1.1.0 bertopic 0.15.0			
Hyperparameters			
Function / Platform	UMAP	HDBSCAN	BERTopic
LexisNexis	metric= 'cosine'	metric= 'euclidean', cluster_selection_method ='eom'	nr_topics=8
Web of Science			nr_topics=9
Reddit			nr_topics=4

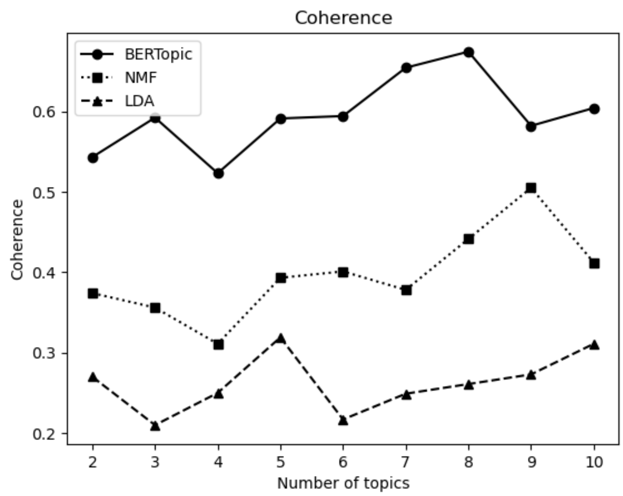

FIGURE 2. **Comparison of coherence values with baseline models on LexisNexis data.**

A. COMPARISON OF COHERENCE VALUES BETWEEN BERTOPIC AND BASELINE MODELS

Coherence serves as a measure for assessing the interpretability, consistency, and meaningfulness of topic modeling outcomes. Typically, a higher coherence value in the results of topic modeling is regarded as a sign of a more effective topic model [26]. Additionally, the use of coherence allows for the extraction of the optimal number of topics.

To ensure the validity of the analysis using BERTopic in each domain and to determine the optimal number of topics, the authors first compared it with baseline models based on coherence values (Fig. 2, Fig. 3, Fig. 4).

Although differences in values were observed, it was confirmed that BERTopic exhibited the highest coherence values for all three target platforms. It was also confirmed that, for each platform, using 8, 9, and 4 topics was suitable for extracting the optimal representation.

The tendency appeared in previous studies comparing topic models. BERTopic demonstrates its value as SOTA by generating insights from short and unstructured text, ensuring high stability and diversity across various domains [20], [27]. Consequently, it outperforms recently introduced models that lack application in diverse domains.

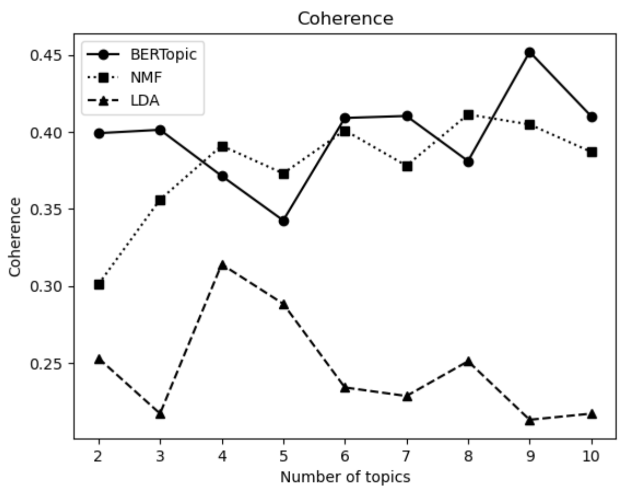

FIGURE 3. Comparison of coherence values with baseline models on Web of Science data.

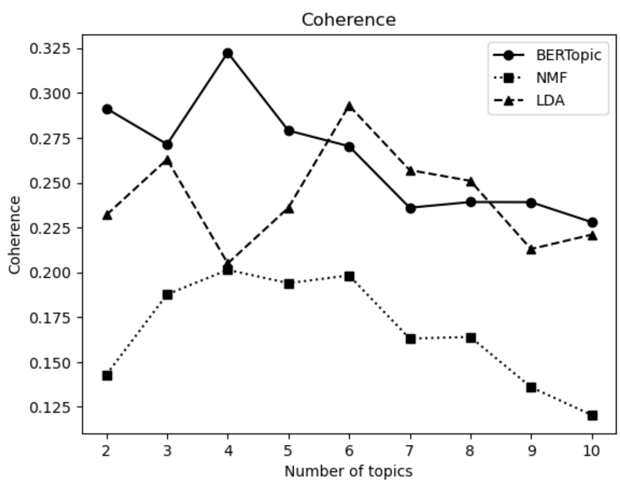

FIGURE 4. Comparison of coherence values with baseline models on Reddit data.

B. TOPIC ANALYSIS FROM LEXISNEXIS DATA

An analysis of LexisNexis data was conducted to gain insights into the media's perception and knowledge of Bitcoin. Each topic presentation was evaluated using a c-TF-IDF score.

Topic 0 was denoted as "Bitcoin as a digital currency" through "cryptocurrency," "currency," "digital," and "bank." There are ongoing discussions regarding whether Bitcoin or other cryptocurrency technologies can function as digital currencies and potentially replace traditional currencies. Grinberg [28] addressed various aspects of Bitcoin, including its innovative nature as a decentralized digital currency, its relationship with the United States (U.S.) dollar as a global reserve currency, and the legal challenges that Bitcoin faces in terms of regulation and adoption. Dwyer [29] emphasized that Bitcoin can be transferred peer-to-peer without bank intervention. The authors also mentioned the advantage of Bitcoin as a digital currency that prevents double spending of wire transfer fees using open-source software.

Topic 1 was designated as "Energy consumption of Bitcoin mining" through "energy," "mining," "power," "tesla," "cryptocurrency," and "electricity." The mining of cryptocurrency using the proof-of-work (PoW) methodology and its associated energy consumption have been topics of ongoing discussion among experts in related fields. Li et al. [30] investigated the energy consumption of cryptocurrency mining, and emphasized the need for energy conservation and sustainable development despite the promising potential of blockchain technology. O'Dwyer and Malone [31] examined the energy consumption acquired for bitcoin mining, and discovered that the aggregate energy consumption utilized in Bitcoin mining was equivalent to that of Ireland.

Topic 2 was summarized as "Cryptocurrency-friendly countries" through "salvador," "ukraine," "country," "government," "bitcoin," "cryptocurrency," "bukele," "dollar," and "Venezuela." An increasing number of countries are designating cryptocurrency as legal tender or expressing a cryptocurrency-friendly stance. For instance, President Nayib Bukele of El Salvador made the move of recognizing Bitcoin as a legal currency, Ukraine has legalized Bitcoin, and Venezuela is seeing a rise in cryptocurrency usage [32], [33], [34].

Topic 3 was named "Taxation of cryptocurrency income in the United States" through "tax", "cryptocurrency," "irs," "government," "income," "bill," and "taxpayer." The United Stastes Treasury Department has required the Internal Revenue Service (IRS) to report cryptocurrency transactions worth more than $10,000 [35], [36]. In addition, plans for cryptocurrency taxation are being continuously discussed globally [37], [38].

Topic 4 was labeled "Conflicts between Berkshire Hathaway and cryptocurrency companies" through "robinhood," "gemini," "buffett," "winklevoss," "investor," and "berkshire." Prominent stock investors Warren Buffett and Charlie Munger are known for their negative views on cryptocurrency [39]. As a result, their remarks often draw opposition from companies in the cryptocurrency industry [40].

Topic 5 was described as "Singapore's cryptocurrency-related regulations" through "Singapore," "mas," "binance," and "regulation." Binance, the largest cryptocurrency exchange, made efforts to obtain exchange approval from the Singapore government since 2020, however, it eventually ceased its services in 2022 due to the Monetary Authority of Singapore's (MAS) strict regulations regarding the approval of cryptocurrency exchanges [41], [42].

Topic 6 was designated as "Cryptocurrency as a means of donation" through "donor," "donation," "charity," "donate," "cryptocurrency," and "dafs." The faster and easier remittance processes offered by cryptocurrency technology compared to those in conventional finance, as well as the transparency of transactions, have led to the emergence of a donation culture using cryptocurrency [43], [44]. As a representative example, during the Russian-Ukrainian War, donations using

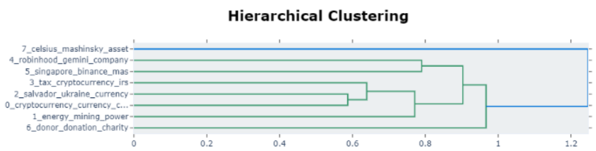

FIGURE 5. Hierarchical clustering from LexisNexis data.

cryptocurrency have been collected on an international scale [45].

Topic 7 was presented as "Cryptocurrency lending platform's bankruptcy" through "celsius," "mashinsky," "customer," "lender," "bankruptcy," "withdrawal," and "deposit." The bankruptcy of multiple cryptocurrency lenders following the collapse of Terra and Luna in 2022 resulted in significant losses for cryptocurrency investors, leading to widespread distrust and negative effects on the cryptocurrency market [46], [47]. These events highlight the need for international cooperation among relevant organizations to establish standards and promote collaboration [48].

Results from LexisNexis revealed that media predominantly covers major issues related to the cryptocurrency market (Fig. 5, Table 2). Beginning with the function of cryptocurrency as a digital currency, both the advantages and disadvantages have been covered extensively. Additionally, the social and legal challenges faced by cryptocurrency have been identified, providing insight into problems that must be addressed in the future.

C. TOPIC ANALYSIS FROM WEB OF SCIENCE DATA

An analysis of Web of Science data was conducted to understand the perception and knowledge of Bitcoin in academia. Each topic presentation was evaluated using a c-TF-IDF score.

Topic 0 was denoted as "Contributions and applications of blockchain technology" through "blockchain," "technology," "system," "node," and "application." Although M. Hashemi Joo et al. [49] addressed the use of blockchain technology in cryptocurrency, it is important to note that blockchain technology can be utilized in a wide range of fields, and research regarding its use in other industries is actively underway [50], [51], [52].

Topic 1 was designated as "Bitcoin as an investment asset" through "market," "cryptocurrency," "volatility," "asset," "gold," "price," and "portfolio." Cryptocurrency continues to attract more attention as an investment asset rather than a real-life application of technology [53]. Because cryptocurrencies are now considered a new asset class, governments around the world are beginning to introduce relevant regulations. In this trend, studies have been conducted on whether Bitcoin can be included in the investment portfolio and used as a hedging instrument [54], [55], [56].

Topic 2 was summarized as "Bitcoin's PoW mechanism" through "mining," "btc," "miner," "pool," "block," "energy," and "reward." PoW is a consensus algorithm

TABLE 2. Topic representations of bitcoin from lexisnexis.

Topic number	Main theme	Topic representations (c-TF-IDF score)
0	Bitcoin as a digital currency	('cryptocurrency', 0.023), ('currency', 0.0226), ('company', 0.0218), ('bitcoin', 0.0195), ('digital', 0.0190), ('market', 0.0178), ('bank', 0.0158), ('industry', 0.0154), ('news', 0.0148), ('price', 0.0148)
1	Energy consumption of Bitcoin mining	('energy', 0.0678), ('mining', 0.0502), ('power', 0.0414), ('tesla', 0.0369), ('bitcoin', 0.0356), ('musk', 0.0344), ('cryptocurrency', 0.0299), ('electricity', 0.0273), ('industry', 0.0270), ('electric', 0.0250)
2	Cryptocurrency-friendly countries	('salvador', 0.0435), ('ukraine', 0.0341), ('currency', 0.0322), ('country', 0.0318), ('government', 0.0312), ('bitcoin', 0.0280), ('cryptocurrency', 0.0273), ('bukele', 0.0246), ('dollar', 0.0218), ('venezuela', 0.0210)
3	Taxation of cryptocurrency income in the United States	('tax', 0.0997), ('cryptocurrency', 0.0379), ('irs', 0.0289), ('government', 0.0279), ('currency', 0.0234), ('income', 0.0221), ('industry', 0.0201), ('state', 0.0200), ('bill', 0.01917), ('taxpayer', 0.0182)
4	Conflicts between Berkshire Hathaway and cryptocurrency companies	('robinhood', 0.0687), ('gemini', 0.0607), ('company', 0.0360), ('buffett', 0.0356), ('trading', 0.0337), ('stock', 0.0322), ('winklevoss', 0.0312), ('securities', 0.0281), ('investor', 0.0269), ('berkshire', 0.0255)
5	Singapore's cryptocurrency-related regulations	('singapore', 0.0845), ('binance', 0.0636), ('mas', 0.0344), ('finance', 0.0295), ('digital', 0.0245), ('cryptocurrency', 0.0245), ('industry', 0.0244), ('banking', 0.0240), ('regulation', 0.0235), ('technology', 0.0234)
6	Cryptocurrency as a means of donation	('donor', 0.0648), ('donation', 0.0561), ('charity', 0.0521), ('charitable', 0.0408), ('gift', 0.0363), ('donate', 0.0356), ('cryptocurrency', 0.0332), ('fidelity', 0.0327), ('organization', 0.0325), ('dafs', 0.0291)
7	Cryptocurrency lending platform's bankruptcy	('celsius', 0.2181), ('mashinsky', 0.0610), ('asset', 0.0450), ('customer', 0.0393), ('lender', 0.0389), ('company', 0.0378), ('bankruptcy', 0.0359), ('cryptocurrency', 0.0355), ('withdrawal', 0.0337), ('deposit', 0.0324)

that reflects task participation by iteratively searching for a hash value below the target threshold [57]. Because this process consumes a significant amount of power, there is ongoing discussion and research to reduce its energy consumption while maintaining the security and integrity of the system [58].

Topic 3 was named "Bitcoin price prediction through sentiment analysis" through "model," "price," "prediction," "sentiment," "learning," "volatility," "forecast," and "machine." Among machine learning models, studies have been actively conducted to predict the price of Bitcoin, especially using sentiment analysis. In [59], sentiment scores attained using Valence Aware Dictionary and sEntiment Reasoner (VADER) over 2019 and 2020 were found to correlate with short-term trends in Bitcoin prices. In [14], the trend of Bitcoin prices was predicted by combining technical and sentiment analyses.

Topic 4 was labeled "Studies on the factors affecting the Bitcoin price" through "study," "research," "btc," "analysis," and "factor." Because the cryptocurrency market experienced rapid growth, research on Bitcoin price variables has been conducted actively. R. Hakim das Neves [60] found interest in Bitcoin to precede price hikes, whereas T. Panagiotidis et al. [61] indicated that changes in interest rates and exchange rates also affect the price of Bitcoin.

Topic 5 was described as "The digital signature algorithm in cryptocurrency wallets" through "signature," "wallet," "key," "protocol," "security," and "ecdsa." As the possibility of losing cryptocurrency due to bankruptcy or hacking increased, the security of cryptocurrency wallets became more important [62], [63]. Consequently, research has been conducted on utilizing digital signature algorithms such as the Elliptic Curve Digital Signature Algorithm (ECDSA) in cryptocurrency wallets [64], [65].

Topic 6 was stated as "The need of regulations in token Initial coin offering (ICO)" through "ico," "token," "icos," "offering," "capital," "coin," "scam," and "regulation." ICO is a method of raising investment funds for a cryptocurrency project by transferring a portion of a newly developed cryptocurrency to investors in exchange for cash or other cryptocurrencies [66]. Whereas the ICO process has led to the creation of successful tokens, there have also been instances of scams and abuse, which have highlighted the need for appropriate regulation [67], [68], [69].

Topic 7 was presented as "Hacking attacks on cryptocurrency" through "ransomware," "ransom," "malware," "attack," and "victim." Despite its recent growth, the cryptocurrency market remains vulnerable to hacking and money laundering due to the inherent nature of its coding base structure, leading to a constant need for improved security measures [70], [71], [72].

Topic 8 was classified as "Resolving transaction processing speed and scalability problems through blockchain sharding" through "shard," "sharding," "blockchain," "transaction," "performance," and "scalability." Sharding, which was adopted as a solution to the slow transaction speed and scalability trilemma in blockchain, involves dividing the overhead a transaction process into smaller groups of nodes [73]. Studies on blockchain sharding are actively being conducted as the cryptocurrency market grows [74], [75].

Insights obtained from the Web of Science data have been confirmed to primarily focus on the technology, phenomena, and significance of blockchain in an academic context (Fig. 6, Table 3). The applications and conceptions of blockchain-related technology have been extensively discussed, shedding light on the technical challenges that cryptocurrencies must address and improve.

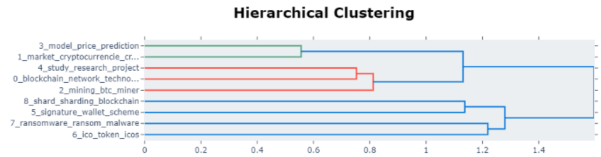

FIGURE 6. Hierarchical clustering from web of science data.

D. TOPIC ANALYSIS FROM REDDIT DATA

An analysis of Reddit data was conducted to understand the perception and knowledge of Bitcoin in social media. Each topic presentation was evaluated using a c-TF-IDF score.

Topic 0 was denoted as "Cryptocurrency investment recommendations" through "bitcoin," "btc," "buy," "money," "dip," and "want." The term "buy the dip" is a phrase commonly used in investing and trading, referring to the strategy of purchasing an asset at a lower price point during a market downturn, with the expectation that the asset will eventually recover and increase in value [76]. Expressions such as "buy," "money," and "want" can also be seen as words that encourage the purchase of bitcoin.

Topic 1 was referred to as "Negative view of the scam coin projects" through "shitcoin," "altcoin," "meatcoin," "fuck," "trashcoin," and "bullshit." There are negative opinions towards developers who attract money early in development or through the promotion of scam coins, only to abandon or cease development afterward. The derived words can be seen as expressions with negative connotations on the issues.

Topic 2 was summarized as "Cryptocurrency exchange recommendations" through "gemini," "coinbase," "trader," "use," "gdax," "fee," "blockfi," and "exchange." As the cryptocurrency market grew, the number of cryptocurrency exchanges also increased significantly. However, investors have suffered significant financial losses with the bankruptcy of certain exchanges [77]. As a result, users now have a wide range of options to choose from in terms of fees, convenience, and stability when selecting a cryptocurrency exchange. Extracted words can be perceived as the advantages and recommendations associated with a specific exchange.

Topic 3 was named "Cryptocurrency testnet and mainnet" through "testnet," "mainnet," "test," "connect," "network," "change," and "server." A blockchain mainnet is a live network that runs a blockchain project, whereas a testnet is a temporary network used during the development phase to build an independent mainnet [78]. News pertaining to a mainnet or testnet frequently have a positive effect on price. Obtained words may reflect investors' hopes that the launch of a mainnet will have a positive impact on prices.

Data from Reddit tended to feature words related to investor sentiment, rather than serious discussions or explanations of Bitcoin (Fig. 7, Table 4). Posts often contained favorable information such as updates on the mainnet and testnet, which can affect prices, as well as recommendations

TABLE 3. Topic representations of bitcoin from web of science.

Topic number	Main theme	Topic representations (c-TF-IDF score)
0	Contributions and applications of blockchain technology	('blockchain', 0.0665), ('network', 0.0376), ('technology', 0.0367), ('transaction', 0.0352), ('system', 0.0282), ('node', 0.0281), ('base', 0.0268), ('propose', 0.0267), ('application', 0.0243), ('paper', 0.0235)
1	Bitcoin as an investment asset	('market', 0.0580), ('cryptocurrencie', 0.0488), ('cryptocurrency', 0.0451), ('bitcoin', 0.0447), ('volatility', 0.0367), ('asset', 0.0345), ('return', 0.0331), ('gold', 0.0330), ('price', 0.0321), ('portfolio', 0.0292)
2	Bitcoin's PoW mechanism	('mining', 0.0616), ('btc', 0.0448), ('miner', 0.0438), ('pool', 0.0432), ('block', 0.0258), ('attack', 0.0246), ('energy', 0.0237), ('reward', 0.0218), ('model', 0.0217), ('mof', 0.0215)
3	Bitcoin price prediction through sentiment analysis	('model', 0.0844), ('price', 0.0669), ('prediction', 0.0571), ('sentiment', 0.0549), ('learning', 0.0397), ('predict', 0.03626), ('datum', 0.0327), ('volatility', 0.0315), ('machine', 0.0311), ('forecast', 0.0303)
4	Studies on the factors affecting the Bitcoin price	('study', 0.0415), ('research', 0.0394), ('project', 0.0275), ('btc', 0.0272), ('analysis', 0.0239), ('use', 0.0223), ('result', 0.0210), ('bitcoin', 0.0206), ('factor', 0.0195), ('datum', 0.0183)
5	The digital signature algorithm in cryptocurrency wallets	('signature', 0.1360), ('wallet', 0.1180), ('scheme', 0.1046), ('key', 0.1001), ('protocol', 0.0524), ('user', 0.0494), ('security', 0.0462), ('ecdsa', 0.0433), ('propose', 0.0391), ('attack', 0.0317)
6	The need of regulations in token ICO	('ico', 0.1540), ('token', 0.1497), ('icos', 0.0952), ('offering', 0.0593), ('capital', 0.0436), ('market', 0.0375), ('coin', 0.0368), ('scam', 0.0357), ('regulation', 0.0328), ('raise', 0.0318)
7	Hacking attacks on cryptocurrency	('ransomware', 0.2529), ('ransom', 0.1012), ('malware', 0.0726), ('attack', 0.0705), ('file', 0.0686), ('victim', 0.0662), ('detection', 0.0476), ('crypto', 0.0460), ('payment', 0.0361), ('computer', 0.0341)
8	Resolving transaction processing speed and scalability problems through blockchain sharding	('shard', 0.2474), ('sharding', 0.1796), ('blockchain', 0.0695), ('network', 0.0564), ('transaction', 0.0552), ('system', 0.0523), ('performance', 0.0475), ('validator', 0.0466), ('scalability', 0.0442), ('base', 0.0396)

for cryptocurrency exchanges. Consequently, social media can be observed to largely highlight information that investors find significant and relevant.

E. DYNAMIC TOPIC MODEL (DTM)

A dynamic topic model is a generative model that enables the analysis of topic evolution within a collection of documents over time [79]. Dynamic topic modeling generates topics over time, which can then be visualized using the Plotly Python library to track their frequency and evolution.

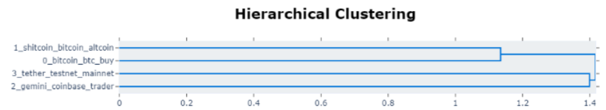

FIGURE 7. Hierarchical clustering from Reddit data.

TABLE 4. Topic representations of bitcoin from REDDIT.

Topic number	Main theme	Topic representations (c-TF-IDF score)
0	Cryptocurrency investment recommendations	('bitcoin', 0.0553), ('btc', 0.0519), ('buy', 0.0298), ('money', 0.0270), ('block', 0.0246), ('time', 0.0245), ('go', 0.0232), ('dip', 0.0226), ('transaction', 0.0225), ('want', 0.0223)
1	Negative view of the scam coin projects	('shitcoin', 1.099), ('bitcoin', 0.7072), ('altcoin', 0.2614), ('mean', 0.0699), ('meatcoin', 0.0531), ('fuck', 0.0530), ('trashcoin', 0.0528), ('bullshit', 0.0527), ('bitcooooin', 0.0519), ('bitcoinfixesthis', 0.0517)
2	Cryptocurrency exchange recommendations	('gemini', 1.2336), ('coinbase', 0.2068), ('trader', 0.0957), ('use', 0.0833), ('gdax', 0.0795), ('fee', 0.0792), ('blockfi', 0.0740), ('exchange', 0.0707), ('horoscope', 0.0624), ('active', 0.0623)
3	Cryptocurrency testnet and mainnet	('tether', 0.7977), ('testnet', 0.5034), ('mainnet', 0.3266), ('test', 0.1969), ('connect', 0.1489), ('go', 0.1391), ('network', 0.1363), ('ln', 0.1228), ('change', 0.1036), ('server', 0.1025)

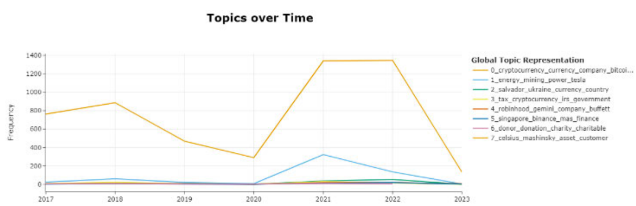

FIGURE 8. Temporal visualization of topics from LexisNexis data.

The changes in topics over time within LexisNexis data are depicted in Fig. 8. Topic 0, which focuses on Bitcoin as a digital currency, is consistently associated with the highest percentage every year, indicating sustained interest in Bitcoin itself. Furthermore, by examining the movement of Topic 0, it becomes apparent that the level of interest in Bitcoin correlates with its price movements [80]. In detail, interest in cryptocurrencies can be observed to rise during periods of explosive price growth, such as late 2017 to early 2018 and 2021, and falls during periods of price decline. Similarly, Topic 1, which pertains to Bitcoin mining, exhibited a sharp increase in 2021, coinciding with the peak of Bitcoin prices.

The evolution of topics over time in Web of Science data is illustrated in Fig. 9. Unlike LexisNexis data, which was concentrated on a single topic, the data appear to be

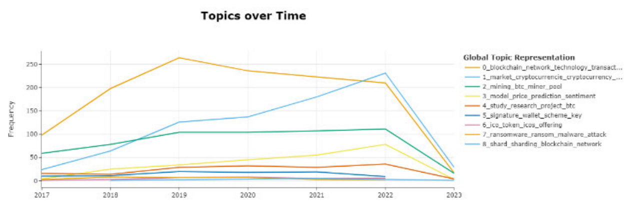

FIGURE 9. Temporal visualization of topics from web of science data.

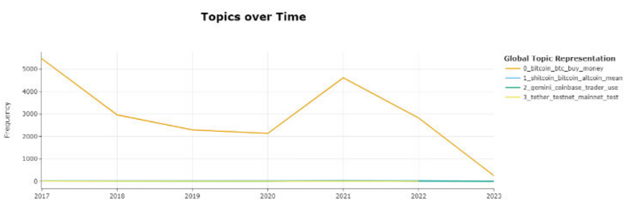

FIGURE 10. Temporal visualization of topics from Reddit data.

relatively evenly distributed across multiple topics. Moreover, the volume of data appears to fluctuate independently of Bitcoin price movements. Although many studies related to blockchain technology have been conducted prior to 2021, the number of papers related to the cryptocurrency market surpassed that of blockchain technology in 2022, which can be interpreted as a result of the growing market. Research on Bitcoin mining has continued with a similar volume of studies. The next most studied topics were those related to topic 3 (predicting Bitcoin prices using sentiment analysis) and topic 4 (factors affecting Bitcoin prices), which can be linked to the previously mentioned growth of the cryptocurrency market.

Fig. 10 depicts the changes in topics over time within Reddit data. Because the total volume of data was over 9 million, the frequency was cut into units of 1,000 for visibility. According to [81], the use of social media is influenced to a significant extent by the connections and interactions between users within social networks. Consequently, this social influence has a positive effect on the level of participation among users. In this context, the majority of Reddit posts on Bitcoin relate to purchase and investment recommendations. The total volume of posts, which can be seen to correlate with market interest, covariates more with Bitcoin prices than in the case of LexisNexis data.

V. CONCLUSION

Since Bitcoin was first created by Satoshi in 2008, the cryptocurrency market has grown rapidly, with interest in it continuing to increase [2], [82]. Cryptocurrency is now recognized as a type of asset, and related laws and finance have emerged [54], [55], [56]. Therefore, it is essential to perform data mining and attain knowledge pertaining to cryptocurrencies. In [83], topic modeling was performed using online Bitcoin forum data. In [84], topic modeling was performed using Twitter data and LDA to investigate the concerns and sentiment analyses of international users. However, these studies provided limited insights due to their use of data from single source.

To overcome the limitations of existing studies, the present study collected unstructured text data from LexisNexis, Web of Science, and Reddit to represent media, academia, and the general public, respectively. Additionally, an analysis of main themes was conducted using BERTopic, a SOTA model for topic modeling, to obtain knowledge. Finally, DTM was applied to study discourse on Bitcoin over time on each platform, thereby gaining insights and identifying differences.

The present study offers several possible applications. The findings suggest that the use of three distinct platforms, LexisNexis, Web of Science, and Reddit, with Bitcoin queries can provide a comprehensive understanding of sentiments in cryptocurrency. Unique characteristics of discourse were identified for each platform. Specifically, the news mainly covered major issues related to cryptocurrency, academic journals focused on practical improvements in line with technological developments, and social media discussions primarily revolved around investor psychology and the communication of information from an investor's perspective.

However, this study has several limitations that must be solved in future research. This study collected and processed solely English-language data. Because each country has different characteristics and recognition on cryptocurrencies [84], more generalized results could be obtained by conducting research that reflects this diversity. Additionally, this study collected and analyzed data from March 1, 2017, to March 1, 2023, whereas cryptocurrencies have existed for a longer duration. Therefore, future studies may cover longer timespans to grant a more comprehensive understanding of cryptocurrency.

ACKNOWLEDGMENT

The authors appreciate Editage (www.editage.co.kr) for their English editing service.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Bus. Rev.*, p. 21260, Nov. 2008.

[2] D.-E. Diaconaşu, S. Mehdian, and O. Stoica, "An analysis of investors' behavior in Bitcoin market," *PLoS ONE*, vol. 17, no. 3, Mar. 2022, Art. no. e0264522, doi: 10.1371/journal.pone.0264522.

[3] D. Vujičić, D. Jagodić, and S. Randić, "Blockchain technology, Bitcoin, and Ethereum: A brief overview," in *Proc. 17th Int. Symp. INFOTEH-JAHORINA (INFOTEH)*, Sarajevo, Bosnia Herzegovina, Mar. 2018, pp. 1–6.

[4] P. Ciaian, M. Rajcaniova, and D. Kancs, "Virtual relationships: Short- and long-run evidence from Bitcoin and altcoin markets," *J. Int. Financial Markets, Inst. Money*, vol. 52, pp. 173–195, Jan. 2018, doi: 10.1016/j.intfin.2017.11.001.

[5] A. S. Kumar and T. Ajaz, "Co-movement in crypto-currency markets: Evidences from wavelet analysis," *Financial Innov.*, vol. 5, no. 1, pp. 1–17, Jul. 2019, doi: 10.1186/s40854-019-0143-3.

[6] B. V. Barde and A. M. Bainwad, "An overview of topic modeling methods and tools," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Madurai, India, Jun. 2017, pp. 745–750.

[7] L. Hong and B. D. Davison, "Empirical study of topic modeling in Twitter," in *Proc. 1st Workshop Social Media Anal.*, Washington, DC, USA, Jul. 2010, pp. 80–88.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[9] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022, *arXiv:2203.05794*.

[10] M. Ortu, S. Vacca, G. Destefanis, and C. Conversano, "Cryptocurrency ecosystems and social media environments: An empirical analysis through Hawkes' models and natural language processing," *Mach. Learn. Appl.*, vol. 7, Mar. 2022, Art. no. 100229, doi: 10.1016/j.mlwa.2021.100229.

[11] B. Mcmillan, J. Myers, A. Nguyen, D. Robinson, and M. Kennard, "Analysis and comparison of natural language processing algorithms as applied to Bitcoin conversations on social media," *J. Investing*, vol. 31, no. 2, pp. 38–59, Jan. 2022, doi: 10.3905/joi.2021.1.213.

[12] J. V. Critien, A. Gatt, and J. Ellul, "Bitcoin price change and trend prediction through Twitter sentiment and data volume," *Financial Innov.*, vol. 8, no. 1, p. 45, May 2022, doi: 10.1186/s40854-022-00352-7.

[13] O. Sattarov, H. S. Jeon, R. Oh, and J. D. Lee, "Forecasting Bitcoin price fluctuation by Twitter sentiment analysis," in *Proc. Int. Conf. Inf. Sci. Commun. Technol. (ICISCT)*, Karachi, Pakistan, Nov. 2020, pp. 1–4.

[14] H. S. Jung, S. H. Lee, H. Lee, and J. H. Kim, "Predicting Bitcoin trends through machine learning using sentiment analysis with technical indicators," *Comput. Syst. Sci. Eng.*, vol. 46, no. 2, pp. 2231–2246, 2023, doi: 10.32604/csse.2023.034466.

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

[16] B. Yin and C.-H. Yuan, "Detecting latent topics and trends in blended learning using LDA topic modeling," *Educ. Inf. Technol.*, vol. 27, no. 9, pp. 12689–12712, Nov. 2022, doi: 10.1007/s10639-022-11118-0.

[17] E. Polyzos and F. Wang, "Twitter and market efficiency in energy markets: Evidence using LDA clustered topic extraction," *Energy Econ.*, vol. 114, Oct. 2022, Art. no. 106264, doi: 10.1016/j.eneco.2022.106264.

[18] C. Sharma, S. Sharma, and Sakshi, "Latent Dirichlet allocation (LDA) based information modelling on BLOCKCHAIN technology: A review of trends and research patterns used in integration," *Multimedia Tools Appl.*, vol. 81, no. 25, pp. 36805–36831, Oct. 2022, doi: 10.1007/s11042-022-13500-z.

[19] S. Avasthi, R. Chauhan, and D. P. Acharjya, "Topic modeling techniques for text mining over a large-scale scientific and biomedical text corpus," *Int. J. Ambient Comput. Intell.*, vol. 13, no. 1, pp. 1–18, Jan. 2022, doi: 10.4018/ijaci.293137.

[20] R. Egger and J. Yu, "A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts," *Frontiers Sociol.*, vol. 7, May 2022, Art. no. 886498, doi: 10.3389/fsoc.2022.886498.

[21] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese bert-networks," 2019. *arXiv:1908.10084*.

[22] I. Assent, "Clustering high dimensional data," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 2, no. 4, pp. 340–350, Jun. 2012, doi: 10.1002/widm.1062.

[23] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

[24] L. McInnes, J. Healy, and S. Astels, "HDBSCAN: Hierarchical density based clustering," *J. Open Source Softw.*, vol. 2, no. 11, p. 205, Mar. 2017, doi: 10.21105/joss.00205.

[25] M. Honnibal and I. Montani, "SpaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," *To Appear*, vol. 7, no. 1, pp. 411–420, 2017.

[26] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proc. 8th ACM Int. Conf. Web Search Data Mining*, New York, NY, USA, Feb. 2015, pp. 399–408.

[27] C. Meaney, M. Escobar, T. A. Stukel, P. C. Austin, and L. Jaakkimainen, "Comparison of methods for estimating temporal topic models from primary care clinical text data: Retrospective closed cohort study," *JMIR Med. Informat.*, vol. 10, no. 12, Dec. 2022, Art. no. e40102.

[28] R. Grinberg, "Bitcoin: An innovative alternative digital currency," *Hastings Sci. Technol. Law J.*, vol. 4, p. 160, Dec. 2011.

[29] G. P. Dwyer, "The economics of Bitcoin and similar private digital currencies," *J. Financial Stability*, vol. 17, pp. 81–91, Apr. 2015, doi: 10.1016/j.jfs.2014.11.006.

[30] J. Li, N. Li, J. Peng, H. Cui, and Z. Wu, "Energy consumption of cryptocurrency mining: A study of electricity consumption in mining cryptocurrencies," *Energy*, vol. 168, pp. 160–168, Feb. 2019, doi: 10.1016/j.energy.2018.11.046.

[31] K. J. O'Dwyer and D. Malone, "Bitcoin mining and its energy footprint," in *Proc. 25th IET Irish Signals Syst. Conf. China-Ireland Int. Conf. Inf. Commun. Technol. (ISSC/CIICT)*, Limerick, Ireland, Jun. 2014, pp. 280–285.

[32] E. Livni and O. Lopez, "El Salvador's adoption of Bitcoin is off to a rocky start," The New York Times, Sep. 7, 2021. Accessed: Aug. 4, 2023. [Online]. Available: https://www.nytimes.com/2021/09/07/business/el-salvador-Bitcoin.html?searchResultPosition=2

[33] E. Barrett, "Ukraine already trades more crypto than fiat currency. Now assets like Bitcoin are officially legal," Fortune, Feb. 18, 2020. Accessed: Aug. 4, 2023. [Online]. Available: https://fortune.com/2022/02/18/ukraine-legalizes-cryptocurrency-Bitcoin-russia-digital-assets/

[34] Triple A, Singapore. (2021). *Cryptocurrency Information About Venezuela*. Accessed: Aug. 4, 2023. [Online]. Available: https://triple-a.io/crypto-ownership-venezuela-2021/

[35] L. Davison and C. Condon, "Treasury calls for crypto transfers over $10,000 to be reported to IRS," Bloomberg, May 20, 2021. Accessed: Aug. 4, 2023. [Online]. Available: https://www.bloomberg.com/news/articles/2021-05-20/treasury-calls-for-crypto-transfers-over-10-000-reported-to-irs#xj4y7vzkg

[36] D. A. Liedel, "The taxation of Bitcoin: How the IRS views cryptocurrencies," *Drake Law Rev.*, vol. 66, no. 1, pp. 107–146, 2018.

[37] K. Solodan, "Legal regulation of cryptocurrency taxation in European countries," *Eur. J. Law Public Admin.*, vol. 6, no. 1, pp. 64–74, Sep. 2019.

[38] M. Lerer, "The taxation of cryptocurrency: Virtual transactions bring real-life tax implications," *CPA J.*, vol. 89, no. 1, pp. 40–43, Jan. 2019.

[39] Y. Liu, Z. Yang, and Y. Benslimane, "Bitcoin data analysis using deep learning and statistical modeling," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Kuala Lumpur, Malaysia, Dec. 2022, pp. 127–131.

[40] C. Jones, "Bitcoin debate: Warren Buffett bear vs. Winklevoss twins bull," Forbes, Feb. 23, 2018. Accessed: Aug. 4, 2023. [Online]. Available: https://www.forbes.com/sites/chuckjones/2018/02/23/Bitcoin-debate-warren-buffett-bear-vs-winklevoss-twins-bull/?sh=48452fd51331

[41] G. Ahlstrand and E. Gkritsi, "Binance Singapore drops crypto license plans in city-state," CoinDesk, Dec. 13, 2021. Accessed: Aug. 4, 2023. [Online]. Available: https://www.coindesk.com/policy/2021/12/13/binance-singapore-drops-crypto-license-plans-in-city-state/

[42] J. Lim, "MAS orders crypto exchange platform Binance.com to stop services in Singapore," The Straits Times, Sep. 9, 2021. Accessed: Aug. 4, 2023. [Online]. Available: https://www.straitstimes.com/business/banking/binancecom-placed-on-mas-investor-alert-list

[43] A. Singh, R. Rajak, H. Mistry, and P. Raut, "Aid, charity and donation tracking system using blockchain," in *Proc. 4th Int. Conf. Trends Electron. Informat. (ICOEI)*, Tirunelveli, India, Jun. 2020, pp. 457–462.

[44] E. Shaheen, M. A. Hamed, W. Zaghloul, E. A. Mostafa, A. E. Sharkawy, A. Mahmoud, A. Labeb, M. O. A. Enany, and G. Attiya, "A track donation system using blockchain," in *Proc. Int. Conf. Electron. Eng. (ICEEM)*, Menouf, Egypt, Jul. 2021, pp. 1–7.

[45] B. Lindrea, "Ukraine netted $70M in crypto donations since start of Russia conflict," Cointelegraph, Feb. 27, 2023. Accessed: Aug. 4, 2023. [Online]. Available: https://cointelegraph.com/news/ukraine-netted-70m-in-crypto-donations-since-start-of-russia-conflict

[46] Forbes. (Sep. 20, 2022). *What Really Happened to LUNA Crypto?* Accessed: Aug. 4, 2023. [Online]. Available: https://www.forbes.com/sites/qai/2022/09/20/what-really-happened-to-luna-crypto/?sh=63540a624ff1

[47] S. Lee, J. Lee, and Y. Lee, "Dissecting the Terra-LUNA crash: Evidence from the spillover effect and information flow," *Finance Res. Lett.*, vol. 53, May 2023, Art. no. 103590, doi: 10.1016/j.frl.2022.103590.

[48] A. Briola, D. Vidal-Tomás, Y. Wang, and T. Aste, "Anatomy of a Stablecoin's failure: The Terra-Luna case," *Finance Res. Lett.*, vol. 51, Jan. 2023, Art. no. 103358, doi: 10.1016/j.frl.2022.103358.

[49] M. H. Joo, Y. Nishikawa, and K. Dandapani, "Cryptocurrency, a successful application of blockchain technology," *Managerial Finance*, vol. 46, no. 6, pp. 715–733, Aug. 2019, doi: 10.1108/mf-09-2018-0451.

[50] R. B. Fekih and M. Lahami, "Application of blockchain technology in healthcare: A comprehensive study," in *Proc. 18th Int. Conf. Smart Homes Health Telematics (ICOST)*, Hammamet, Tunisia, 2020, pp. 268–276.

[51] M. H. Miraz and M. Ali, "Applications of blockchain technology beyond cryptocurrency," 2018, *arXiv:1801.03528*.

[52] P. Tasatanattakool and C. Techapanupreeda, "Blockchain: Challenges and applications," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Chiang Mai, Thailand, Jan. 2018, pp. 473–475.

[53] D. G. Baur, K. Hong, and A. D. Lee, "Bitcoin: Medium of exchange or speculative assets?" *J. Int. Financial Markets, Inst. Money*, vol. 54, pp. 177–189, May 2018, doi: 10.1016/j.intfin.2017.12.004.

[54] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, and L. I. Hagfors, "On the Hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?" *Finance Res. Lett.*, vol. 20, pp. 192–198, Feb. 2017, doi: 10.1016/j.frl.2016.09.025.

[55] A. Kliber, P. Marszałek, I. Musiałkowska, and K. Świerczyńska, "Bitcoin: Safe haven, Hedge or diversifier? Perception of Bitcoin in the context of a country's economic situation—A stochastic volatility approach," *Phys. A, Stat. Mech. Appl.*, vol. 524, pp. 246–257, Jun. 2019, doi: 10.1016/j.physa.2019.04.145.

[56] S. Corbet, A. Meegan, C. Larkin, B. Lucey, and L. Yarovaya, "Exploring the dynamic relationships between cryptocurrencies and other financial assets," *Econ. Lett.*, vol. 165, pp. 28–34, Apr. 2018, doi: 10.1016/j.econlet.2018.01.004.

[57] R. Zhang and W. K. V. Chan, "Evaluation of energy consumption in block-chains with proof of work and proof of stake," *J. Phys., Conf. Ser.*, vol. 1584, no. 1, Jul. 2020, Art. no. 012023, doi: 10.1088/1742-6596/1584/1/012023.

[58] N. Lasla, L. Al-Sahan, M. Abdallah, and M. Younis, "Green-PoW: An energy-efficient blockchain proof-of-work consensus algorithm," *Comput. Netw.*, vol. 214, Sep. 2022, Art. no. 109118, doi: 10.1016/j.comnet.2022.109118.

[59] T. Pano and R. Kashef, "A complete VADER-based sentiment analysis of Bitcoin (BTC) tweets during the era of COVID-19," *Big Data Cogn. Comput.*, vol. 4, no. 4, p. 33, Nov. 2020, doi: 10.3390/bdcc4040033.

[60] R. H. D. Neves, "Bitcoin pricing: Impact of attractiveness variables," *Financial Innov.*, vol. 6, no. 1, pp. 1–18, Apr. 2020, doi: 10.1186/s40854-020-00176-3.

[61] T. Panagiotidis, T. Stengos, and O. Vravosinos, "The effects of markets, uncertainty and search intensity on Bitcoin returns," *Int. Rev. Financial Anal.*, vol. 63, pp. 220–242, May 2019, doi: 10.1016/j.irfa.2018.11.002.

[62] J. Korn, "Record $3.8 billion stolen in crypto hacks last year, report says," CNN, Feb. 1, 2022. Accessed: Aug. 4, 2023. [Online]. Available: https://edition.cnn.com/2023/02/01/tech/crypto-hacks-2022/index.html

[63] D. Nelson and N. De, "FTX U.S. temporarily froze crypto withdrawals, adding to chaos of bankruptcy proceedings," Coindesk, Nov. 12, 2022. Accessed: Aug. 4, 2023. [Online]. Available: https://www.coindesk.com/business/2022/11/11/ftx-us-freezes-crypto-withdrawals-sending-millions-in-assets-to-bankruptcy-limbo/

[64] D. Johnson, A. Menezes, and S. Vanstone, "The elliptic curve digital signature algorithm (ECDSA)," *Int. J. Inf. Secur.*, vol. 1, no. 1, pp. 36–63, Aug. 2001, doi: 10.1007/s102070100002.

[65] R. Gennaro, S. Goldfeder, and A. Narayanan, "Threshold-optimal DSA/ECDSA signatures and an application to Bitcoin wallet security," in *Proc. 14th Int. Conf. Appl. Cryptogr. Netw. Secur. (ACNS)*, vol. 14, Guildford, U.K., 2016, pp. 156–174.

[66] A. Feign, "What is an ICO?" Coindesk, Dec. 12, 2022. [Online]. Available: https://www.coindesk.com/learn/what-is-an-ico/

[67] D. Boreiko and N. K. Sahdev. (2018). *To ICO or Not to ICO-Empirical Analysis of Initial Coin Offerings and Token Sales*. [Online]. Available: https://ssrn.com/abstract=3209180

[68] L. Rhue. (2018). *Trust is All You Need: An Empirical Exploration of Initial Coin Offerings (ICOs) and ICO Reputation Scores*. [Online]. Available: https://ssrn.com/abstract=3179723

[69] O. A. Karpenko, T. K. Blokhina, and L. V. Chebukhanova, "The initial coin offering (ICO) process: Regulation and risks," *J. Risk Financial Manage.*, vol. 14, no. 12, p. 599, Dec. 2021, doi: 10.3390/jrfm14120599.

[70] Y. Tsuchiya and N. Hiramoto, "How cryptocurrency is laundered: Case study of coincheck hacking incident," *Forensic Sci. Int., Rep.*, vol. 4, Nov. 2021, Art. no. 100241, doi: 10.1016/j.fsir.2021.100241.

[71] K. Grobys, "When the blockchain does not block: On hackings and uncertainty in the cryptocurrency market," *Quant. Finance*, vol. 21, no. 8, pp. 1267–1279, Aug. 2021, doi: 10.1080/14697688.2020.1849779.

[72] U. W. Chohan. (2018). *The Problems of Cryptocurrency Thefts and Exchange Shutdowns*. [Online]. Available: https://ssrn.com/abstract=3131702

[73] L. Wang, "The challenge and prospect of scalability of blockchain technology," in *Proc. 5th Int. Conf. Comput. Sci. Artif. Intell.*, Beijing, China, Dec. 2021, pp. 296–301.

[74] M. Zamani, M. Movahedi, and M. Raykova, "RapidChain: Scaling blockchain via full sharding," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Toronto, ON, Canada, Oct. 2018, pp. 931–948.

[75] H. Dang, T. T. A. Dinh, D. Loghin, E.-C. Chang, Q. Lin, and B. C. Ooi, "Towards scaling blockchain systems via sharding," in *Proc. Int. Conf. Manage. Data*, Amsterdam, The Netherlands, Jun. 2019, pp. 123–140.

[76] S. Bonini, T. Shohfi and M. Simaan. (2022). *Buy the Dip?* [Online]. Available: https://ssrn.com/abstract=3835376

[77] N. Sherman and J. Tidy, "Crypto giant FTX collapses into bankruptcy," BBC, Nov. 11, 2022. Accessed: Aug. 4, 2023. [Online]. Available: https://www.bbc.com/news/business-63601213

[78] Immunebytes, New Delhi, India. (2022). *Mainnet vs Testnet in Blockchain*. Accessed: Aug. 4, 2023. [Online]. Available: https://www.immunebytes.com/blog/mainnet-vs-testnet-in-blockchain/

[79] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proc. ICML*, Pittsburgh, PA, USA, 2006, pp. 113–120.

[80] A.-D. Vo, "Sentiment analysis of news for effective cryptocurrency price prediction," *Int. J. Knowl. Eng.*, vol. 5, no. 2, pp. 47–52, Dec. 2019, doi: 10.18178/ijke.2019.5.2.116.

[81] S. Boulianne, "Social media use and participation: A meta-analysis of current research," *Inf., Commun. Soc.*, vol. 18, no. 5, pp. 524–538, May 2015, doi: 10.1080/1369118x.2015.1008542.

[82] S. Bourgi, "Institutional investors increase their crypto holdings for 5th straight week," Cointelegraph, Sep. 20, 2021. Accessed: Aug. 4, 2023. [Online]. Available: https://cointelegraph.com/news/institutional-investors-increase-their-crypto-holdings-for-5th-straight-week

[83] Y. B. Kim, J. Lee, N. Park, J. Choo, J.-H. Kim, and C. H. Kim, "When Bitcoin encounters information in an online forum: Using text mining to analyse user opinions and predict value fluctuation," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177630, doi: 10.1371/journal.pone.0177630.

[84] S. Bibi, S. Hussain, and M. I. Faisal, "Public perception based recommendation system for cryptocurrency," in *Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Islamabad, Pakistan, Jan. 2019, pp. 661–665.

HAE SUN JUNG is currently pursuing the Ph.D. degree with the Department of Applied Artificial Intelligence, Sungkyunkwan University. His research interests include natural language processing, computer vision, deep learning, and machine learning.

HAEIN LEE is currently pursuing the Ph.D. degree with the Department of Applied Artificial Intelligence and the Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. Her research interests include natural language processing, deep learning, and machine learning.

JANG HYUN KIM is currently a Professor with the Department of Human-Artificial Intelligence Interaction, the Department of Interaction Science, and the Department of Applied Artificial Intelligence, Sungkyunkwan University. He has authored more than 50 articles in major journals, such as *Information Processing and Management*, *Telematics and Informatics*, *Cities*, *Government Information Quarterly*, *Technological Forecasting and Social Change*, and *Journal of Computer-Mediated Communication*. His research interests include social/semantic data analysis, social media, and future media.