

Cluster Visualized Topic Modeling Paradigms for Recognition of Health-Related Topics Through a Machine Learning



Yerragudipadu Subbarayudu^{1*}, Alladi Sureshbabu²

¹ Computer Science and Engineering, Jawaharlal Nehru Technological University, Anantapur 515002, India

² Computer Science and Engineering, JNTUA College of Engineering, Anantapur 515002, India

Corresponding Author Email: subbu.jntua@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.290320>

ABSTRACT

Received: 20 July 2023

Revised: 17 November 2023

Accepted: 12 March 2024

Available online: 20 June 2024

Keywords:

blockchain technology, machine learning, ledger, consensus, topic modeling, healthcare, decentralized platform, and sentiment analysis

The world can manage its path towards better health thanks to the information, community, and support that medical forums offer in the modern digital environment. Integrating subject modelling on a decentralized platform may be essential and innovative along the way. Topic modelling aids in better understanding user requirements, spotting patterns and trends in the medical sector, and taking proactive measures. A centralized platform is typically used to host health forums, but this has several disadvantages, including a lack of security and privacy for sensitive personal health information, the potential for bias and censorship to serve the vested interests of the central authority, and it is significantly more expensive to implement and maintain than a decentralized platform. We therefore suggest a medical forum with topic modelling housed on a decentralized platform to enhance the existing state of medical forums so that we can better understand the current topics of interest in the medical sector and act proactively. Topic modelling analysis speeds up reaction time and aids in better understanding community needs. Blockchain technology offers enhanced privacy and security for healthcare data. However, there are still challenges in ensuring the privacy of sensitive information when conducting topic modeling on blockchain-based healthcare systems. Further research is needed to address privacy concerns, develop privacy-preserving topic modeling algorithms, and establish robust data access control mechanisms. Social media platforms generate a massive amount of healthcare-related content, including posts, comments, and discussions. Without topic modeling, sorting through this overwhelming volume of data becomes a significant challenge. It can lead to information overload, making it hard to identify key trends, topics, or critical issues. The absence of topic modeling in the analysis of healthcare topics on social media results in a lack of structure, organization, and systematic exploration of the information available. Topic modeling provides a valuable solution by automatically identifying, categorizing, and analyzing the diverse range of healthcare-related discussions, enabling more insightful and efficient understanding of the landscape. Current topic modeling approaches often assume static topics and may not capture temporal dynamics and emerging topics in real-time. Research is needed to develop dynamic topic modeling techniques i.e. Cluster Visualized BTM and Cluster Visualized Hierarchical Dirichlet Process that can adapt to evolving healthcare topics and provide timely insights for decision-making in blockchain-based healthcare systems. The forum's host also offers several benefits like privacy, security, affordability, and less bias and restriction. The submitted information is not utilized by a central authority with personal interests.

1. INTRODUCTION

The healthcare industry is significant for both developed and developing countries. The introduction of cutting-edge computer technologies inside the medical and healthcare sectors has significantly enhanced the sector's overall capabilities. The prompt diagnosis and treatment of a variety of health-related issues may be made possible by these developments in computer technology for physicians and other relevant health providers [1]. The gap in the existing literature on healthcare topics in social media lies in the limited adoption of advanced topic modeling techniques to

systematically analyze and interpret the diverse and dynamic nature of health-related discussions. By incorporating topic modeling, researchers can bridge this gap, enabling a more nuanced, structured, and insightful exploration of healthcare topics in the ever-evolving landscape of social media.

Medical forums are online discussion groups where healthcare professionals, students, patients, and anybody with an interest in medicine can interact, exchange knowledge, and seek assistance. These discussion boards act as online communities where users may share knowledge, ask questions, and offer support about a range of medical topics. Fostering cooperation and education among medical experts is one of the

main goals of medical forums. Doctors, nurses, and other healthcare professionals can connect with peers from around the globe to discuss difficult situations, exchange clinical wisdom, and get professional advice. Professionals can broaden their knowledge, improve their ability to diagnose and treat patients, and keep up with the most recent developments in their professions thanks to our collaboration. Medical forums also allow patients and carers a place to share their experiences, ask questions, and get support from other people dealing with the same health issues. Patients can talk with peers who may have direct experience with comparable disorders about their symptoms, treatment choices, and self-care techniques. Peer-to-peer assistance may offer people navigating their health journeys emotional support, useful advice, and a feeling of community. Medical forums can offer helpful advice and support, but it's vital to be aware that there may also be downsides. These include the potential for inaccurate information, privacy issues, and the requirement for appropriate moderation to guarantee civil and fact-based conversations.

The procedures already in place in medical forums give individuals and healthcare professionals a useful arena for conversation, advice-seeking, and knowledge-sharing. To meet the requirements of the medical community, these systems provide a variety of features and functionalities. Reddit, a well-known social media site with several medically related subreddits, is one notable example. Users may take part in conversations, post questions, and exchange stories about a range of medical issues. Twitter is a social media site where users can share quick updates that can include public health-related information. Platforms like Twitter hold promise for wider adoption in public health applications since they are real-time and can be mined as such [2, 3].

Blockchain maintains a decentralized ledger that is append-only, meaning that once data is added, it cannot be altered or deleted. This immutability ensures that medical records and forum discussions stored on the blockchain remain tamper-proof, enhancing the integrity of the information. Blockchain employs advanced cryptographic techniques to secure transactions and data. Each participant in the network has a unique cryptographic key, ensuring that only authorized individuals can access and contribute to the information within the medical forum.

2. RELATED WORKS

Contemporary research has been done on the topic and the resources used and the major findings are provided. With the rise in the importance of secure and robust medical forums, the importance of blockchain and machine learning algorithms has skyrocketed. Rapid response through improved surveillance is important to combat emerging infectious diseases [4]. This calls for the use of blockchain integrated with machine learning algorithms such as sentiment analysis and topic modelling. Topic modeling is one of the popular techniques for information retrieval based on themes from biomedical documents. Topic modeling techniques are utilized for the summarization of a large collection of text documents. Probabilistic topic modeling techniques are used to identify the core topics from the biomedical text collection of documents [5, 6].

Sentiment analysis according to the study by Wang et al. [7], entails ascertaining the sentiment or opinion represented in

text, which is essential for a variety of applications, including social media analysis and customer feedback analysis. This blockchain technique tries to improve the transparency and dependability of sentiment analysis findings. Blockchain technology integration guarantees data security and integrity. The technique solves issues with data tampering and sentiment analysis's reliability by utilizing blockchain. By utilizing the transparency and immutability of blockchain, it seeks to increase the accuracy and dependability of sentiment analysis findings. Thus, to grant access to the facility, we employ the most widely used technique, which is to make use of the idea of online forums. This takes us to the process of effectively gathering and using data to better people's lives.

The use of online forums and communities for health research is a new strategy that makes effective use of online platforms to obtain important information and facts about health-related issues [8-10]. People can use these forums to discuss their experiences, worries, and knowledge of various medical diseases and treatments. Researchers may reach a wide range of different participants by using online discussion forums and communities, enabling broader and more inclusive research investigations. These platforms provide a special setting for people to openly express their ideas and emotions, which may result in more accurate and insightful data. In the process of better understanding the idea of using online tweets along with topic modelling and sentiment analysis the research on the same was conducted. The research was on the effects of online education which can potentially be extended to understand the retrieval of topic modelling and sentiment analysis trends of a medical forum.

Because of the epidemic, Shahi et al. [11] suggested that the research concentrate on tweet analysis to acquire insights into people's thoughts, feelings, and conversations around the shift to online schooling. To glean useful data from the Tweets, the researchers use topic modelling and sentiment analysis tools. Progressively, machine intelligence will mimic human intelligence in complex, independent ways. Most of the time, early medical artificial intelligence systems relied on experts to educate computers clinical knowledge by encoding it as logic rules for certain clinical scenarios. More capable machine learning algorithms train themselves to recognize and weight important data components, including pixels from medical images or raw data from electronic health records (EHRs), to understand these principles [12-14].

On the other hand, topic modelling seeks to identify the underlying themes or subjects that are present in the tweets. Researchers acquire insights into the primary areas of interest and concerns by spotting reoccurring patterns and prevalent topics. The study makes use of the enormous quantity of tweets that are publicly accessible to get a thorough understanding of the attitude and themes surrounding online schooling. The inclusion of a decentralized medical forum has several benefits for online debates on healthcare. The use of online discussion forums and communities for health research, as well as the literature survey conducted in "A Sentiment Analysis Method Based on a Blockchain-Supported Long Short-Term Memory Deep Network" and "The Use of Online Discussion Forums and Communities for Health Research," support this approach and offer a deeper understanding of patient experiences, sentiments, and concerns regarding medical conditions [15].

Healthcare professionals and academics may improve patient care, develop targeted therapies, and increase patient satisfaction by utilizing sentiment analysis and topic

modelling tools. Sentiment analysis aids in assessing patient sentiments, addressing concerns, and fostering meaningful dialogue, promoting patient engagement and trust [16-19]. Topic modelling facilitates the identification of emerging trends and topics within the medical forum, allowing healthcare professionals to address emerging patient needs. Additionally, real-time analysis of patient feelings and experiences enables healthcare personnel to see pressing problems, unpleasant encounters, or unmet requirements, allowing them to move quickly to address these worries and perhaps enhance patient outcomes and satisfaction.

Additionally, a decentralized medical forum powered by blockchain technology provides data protection and privacy. Benefiting from blockchain's immutability and cryptographic procedures, which ensure the security and integrity of sensitive information, patients maintain ownership over their data. Patients are encouraged to openly discuss their experiences and concerns on the medical forum as a result, which encourages patient confidence and leads to more thorough research results and well-informed healthcare decisions [20]. Online healthcare conversations can benefit from several factors when topic modelling and sentiment analysis are combined with a decentralized medical forum. It offers a better comprehension of patient experiences, makes it easier to spot new patterns, allows for prompt feedback and intervention, and assures data security and privacy [21-23]. These advantages support better patient care, evidence-based treatments, and patient involvement, which eventually advance medical study.

2.1 Research gaps

While topic modeling in the context of healthcare and blockchain technology holds great potential, there are several research gaps that warrant further investigation. Some of the key research gaps include.

a) Scalability and Efficiency: Topic modeling algorithms, particularly those based on probabilistic models, can be computationally intensive, making them less efficient when applied to large-scale healthcare datasets. Research is needed to develop scalable and efficient topic modeling techniques that can handle the increasing volume and complexity of healthcare data stored on blockchain networks.

b) Privacy and Security Considerations: Blockchain technology offers enhanced privacy and security for healthcare data. However, there are still challenges in ensuring the privacy of sensitive information when conducting topic modeling on blockchain-based healthcare systems. Further research is needed to address privacy concerns, develop privacy-preserving topic modeling algorithms, and establish robust data access control mechanisms.

c) Interoperability and Data Integration: Healthcare data is typically fragmented and stored across various systems and organizations. Integrating and harmonizing data from different sources for topic modeling on blockchain networks presents significant challenges. Research is needed to develop effective methods and standards for data interoperability and integration to enable seamless topic modeling across disparate healthcare data sources.

d) Topic Interpretability and Domain-Specific Models: Topic modeling algorithms generate topics based on statistical patterns in the data. However, healthcare-specific nuances and domain-specific terminology may not be adequately captured by generic topic models. Further research is needed to develop

domain-specific topic models that can capture the intricacies of healthcare topics and provide interpretable and actionable insights.

e) Real-Time Analysis and Dynamic Topics: Healthcare data is dynamic and continuously evolving. Current topic modeling approaches often assume static topics and may not capture temporal dynamics and emerging topics in real-time. Research is needed to develop dynamic topic modeling techniques that can adapt to evolving healthcare topics and provide timely insights for decision-making in blockchain-based healthcare systems.

f) User-Centric Topic Analysis: Topic modeling techniques should be user-centric, taking into account the specific needs and perspectives of healthcare professionals, researchers, and patients. Research is needed to develop interactive and customizable topic modeling interfaces that allow users to explore and analyze healthcare topics based on their specific interests, expertise, and objectives.

Blockchain enables patients to have greater control over their medical data. Patients can grant permission for specific healthcare providers or individuals to access their information, enhancing privacy and reducing the risk of unauthorized access. Blockchain technology can offer a more secure and trustworthy environment for medical forums, promoting data integrity, and safeguarding sensitive health information from unauthorized access or tampering.

The implications of blockchain research for healthcare professionals and patients include improved data security, streamlined collaboration, increased patient control, enhanced trust, and more efficient and transparent healthcare processes. The potential benefits span across various aspects of healthcare delivery and patient experience.

Addressing these research gaps will contribute to the advancement of topic modeling techniques in the context of healthcare and blockchain technology, enabling more effective analysis and understanding of healthcare topics and facilitating better decision-making and innovation in the healthcare industry.

3 PROPOSED ARCHITECTURES

Blockchain technology might be used to build a decentralized medical forum where doctors, researchers, and patients could interact and exchange knowledge without the need for a central authority or middleman. The forum would be safer, open, and accessible due to its decentralized structure. Users would have ownership over their own data and be able to safely share it with others on the site through a blockchain-based medical forum. Sharing scientific discoveries, technological advances in medicine, and best practices would be made easier as a result. The forum could also make it possible to share medical documents securely and privately and track patient outcomes.

The validity and integrity of the information published on the forum would be guaranteed using blockchain technology. Each piece of data would be immutable and time-stamped, making it impossible to alter or erase. This would make it simpler to check the accuracy of the material and stop the spread of fraud or false information. Additionally, a decentralized medical forum powered by blockchain technology may make financing for medical research more effective and open. Researchers might obtain money directly from donors and follow the development of their work on the

forum by utilizing blockchain-based smart contracts. Additionally, donors might follow the results of their donations and have a better understanding of how their money is being used. Overall, a decentralized medical forum using blockchain technology has the potential to transform the way medical knowledge is shared, research is conducted, and funding is allocated. It could lead to a more collaborative and transparent healthcare ecosystem, with better health outcomes for patients.

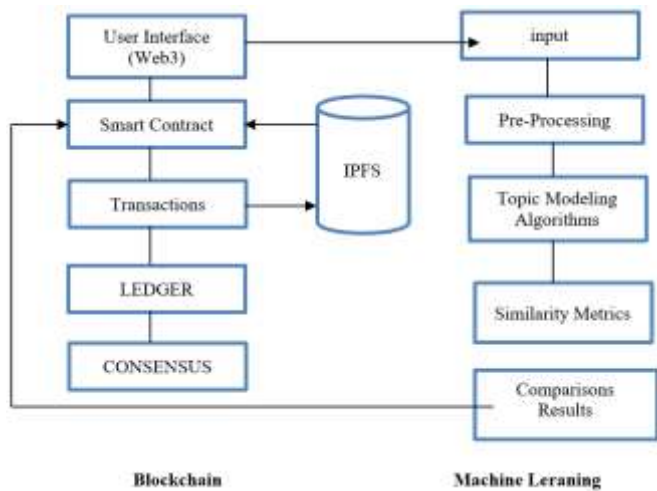


Figure 1. Control flow graph of proposed system with blockchain

In Figure 1, the user gets options to view post, create post, delete post, comment post. The user will also have an option to view profiles of different users, as it is hosted on blockchain the whole data is visible. Simply put, smart contracts are blockchain-based programs that are executed when certain criteria are satisfied. They are often used to automate the implementation of an agreement so that all parties may be confident of the conclusion right away, without the need for an intermediary or additional delay. They can also automate a process such that when circumstances are satisfied, the following action is executed. Simple "if/when...then" phrases that are typed into code and placed on a blockchain are how smart contracts operate. When predefined circumstances have been verified to have been satisfied, a network of computers will carry out the actions. These can entail paying out money to the right people, registering a car, sending out notices, or writing a ticket.

Smart contracts in healthcare-related social media provide a secure and automated framework for implementing privacy controls, executing data sharing agreements, and establishing reputation systems. These functionalities contribute to a more transparent, trustworthy, and patient-centric approach to healthcare discussions in the online social sphere.

When the transaction is finished, the blockchain is then updated. That means the transaction cannot be changed, and only parties who have been granted permission can see the results. The procedure that the smart contract does once a user makes a request is the most crucial. The smart contract adds the transaction to the ledger and provides consensus, another name for proof of security. The user receives a response from the smart contract after processing is complete. The smart contract is used to provide input for the machine learning model, and the user receives the output in return. Blockchain and Machine Learning gives the best solutions together in

performing various tasks in the Smart Health care system [24].

Blockchain technology uses decentralization and cryptographic techniques to ensure strong data protection. To prevent unauthorized access or manipulation, it makes sure that user data and talks are encrypted and stored over a dispersed network of computers. The blockchain gives individuals more control over their personal data. Thanks to the use of encryption and user-controlled access methods, users may choose what information they wish to share and with whom. As a result, people have enhanced protections for their privacy while sharing and discussing medical information. Because blockchain is decentralized, no central authority or middleman is required. The blockchain serves as a transparent and immutable ledger for all forum transactions and interactions. Because everyone can verify the accuracy of the debates and suggestions, this fosters greater participant confidence. The idea behind blockchain is that each block uses a cryptographic technique to maintain a reference to the one before it. Blockchain is kept on network devices (computers), known as nodes, each of which has a copy of the whole blockchain, as opposed to a central server like other internet services [25, 26].

In Figure 1, to explore how smart contracts, transactions, and ledgers can be applied to healthcare topics in social media. Smart Contracts: Smart contracts are self-executing agreements coded on a blockchain that automatically execute predefined conditions and actions.

In the context of healthcare topics in social media, smart contracts can facilitate various processes and interactions.

Privacy Controls: Smart contracts can enable users to define privacy preferences for their healthcare data shared on social media. For example, a smart contract can enforce restrictions on who can access and use the data, ensuring privacy and consent management.

Data Sharing Agreements: Smart contracts can facilitate secure data sharing agreements between healthcare organizations, researchers, and social media platforms. These contracts can specify terms and conditions for sharing, accessing, and utilizing healthcare data shared on social media platforms.

Reputation Systems: Smart contracts can establish reputation systems for social media users discussing healthcare topics. The contracts can track and evaluate the credibility and reliability of users' contributions based on various factors, such as accuracy, expertise, and trustworthiness.

Transactions:

In the context of healthcare topics in social media, transactions refer to the interactions and exchanges of value or information that occur on the platform. Blockchain technology enables secure and transparent transactions.

Data Exchange: Users can exchange healthcare-related information, such as research findings, medical insights, or patient experiences, through social media platforms. Blockchain can ensure the integrity and authenticity of these transactions, preventing tampering or unauthorized modifications.

Tokenized Incentives: Blockchain-based tokens or cryptocurrencies can be used as incentives within social media platforms to reward users who contribute valuable healthcare-related content or participate in discussions. These tokens can be earned, transferred, or redeemed for various benefits or services.

Micropayments: Blockchain technology enables seamless

micropayments for accessing premium healthcare content or services on social media platforms. Users can pay small amounts directly to content creators or service providers, facilitating fair compensation and monetization of valuable healthcare information.

Ledger:

A ledger is a distributed database that records and stores transactions in a chronological and transparent manner. In healthcare topics in social media, a ledger can provide a secure and tamper-proof record of relevant activities:

Data Integrity: The ledger can record the transactions related to healthcare discussions, research collaborations, or data sharing on social media platforms. This provides an immutable and transparent history of interactions, ensuring data integrity and accountability.

Auditability: Ledgers enable auditors, regulators, or researchers to verify and audit transactions and activities related to healthcare topics on social media platforms. This enhances transparency, compliance, and trust in the platform's operations.

Provenance Tracking: The ledger can track the origin and history of healthcare-related content shared on social media. This helps verify the authenticity of information and trace back the sources of knowledge or claims made in discussions.

By leveraging smart contracts, transactions, and ledgers, healthcare topics in social media can benefit from enhanced privacy controls, secure data sharing agreements, reliable reputation systems, transparent transactions, and trustworthy records. These blockchain-enabled features contribute to the integrity, efficiency, and trustworthiness of healthcare discussions and interactions on social media platforms.

In Figure 1, the Interplanetary File System (IPFS) is a decentralized and distributed file system that aims to create a global network of interconnected nodes where files are stored and accessed using content-based addressing. IPFS can be leveraged in the context of healthcare topics in social media to store and retrieve data related to discussions, articles, or documents while incorporating topic modeling algorithms. Here's an explanation of how IPFS works in this scenario.

a) Storing Data on IPFS: Content-Based Addressing: IPFS uses content-based addressing, where files are identified and addressed based on their content rather than their location. Each file is assigned a unique cryptographic hash, derived from its content, which serves as its identifier. Content Addressed Storage: When storing data on IPFS, the files are divided into blocks, and each block is identified by its hash. The blocks are distributed across the IPFS network, and their availability is ensured by replication among participating nodes. Distributed Network: IPFS forms a decentralized network of interconnected nodes, where each node contributes storage and bandwidth. This distribution prevents a single point of failure and provides resilience to the system.

b) Linking IPFS and Topic Modeling: IPFS as Storage: Utilize IPFS to store the preprocessed data, including the social media posts, articles, or discussions related to healthcare topics. Each file is stored and referenced using its unique content-based address (hash). Topic Modeling on IPFS Data: Implement topic modeling algorithms on the data stored in IPFS. Retrieve the necessary files from IPFS, preprocess them, and perform topic modeling to identify and analyze the healthcare-related topics within the social media data.

c) Retrieving Data from IPFS: Content-Based Retrieval: Retrieve the required data from IPFS by specifying the content-based address (hash) of the files. This ensures that the

exact version of the file is retrieved, regardless of its location in the IPFS network. Decentralized Access: Since IPFS is a decentralized network, the data can be accessed from any participating node in the network, eliminating the need for a central server. By utilizing IPFS for storing and accessing healthcare-related data, and incorporating topic modeling algorithms for analyzing the content, we can create a decentralized and distributed system for healthcare topics in social media. This approach allows for the secure and efficient storage, retrieval, and analysis of healthcare-related discussions, articles, or documents, while ensuring data integrity and availability across the IPFS network.

By organizing the content under these headers, we can provide a structured overview of how blockchain, smart contracts, transactions, ledgers, and integration with IPFS collectively contribute to enhancing data security, privacy controls, and reputation systems in healthcare discussions on social media.

Smart contracts are typically coded using specific programming languages. In the case of healthcare discussions on social media, platforms like Ethereum commonly use languages such as Solidity for smart contract development. Smart contracts can be coded to implement patient-controlled access. For example, a simple Solidity function might allow patients to manage access permissions by updating a list of authorized addresses that can interact with their healthcare data.

The choice of conditions depends on the specific use case and requirements of the healthcare forum on social media. Implementing these conditions ensures that smart contracts execute in a way that aligns with the privacy, data sharing, and reputation system goals set by the community and platform stakeholders.

Smart contracts and user identities can utilize public-key cryptography. Each participant has a public key for identification and a private key for secure access. Cryptographic hash functions secure data integrity. Blockchain platforms use hashes to link blocks, ensuring that any change in the data results in a different hash. Digital signatures verify the authenticity of transactions. Participants sign their transactions with their private keys, allowing others to confirm the legitimacy of the sender.

In healthcare scenarios, where trusted identities are crucial, a consensus mechanism like PoA (Proof-of-Authority) can be employed. Authority nodes, identified entities in the healthcare field, validate transactions. For permissioned networks, PBF (Practical Byzantine Fault Tolerance) can provide a consensus mechanism that ensures agreement among nodes in the network even in the presence of malicious actors. DPoS (Delegated Proof-of-Stake) allows stakeholders to vote for delegates who validate transactions. This can enhance scalability and efficiency while maintaining a level of decentralization.

In PoA or permissioned blockchains, the risk of a 51% attack is minimized as consensus is achieved through trusted entities. For public blockchains, continuous monitoring and rapid response to unusual activity can help mitigate the risk. Identity verification through public-key cryptography and the use of reputation systems can help prevent or minimize the impact of Sybil attacks. Encrypting data before storing it on IPFS and ensuring that only authorized parties have access to the decryption keys.

IPFS is preferred over traditional cloud storage and other decentralized storage systems due to its decentralized, content-

addressed, and privacy-preserving nature. Its use contributes directly to the goals of the proposed architecture, supporting data integrity, efficient content retrieval, privacy control, and reducing reliance on centralized infrastructure Figure 1.

4. METHODOLOGY

Blockchain technology has the potential to revolutionize the healthcare industry and can be related to healthcare topics in social media in several ways. **Data Integrity and Security:** Blockchain ensures the integrity and security of healthcare data shared on social media platforms. By leveraging the cryptographic properties of blockchain, healthcare information can be securely stored, encrypted, and shared, reducing the risk of data breaches and unauthorized access. **Patient Privacy and Consent:** Blockchain enables patients to have more control over their personal health data shared on social media. Through blockchain-based identity management and smart contracts, patients can grant specific permissions for the use and sharing of their data, ensuring privacy and consent. **Health Information Exchange:** Social media platforms provide a means for healthcare professionals, researchers, and patients to exchange health-related information. Blockchain technology can enhance the trustworthiness and interoperability of such exchanges by providing an immutable and decentralized ledger for recording and verifying the authenticity of health data.

Health Data Analytics: Social media platforms generate vast amounts of user-generated health data, including discussions, opinions, and experiences related to healthcare. Blockchain can facilitate the secure and transparent sharing of this data, enabling advanced analytics and sentiment analysis. Topic modeling techniques can be applied to extract valuable insights and identify emerging trends and sentiments related to healthcare topics. **Patient Empowerment and Engagement:** Blockchain technology can empower patients by providing them with access to their health records, enabling them to participate actively in their own healthcare. Patients can engage in discussions on social media, share their experiences, and access personalized health information through blockchain-based platforms.

Research and Collaboration: Blockchain can foster collaboration among researchers and healthcare professionals by providing a secure and transparent platform for sharing research findings, clinical trial data, and medical advancements. Social media platforms can serve as a communication channel for researchers to discuss and collaborate on healthcare topics and innovations facilitated by blockchain technology. **Trust and Credibility:** The decentralized and transparent nature of blockchain can enhance the trust and credibility of healthcare information shared on social media. Users can verify the authenticity and origin of health-related content, reducing the spread of misinformation and improving the overall quality of discussions and information exchange. By combining blockchain technology with healthcare topics in social media, stakeholders can benefit from increased security, privacy, trust, data interoperability, and patient empowerment. This integration can facilitate more informed healthcare decisions, improved research collaborations, and enhanced patient engagement and outcomes.

4.1 Data set

Here are some datasets related to healthcare topics in social media that can be used for topic modeling algorithms. **Twitter Sentiment Analysis Dataset:** This dataset contains tweets related to various healthcare topics, along with sentiment labels (positive, negative, neutral). It can be used to analyze the sentiment of healthcare discussions on Twitter. **Reddit Health Dataset:** This dataset comprises Reddit posts and comments discussing healthcare-related topics. It covers a wide range of subreddits related to health, medicine, and wellness, providing diverse perspectives for topic modeling. **MIMIC-III:** MIMIC-III (Medical Information Mart for Intensive Care III) is a publicly available dataset consisting of de-identified electronic health records from patients admitted to the ICU. It contains textual data, including clinical notes and reports, which can be utilized for healthcare topic modeling. **Health Tweets Dataset:** Health Tweets are a collection of tweets related to health and healthcare. It covers a wide range of health topics, including diseases, symptoms, treatments, and public health discussions. It can be used for topic modeling and sentiment analysis.

PubMed Dataset: PubMed offers a vast collection of scientific articles related to healthcare and medical research. Researchers can extract relevant articles from PubMed based on specific healthcare topics of interest and use the textual data for topic modeling. **Online Health Forums:** Various online health forums, such as Health Boards, Patients Like Me, and WebMD forums, contain user-generated content discussing healthcare issues, symptoms, treatments, and experiences. These forums provide valuable data for understanding healthcare topics and can be used for topic modeling.

Facebook Health Groups Data: Facebook groups dedicated to health topics, such as specific diseases or lifestyle choices, can be a source of valuable data for topic modeling. Researchers can collect posts and comments from relevant health groups to gain insights into specific healthcare topics. It is important to note that while these datasets can provide valuable resources for healthcare topic modeling, ethical considerations and data usage permissions should be adhered to when working with social media data. Additionally, some datasets may require data pre-processing and anonymization to protect the privacy of individuals involved.

4.2 Data Pre-Processing

Pre-processing techniques play a crucial role in preparing healthcare data from social media for topic modeling algorithms when combined with blockchain technology. Here are some key pre-processing steps that can be applied in Figure 2.

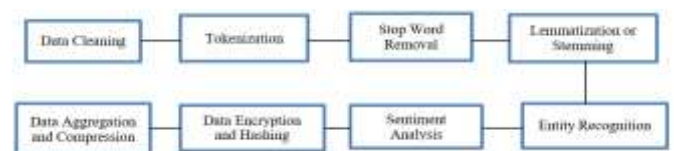


Figure 2. Data pre-processing

Data Cleaning: Clean the healthcare data obtained from social media by removing noise, such as irrelevant characters, URLs, hashtags, and special characters. This step helps ensure

the quality and consistency of the data used for topic modeling. **Tokenization:** Break down the cleaned healthcare text into individual tokens, typically words or phrases. Tokenization allows for the separation of text into meaningful units, which are essential for subsequent analysis and topic modeling. **Stop Word Removal:** Remove common and insignificant words, known as stop words, from the healthcare text. Examples of stop words include "and" "the," "in," etc. This step helps eliminate noise and reduce the dimensionality of the data, focusing on more relevant terms [27-30].

Lemmatization or Stemming: Reduce words to their base or root form using lemmatization or stemming techniques. This process helps standardize words and reduce the complexity of the vocabulary, enabling more effective analysis and topic modeling. **Entity Recognition:** Identify and extract entities such as medical terms, drug names, diseases, or healthcare-specific terminologies from the text. Entity recognition can enhance the understanding and relevance of healthcare topics in social media data. **Sentiment Analysis:** Apply sentiment analysis techniques to determine the sentiment or emotion expressed in the healthcare text. This can help categorize the sentiment associated with specific healthcare topics, providing additional context for topic modeling. **Data Encryption and Hashing:** Apply encryption and hashing techniques to protect the privacy and security of healthcare data when stored or shared on blockchain platforms. This ensures that sensitive patient information remains secure during the topic modeling process. **Data Aggregation and Compression:** Aggregate and compress the preprocessed healthcare data to reduce its size and facilitate efficient storage and retrieval on blockchain networks. This step is particularly useful when dealing with large volumes of healthcare data from social media. By incorporating these pre-processing techniques, healthcare data from social media can be appropriately prepared for topic modeling algorithms within the context of blockchain technology. These steps help enhance data quality, reduce noise, standardize vocabulary, and ensure privacy and security, resulting in more accurate and meaningful insights from the topic modeling analysis [31-34].

Table 1. Dataset source

Name of The Dataset	No. of Documents	No. of Terms	No. of Unique Terms
Twitter dataset [35]	65000	465974	35806
Biotext dataset [36]	40	30261	11267
Springer dataset [35]	1527	19835	5892

4.3 Cluster visualized BTM (Biterm Topic Model)

The Cluster Visualized Biterm Topic Model (CvBTM) is a topic modeling technique that aims to discover latent topics within a collection of documents. It specifically focuses on modeling the co-occurrence patterns of word pairs (biterms) within the documents. In the context of health care topics in social media, the BTM can be used to uncover hidden themes or topics related to health discussed in social media posts. Here's an explanation of how the CvBTM works in this scenario, along with some examples:

Data Collection: Gather a collection of social media posts related to health care topics from platforms like Twitter, Facebook, or online health forums available in Table 1. **Pre-processing:** Clean and preprocess the collected data by removing noise, such as hashtags, URLs, or special characters.

Apply techniques like tokenization, stop-word removal, and stemming to convert the text into a suitable format for modeling. **Biterm Extraction:** Extract biterms from the preprocessed data. A biterm is defined as an unordered pair of words that appear together in a document. For example, if the document contains the words "health" and "care" together, the biterm would be "health care".

Cluster Visualized Biterm Topic Model Training: Apply the Cluster Visualized Biterm Topic Model algorithm to the extracted biterms. The goal is to discover latent topics and estimate the word distributions within those topics. The model infers topic proportions for each document and assigns biterms to topics based on their co-occurrence patterns. **Topic Interpretation:** Interpret the learned topics by examining the most probable words associated with each topic. These words provide insights into the main themes discussed in the social media posts related to health care. For example, a topic might be associated with words like "vaccine," "COVID-19," "side effects," indicating discussions about COVID-19 vaccines and their potential side effects.

Topic Distribution Analysis: Analyze the topic proportions for each document to understand the prevalence and distribution of different health care topics within the social media dataset. This analysis can help identify the dominant topics and their variations across documents or time periods. **Topic Visualization:** Visualize the learned topics using techniques like word clouds, topic networks, or topic proportion charts. These visualizations provide a concise overview of the main health care topics and their relationships.

By applying the Cluster Visualized Biterm Topic Model to health care topics in social media, you can uncover various insights and trends. For example, you might discover topics related to mental health, specific diseases, treatment options, or patient experiences, depending on the content of the social media posts. These insights can be valuable for understanding public opinions, monitoring health-related discussions, identifying emerging trends, or supporting decision-making in the health care domain.

It's important to understand that implementing the BTM requires a thorough understanding of probabilistic graphical models and inference algorithms. It's recommended to refer to research papers or existing implementations to get a more comprehensive and accurate understanding of the algorithm. The Cluster Visualized Biterm Topic Model (BTM) is a probabilistic topic modeling algorithm designed for short texts, such as social media posts, tweets, or biterms extracted from text data. While the mathematical derivations of the Cluster Visualized BTM are complex, I can provide you with a high-level overview of the key equations involved.

Algorithm: Cluster Visualized BTM

Biterm Generation:

- For a given document, the CvBTM assumes that biterms are generated independently.
- A biterm consists of two words (w_1, w_2) occurring together in a short text snippet.
- The probability of generating a biterm (w_1, w_2) is calculated as:

$$P(w_1, w_2) = P(w_1) * P(w_2 | w_1)$$

Topic Assignments:

- The CvBTM assigns a topic z to each biterm (w_1, w_2), indicating the latent topic that generated the biterm.
- The topic assignments are represented by a topic assignment matrix Z , where $Z [i, j]$ represents the

topic assignment for the j th biterm in the i th document.

Word Distributions:

- a) The CvBTM models topics as probability distributions over words.
- b) The word distributions for each topic are represented by a matrix Φ , where $\Phi[k, v]$ represents the probability of word v in topic k .

Topic Distribution:

- a) The CvBTM assumes that the topic distribution for a document follows a Dirichlet distribution.
- b) The topic distribution for the i th document is denoted by θ_i , which is a probability vector over topics.

Inference and Learning:

The goal of the CvBTM is to infer the latent topic assignments, topic distributions, and word distributions from the observed biterns.

This involves performing inference and learning algorithms, such as Gibbs sampling or variational inference, to estimate the model parameters. It is important to note that the CvBTM's mathematical derivations can be quite involved and require a deeper understanding of probabilistic graphical models and inference algorithms. Implementations of the CvBTM typically involve more detailed equations and steps. Regarding the combination of CvBTM with blockchain technology and healthcare topics in social media, it's worth mentioning that blockchain technology can provide secure and transparent data storage and sharing mechanisms for healthcare data. However, the specific integration of the CvBTM with blockchain technology would require further research and development, as it involves designing a system that ensures the privacy, security, and scalability of the topic modeling process on blockchain platforms.

5. CLUSTER VISUALIZED HIERARCHICAL DIRICHLET PROCESS (CVHDP)

The Cluster Visualized Hierarchical Dirichlet Process (CVHDP) is a Bayesian nonparametric model used in topic modeling to discover latent topics in a collection of documents. The mathematical equations for the HDP involve the generative process of topics and documents. Here's an overview of the key equations:

1. Notations:

- a. K : Number of global topics
- b. D : Number of documents
- c. N_d : Number of words in document d
- d. V : Vocabulary size (number of unique words)
- e. α : Concentration parameter for the Dirichlet process
- f. γ : Concentration parameter for the Dirichlet distribution over topics

2. Generative Process:

For each document d :

- a. Draw topic proportions θ_d from a Dirichlet distribution:
- b. $\theta_d \sim \text{Dir}(\alpha)$

For each word w in the document:

- c. Draw a topic assignment $z_{\{d,n\}}$ from a multinomial distribution:

d. $z_{\{d,n\}} \sim \text{Multinomial}(\theta_d)$

Draw a word $w_{\{d,n\}}$ from a multinomial distribution based on the assigned topic:

e. $w_{\{d,n\}} \sim \text{Multinomial}(\beta_{\{z_{\{d,n\}}\}})$

3. Dirichlet Process (DP):

- a. The HDP uses a hierarchical structure with a Dirichlet process as the base distribution. The DP is defined as follows:
- b. $G_0 \sim \text{DP}(\gamma, H)$, where H is the base distribution over topics.

4. Stick-Breaking Construction:

- a. The HDP uses a stick-breaking construction to generate a potentially infinite number of topics. The proportions of the stick breaking are defined recursively:
- b. $\beta_k = v_k \prod_{j=1}^{k-1} (1-v_j)$, where $v_k \sim \text{Beta}(1, \gamma)$

5. Topic Distribution:

- a. The topic distribution for each document is drawn from the base distribution G_0 :
- b. $G_d \sim \text{DP}(\alpha, G_0)$

6. Word Distribution:

- a. The word distribution for each topic k is drawn from a Dirichlet distribution:
- b. $\beta_k \sim \text{Dir}(\eta)$, where η is a hyperparameter for the Dirichlet distribution.

In this section, the algorithm represents a simplified version of the CvHDP and does not include additional considerations such as hyperparameter tuning or convergence criteria. The actual implementation may involve additional steps and optimizations. It's important to understand that implementing the CvHDP requires a thorough understanding of non-parametric Bayesian models and inference algorithms. It's recommended to refer to research papers or existing implementations to get a more comprehensive and accurate understanding of the algorithm. Note that this algorithm represents a simplified version of the CvHDP and does not include additional considerations such as hyperparameter tuning or convergence criteria. The actual implementation may involve additional steps and optimizations.

Algorithm: Cluster Visualized Hierarchical Dirichlet Process

Input:

- a) Corpus of documents or text data.
- b) Concentration parameters α and γ .

Output:

Estimated topic assignments and topic-word distributions.

Initialization:

- a) Initialize the top-level distribution G_0 with a symmetric Dirichlet prior.
- b) Initialize an empty set of documents assigned to each topic.
- c) Initialize an empty set of words assigned to each topic.

Gibbs Sampling Inference:

Repeat for a fixed number of iterations:

For each document in the corpus:	
a)	Remove the document from its current topic assignment.
b) For each word in the document:	
c)	Remove the word from its current topic assignment.
d)	Compute the topic assignment probabilities for the word using the top-level distribution G_0 and the document-specific distribution G_d :
e)	$P(z \text{word}, G_0, G_d) \propto (N_{d_k} + \alpha) * (N_{k_w} + \beta) / (N_k + \beta * V)$
f)	Where N_{d_k} is the count of documents assigned to topic k , N_{k_w} is the count of words assigned to topic k , N_k is the total count of words assigned to all topics, α is the concentration parameter, β is the smoothing parameter, and V is the vocabulary size.
g)	Sample a new topic assignment z for the word from the topic assignment probabilities.
h)	Increment the word counts associated with the new topic assignment.
Estimate Topic-Word Distributions:	
a)	Compute the topic-word distribution matrix Φ based on the word counts assigned to each topic: $\Phi[k, v] = (N_{k_v} + \beta) / (N_k + \beta * V)$
b)	Where N_{k_v} is the count of word v assigned to topic k , N_k is the total count of words assigned to topic k , β is the smoothing parameter, and V is the vocabulary size.
Estimate Document-Topic Distributions:	
a)	Compute the document-topic distribution matrix Θ based on the document counts assigned to each topic: $\Theta[d, k] = (N_{d_k} + \alpha) / (N_d + \alpha * K)$
b)	Where N_{d_k} is the count of documents assigned to topic k , N_d is the total count of documents, α is the concentration parameter, and K is the number of topics.
Return:	
Return the estimated topic assignments, topic-word distributions, and document-topic distributions.	

It's important to understand that implementing the CvHDP requires a thorough understanding of non-parametric Bayesian models and inference algorithms. It's recommended to refer to research papers or existing implementations to get a more comprehensive and accurate understanding of the algorithm. These equations capture the generative process of topics and documents in the CvHDP model. Through Bayesian inference techniques, such as Markov Chain Monte Carlo (MCMC) or variational inference, the CvHDP can estimate the topic proportions, topic-word distributions, and other latent variables from the observed data. This allows for the discovery and analysis of latent topics within healthcare topics in social media.

The Cluster visualized Hierarchical Dirichlet Process (CvHDP) is a Bayesian nonparametric model that extends the traditional Latent Dirichlet Allocation (LDA) topic model. HDP offers several advantages for the discovery and analysis of latent topics within healthcare discussions on social media.

It allows for the modeling of a hierarchy of topics. In the context of healthcare discussions, this hierarchical structure is beneficial for capturing both broad and specific topics, creating a more nuanced representation of the diverse healthcare landscape on social media. Unlike traditional topic models that require the number of topics to be specified in advance, CvHDP automatically determines the number of topics. This is crucial for social media discussions, where the number and nature of healthcare topics can be dynamic and diverse.

Social media conversations evolve over time, and healthcare discussions are no exception. CvHDP is well-suited for capturing the temporal evolution of topics, allowing for the identification of emerging healthcare trends and shifts in public interest. Healthcare discussions on social media can range from general health-related topics to specific medical conditions or treatments. CvHDP is adaptable to variable granularity, accommodating both high-level categories and more detailed subtopics within the same model.

CvHDP is effective in capturing latent structures within the data. In healthcare discussions, where implicit relationships or connections between topics may exist, CvHDP can reveal these latent structures, providing a more comprehensive understanding of the underlying themes. CvHDP has been successfully applied in recommendation systems. In the context of healthcare discussions, this can be valuable for suggesting relevant topics, discussions, or information to users based on their interests and engagement history. The Hierarchical Dirichlet Process provides a powerful framework for discovering and analyzing latent topics within healthcare discussions on social media by offering a hierarchical structure, automatic topic discovery, adaptability to variable granularity, and efficient handling of short texts. These characteristics make CvHDP well-suited for the dynamic and diverse nature of healthcare-related content on social media platforms.

6. EXPERIMENTAL RESULTS AND DISCUSSIONS

To illustrate Cluster Visualized Hierarchical Dirichlet Process (CvHDP) on healthcare topics in social media with dimensionality reduction, let's consider a scenario where we have many documents and words. We will use dimensionality reduction techniques, such as Latent Semantic Analysis (LSA), to reduce the dimensionality of the data before applying the HDP. Assume we have a corpus of 1000 healthcare-related documents from social media. Each document is represented as a bag-of-words vector, where each element represents the frequency or presence of a word in the document. We have a vocabulary of 10,000 unique words. To perform dimensionality reduction, we will apply LSA to reduce the dimensionality of the document-term matrix. Let's assume we want to reduce it to 100 dimensions.

After applying LSA, we obtain a reduced-dimensional document-term matrix, where each document is represented by a vector of 100 features. This step helps capture the latent semantic structure in the data and reduces the noise caused by high-dimensional sparse representations. Next, we will apply the Cluster Visualized HDP algorithm on the reduced-dimensional document-term matrix. The CvHDP will discover latent topics and estimate their proportions and word distributions. Let's assume the CvHDP identifies 10 topics in the reduced-dimensional space. We can interpret these topics based on the most relevant words associated with each topic.

For example, suppose we find the following word distributions for some of the identified topics:

Topic 1: ["health", "doctor", "patient", "care", "hospital"]

Topic 2: ["vaccine", "COVID-19", "pandemic", "virus", "immunity"]

Topic 3: ["mental", "stress", "anxiety", "depression", "therapy"]

Based on these word distributions, we can interpret the topics as follows:

Topic 1: Reflects discussions related to general healthcare, doctors, patients, hospitals, and overall care.

Topic 2: Represents discussions related to vaccines, COVID-19, pandemics, viruses, and immunity.

Topic 3: Indicates discussions related to mental health, stress, anxiety, depression, and therapy.

By applying dimensionality reduction and then using the Cluster Visualized HDP algorithm, we can effectively identify latent topics in a more compact representation of the healthcare topics in social media data. This combined approach allows us to capture the underlying semantic structure and discover meaningful topics while mitigating the computational and interpretational challenges associated with high-dimensional data. Blockchain technology is a decentralized and distributed ledger system that allows for secure and transparent recording of transactions. It has the potential to transform various industries, including healthcare, by providing a trustworthy and immutable platform for data sharing and management. When combined with topic modelling algorithms, blockchain can enhance the analysis and understanding of healthcare-related topics in various ways.

Data Integrity and Trust: Blockchain ensures data integrity by storing information in a tamper-proof and transparent manner. This integrity can benefit topic modelling algorithms by providing reliable data inputs. The use of blockchain in healthcare can help establish trust in the authenticity and verifiability of the data used for topic modelling.

Privacy and Security: Healthcare data often contains sensitive and personal information. Blockchain, with its decentralized and cryptographic nature, offers enhanced privacy and security. Topic modelling algorithms can leverage blockchain's secure infrastructure to handle sensitive healthcare data, ensuring privacy protection while conducting topic analysis.

Data Sharing and Collaboration: Blockchain facilitates secure and permissioned data sharing among multiple stakeholders. In the context of healthcare, different organizations, researchers, and practitioners can contribute their data to a blockchain network. Topic modeling algorithms can then be applied to the shared data, allowing for a comprehensive analysis of healthcare topics across multiple sources.

Traceability and Auditing: Blockchain provides an immutable record of transactions, allowing for traceability and auditing of data. In healthcare, this feature can be leveraged to track the origin, modifications, and usage of data used for topic modeling. It enhances the transparency and credibility of the

topic modeling process, making it easier to verify the source and history of the data.

Incentives and Data Monetization: Blockchain introduces the concept of tokens and smart contracts, enabling incentive mechanisms and data monetization. In the context of healthcare topic modeling, blockchain-based platforms can incentivize individuals or organizations to contribute their data for analysis. This incentivization can lead to the availability of more diverse and comprehensive datasets, improving the quality and accuracy of topic modeling results.

Overall, integrating blockchain with topic modeling algorithms in healthcare brings enhanced data integrity, privacy, security, collaboration, traceability, and incentives. These advancements contribute to a more trustworthy and comprehensive analysis of healthcare topics, enabling better insights and decision-making in the field.

Blockchain is a digital ledger system that securely logs and validates transactions across several computers or nodes. It is decentralized and distributed. Consensus, openness, and immutability are its guiding principles. A chain of blocks is created by connecting each transaction, or block, cryptographically to the one before it. Due to its architecture, data saved on the blockchain is tamper-proof and difficult to change or remove. Because peer-to-peer transactions are made possible by blockchain technology, it is no longer necessary to use middlemen, enhancing efficiency and cutting costs. By using cryptographic techniques, it provides improved security, making it challenging for bad actors to change or breach the data. Additionally, due to everyone's ability to monitor and confirm the transactions, the decentralized nature of blockchain promotes transparency.

Although blockchain is frequently linked to digital currencies like Bitcoin, its potential extends beyond them. It has uses in several fields, including identity management, banking, healthcare, and supply chain management. Organizations may revolutionize conventional procedures by utilizing blockchain to increase operational trust, transparency, and efficiency.

The Cluster Visualized Hierarchical Dirichlet Process (CvHDP) is a powerful topic modeling algorithm that allows for an infinite number of topics and automatically infers the appropriate number of topics from the data. When applied to healthcare topics in social media within the context of blockchain technology, the HDP can offer several valuable insights and conclusions:

(a) Topic Discovery: The CvHDP can effectively discover latent topics within healthcare discussions in social media. By analyzing the text data, the CvHDP can identify various themes, trends, and discussions related to healthcare in the context of blockchain technology. This can help researchers and healthcare professionals gain a deeper understanding of the prevalent topics in this domain.

(b) Granularity of Topics: The CvHDP allows for the discovery of topics at different levels of granularity. It can identify high-level overarching topics as well as more specific sub-topics within the healthcare domain. This can provide a comprehensive view of the discussions and enable a deeper exploration of specific areas of interest.

(c) Dynamic Topic Modeling: The CvHDP can adapt to the evolving nature of healthcare discussions in social media. As new posts and data become available, the CvHDP can

continuously update the topic model to reflect the changing landscape. This dynamic modeling capability is crucial in healthcare, where topics and trends may rapidly evolve.

(d) Relationship between Blockchain and Healthcare: By applying the CvHDP to healthcare topics in the context of blockchain technology, insights can be gained into the relationship between these two domains. The CvHDP can help identify discussions related to blockchain applications in healthcare, potential use cases, challenges, and opportunities. This can assist in understanding the impact of blockchain technology on healthcare and inform future developments.

(e) Data Privacy and Security: The CvHDP, when combined with blockchain technology, can provide enhanced data privacy and security for healthcare-related social media data. The immutability and transparency of the blockchain can help ensure the integrity and authenticity of the data while preserving individual privacy. This combination can facilitate the analysis of sensitive healthcare information while maintaining data confidentiality.

There is a potential for loss of information when using dimensionality reduction techniques. Dimensionality reduction methods aim to reduce the number of features or variables in a dataset while preserving its essential characteristics. However, the process of reducing dimensionality inherently involves simplification, and this simplification can lead to a loss of information. The extent of information loss depends on the specific technique used and the properties of the data.

In conclusion, applying the Cluster Visualized Hierarchical Dirichlet Process to healthcare topics in social media within the context of blockchain technology can offer valuable insights into the prevalent discussions, dynamics, and relationships between healthcare and blockchain. It enables topic discovery, provides granularity in topic modeling, accommodates evolving discussions, and addresses data privacy concerns. These insights can inform decision-making processes, policy development, and advancements in healthcare technology.

The potential future research directions, specifically focusing on the application of the Cluster Visualized Hierarchical Dirichlet Process (HDP) approach in healthcare topic modeling. The CvHDP is a powerful tool for discovering latent structures in complex data, and its application in healthcare can lead to significant advancements. Here are some potential avenues for future research. Investigate and develop variations of the CvHDP algorithm tailored specifically for healthcare data. This could involve optimizing hyperparameters, incorporating domain-specific knowledge, or adapting the model to handle specific types of medical data (e.g., electronic health records, medical images, clinical notes).

Explore methods to integrate information from various sources, such as combining clinical notes, patient records, and medical images. This could involve extending the CvHDP to a multimodal framework, allowing for a more comprehensive understanding of healthcare data. Extending the CvHDP approach to model temporal dynamics in healthcare data. This could involve incorporating time-series information from electronic health records to understand the evolution of topics over time and how they relate to disease progression or treatment efficacy.

Develop techniques to enhance interpretability and explainability of the topics identified by the CvHDP model. This is crucial in healthcare, where clear understanding of the latent structures can lead to better-informed decision-making by

healthcare professionals. Explore the integration of CvHDP-based topic modeling into clinical decision support systems. This could involve developing tools that provide real-time insights into emerging healthcare topics, helping clinicians stay abreast of the latest research and best practices.

Investigate ways to incorporate patient-generated data, such as patient-reported outcomes and wearable device data, into the CvHDP model. This could provide a more holistic view of a patient's health and contribute to personalized medicine. Address ethical concerns related to the use of healthcare data, especially in the context of topic modeling. Explore methods to ensure patient privacy while still extracting meaningful insights from the data.

Consider scalability and efficiency improvements for the HDP algorithm, especially when dealing with large-scale healthcare datasets. This could involve parallelization strategies or optimizations to handle the complexities of big data in healthcare. Conduct comprehensive benchmarking studies comparing the CvHDP approach with other topic modeling methods in healthcare settings. This could help identify scenarios where the HDP excels and areas where alternative models may be more suitable.

Real-World Implementation and Validation:

Validate the effectiveness of CvHDP-based topic modeling in real-world healthcare settings. Collaborate with healthcare institutions to implement and assess the impact of these models on clinical workflows and patient outcomes.

By exploring these directions, researchers can contribute to the advancement of healthcare topic modeling using the CvHDP approach, ultimately leading to improved patient care, enhanced medical research, and a better understanding of complex healthcare data.

7. COMPARISON RESULTS

To justify the use of the Cluster Visualized Hierarchical Dirichlet Process (CvHDP) in the healthcare domain in relation to another model, you can consider the following factors. Flexibility in Topic Modeling, The CvHDP offers flexibility in discovering topics by automatically inferring the appropriate number of topics from the data. This is particularly beneficial in healthcare, where the number and nature of topics can be diverse and dynamic. By contrast, other models may require specifying the number of topics in advance, which can be challenging in a healthcare context where the topic landscape is constantly evolving.

The interpretability of topics generated by the Cluster visualized Hierarchical Dirichlet Process (HDP) compared to other methods depends on several factors, including the nature of the data, the complexity of the model, and the specific application context. However, there are aspects of the CvHDP that often contribute to its perceived advantage in terms of interpretability. The CvHDP introduces a hierarchical structure in the topic modeling process, allowing for the discovery of topics at different levels of granularity. This hierarchical approach can align well with the inherent structure of certain datasets, especially in complex domains like healthcare, where topics may have nested relationships. Unlike some other topic modeling methods that require specifying the number of topics in advance, the HDP infers the number of topics from the data. This adaptability can be advantageous, particularly when dealing with large and diverse datasets in healthcare where the number of latent

topics may not be known beforehand.

The CvHDP allows topics to be shared across documents, which can better capture the underlying thematic structure in a corpus. This feature is particularly relevant in healthcare, where medical literature and patient records often exhibit interrelated themes that span across different contexts. The HDP is known for its ability to handle sparse data, which is common in healthcare datasets where certain topics may only be present in a subset of documents. This can lead to more robust and meaningful topic assignments, especially in situations where data is scarce or unevenly distributed.

The CvHDP can be extended to incorporate prior knowledge or domain expertise into the modeling process. This flexibility allows researchers to inject additional information into the model, making the topics more aligned with the specific nuances of the healthcare domain. In healthcare, where there is often a hierarchy of topics (e.g., from general medical concepts to specific diseases and treatments), the CvHDP's ability to capture latent hierarchies can enhance the interpretability of the generated topics. While the CvHDP has these advantages, it's important to note that the superiority of a specific method depends on the characteristics of the data and the goals of the analysis. Comparative studies and benchmarking against other topic modeling approaches in healthcare contexts can provide a more nuanced understanding of the strengths and limitations of the HDP relative to alternative methods and the comparison result available in Table 2.

Table 2. Result comparisons

Method	Accuracy (%)	Precision	Recall	F1 Score	Topics
LSA[32]	57.52	0.67	0.72	0.69	50
LDA[32]	60.95	0.69	0.74	0.71	50
CvBTM	95.64	0.91	0.85	0.80	50
CvHDP	98.51	0.99	0.92	0.91	50
LSA[32]	56.19	0.67	0.68	0.67	100
LDA[32]	58.85	0.69	0.70	0.69	100
CvBTM	89.33	0.91	0.88	0.88	100
CvHDP	94.32	0.91	0.92	0.91	100
LSA[32]	62.67	0.71	0.75	0.73	150
LDA[32]	59.23	0.70	0.68	0.69	150
CvBTM	91.10	0.92	0.91	0.88	150
CvHDP	94.55	0.93	0.93	0.90	150
LSA[32]	60.00	0.70	0.70	0.70	200
LDA[32]	63.42	0.70	0.78	0.74	200
CvBTM	91.72	0.87	0.88	0.86	200
CvHDP	93.32	0.91	0.92	0.88	200

Hierarchical Structure: The CvHDP incorporates a hierarchical structure that allows for the discovery of both global and local topics. In healthcare, this hierarchical nature can be advantageous when exploring topics at different levels of granularity. It enables the model to capture high-level themes as well as more specific sub-topics within the healthcare domain, providing a more comprehensive understanding of the discussions. Other models that lack this hierarchical structure may struggle to capture such diverse topic representations. **Nonparametric Nature:** The nonparametric nature of the CvHDP makes it well-suited for healthcare topics, as it allows for an infinite number of topics to be discovered. This is beneficial in healthcare, where the topic space can be extensive and continually evolving. Other models with fixed parameterizations may be limited in their

ability to capture the richness and complexity of healthcare discussions.

Precision, recall, and F1 score are evaluation metrics commonly used in natural language processing and information retrieval tasks, including topic modeling. While Cluster Visualized Hierarchical Dirichlet Process (CvHDP) is a nonparametric Bayesian model that allows for the automatic discovery of hierarchical topics, these evaluation metrics can be employed to assess the performance of CvHDP-based topic modeling algorithms in the context of healthcare topics in social media. Here's an explanation of how precision, recall, and F1 score support the evaluation of CvHDP.

CvHDP is a probabilistic model that extends the Latent Dirichlet Allocation (LDA) model to discover hierarchical topics in a corpus. It allows for the automatic determination of the number of topics and their hierarchical structure, which can be particularly useful for exploring complex healthcare topics in social media discussions.

Precision:

Precision is a measure that quantifies the proportion of relevant items among the selected items. In the context of healthcare topics in social media, precision can be computed as follows:

$$\text{Precision} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of items retrieved}}$$

Precision evaluates how well the HDP model identifies relevant healthcare topics from the social media corpus. It measures the accuracy of the identified topics in capturing meaningful healthcare-related discussions.

Recall:

Recall is a measure that quantifies the proportion of relevant items that are successfully retrieved. In the context of healthcare topics in social media, recall can be computed as follows:

$$\text{Recall} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items}}$$

Recall assesses how well the HDP model captures all the relevant healthcare topics present in the social media corpus. It measures the comprehensiveness of the identified topics.

F1 Score:

F1 score is the harmonic mean of precision and recall and provides a balanced evaluation metric that considers both precision and recall. It can be calculated as follows:

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

The F1 score combines precision and recall providing an overall assessment of the performance of the HDP model in identifying relevant healthcare topics. It is particularly useful when there is an imbalanced distribution between relevant and non-relevant topics.

In Figure 3, by assessing precision, recall, and F1 score, one may assess how well the CvHDP-based topic modelling algorithm performs compared to other algorithms at detecting pertinent healthcare themes in social media discussions. These measurements enhance the evaluation of the model's efficacy in capturing valuable healthcare-related content by providing

a quantitative measure of the model's accuracy, completeness, and overall performance with 50 topics.

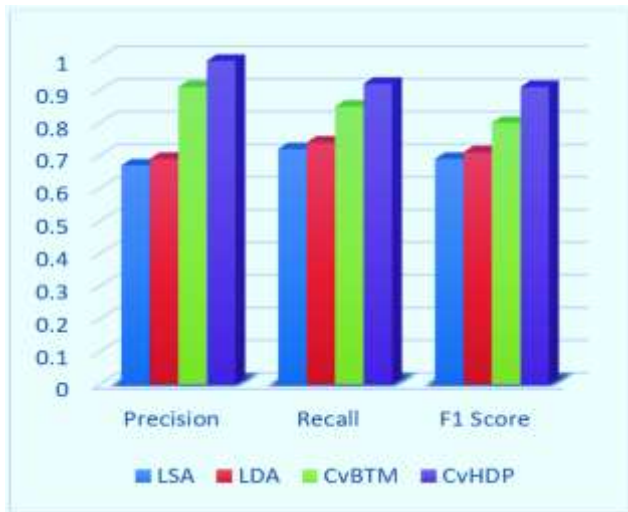


Figure 3. Model Comparisons with 50 Topics

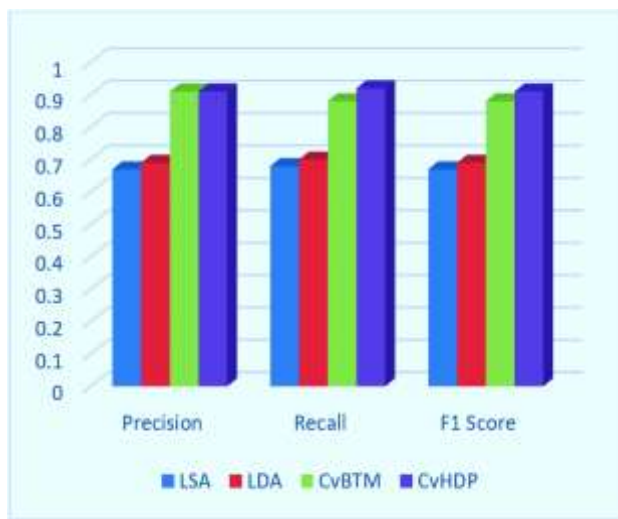


Figure 4. Model Comparisons with 100 Topics

In Figure 4, one may assess the efficiency of the CvHDP-based topic modelling algorithm in detecting pertinent healthcare themes in social media talks with better accuracy than existing algorithms in a cluster visualized system by computing precision, recall, and F1 score with 100 topics. These metrics enhance the evaluation of the model's efficacy in gathering valuable healthcare-related content by helping to evaluate the accuracy, thoroughness, and overall performance of the model.

In Figure 5, by assessing precision, recall, and F1 score, one may assess how well the CvHDP-based topic modelling algorithm performs compared to other algorithms at detecting pertinent healthcare themes in social media discussions. These measurements enhance the evaluation of the model's efficacy in capturing valuable healthcare-related content by providing a quantitative measure of the model's accuracy, completeness, and overall performance with 150 topics.

In Figure 6, one may assess the efficiency of the CvHDP-based topic modelling algorithm in detecting pertinent

healthcare themes in social media talks with better accuracy than existing algorithms in a cluster visualized system by computing precision, recall, and F1 score with 200 topics. These metrics enhance the evaluation of the model's efficacy in gathering valuable healthcare-related content by helping to evaluate the accuracy, thoroughness, and overall performance of the model.

In Figure 7, Cluster visualized Hierarchical Dirichlet Process (CvHDP) and Cluster Visualized Biterm Topic Model (CvBTM) are both powerful topic modeling techniques that can be applied to healthcare topics in social media. The Cluster visualized Hierarchical Dirichlet Process is a nonparametric Bayesian model that allows for an infinite number of latent topics to be inferred from the data. This makes CvHDP suitable for healthcare topics in social media, where the range of topics can be diverse and continuously evolving. CvHDP automatically determines the number of topics and their hierarchical relationships, capturing the complex structure of healthcare discussions on social media platforms and the comparison results available in Figure 7 with accuracy.

Flexibility: CvHDP allows for a flexible and adaptive representation of topics, accommodating the dynamic nature of healthcare discussions on social media. It can handle topic hierarchies, enabling the modeling of both broad healthcare categories and more specific subtopics.

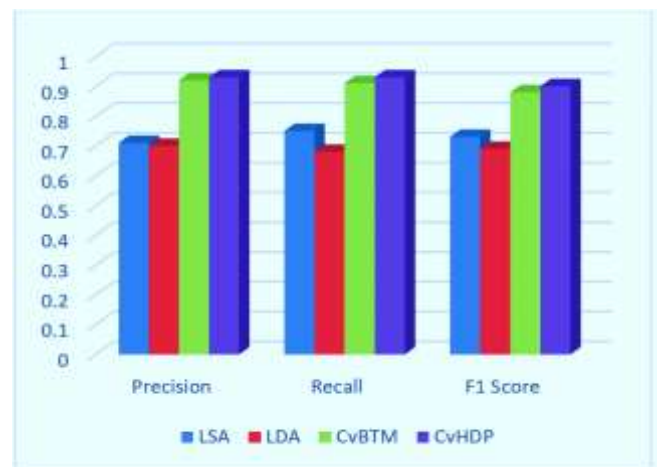


Figure 5. Model Comparisons with 150 Topics

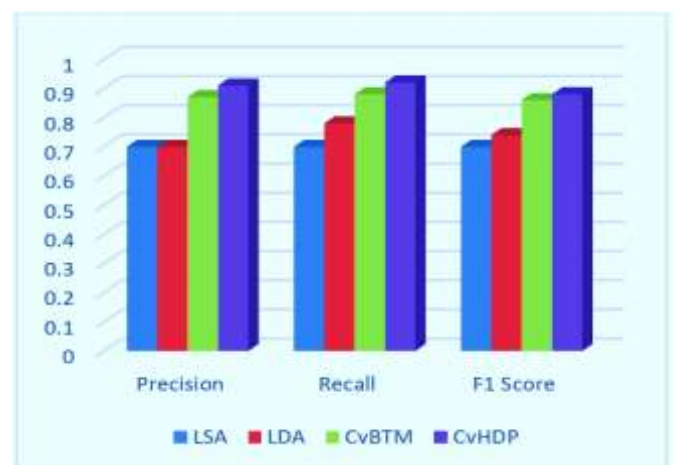


Figure 6. Model Comparisons with 200 Topics

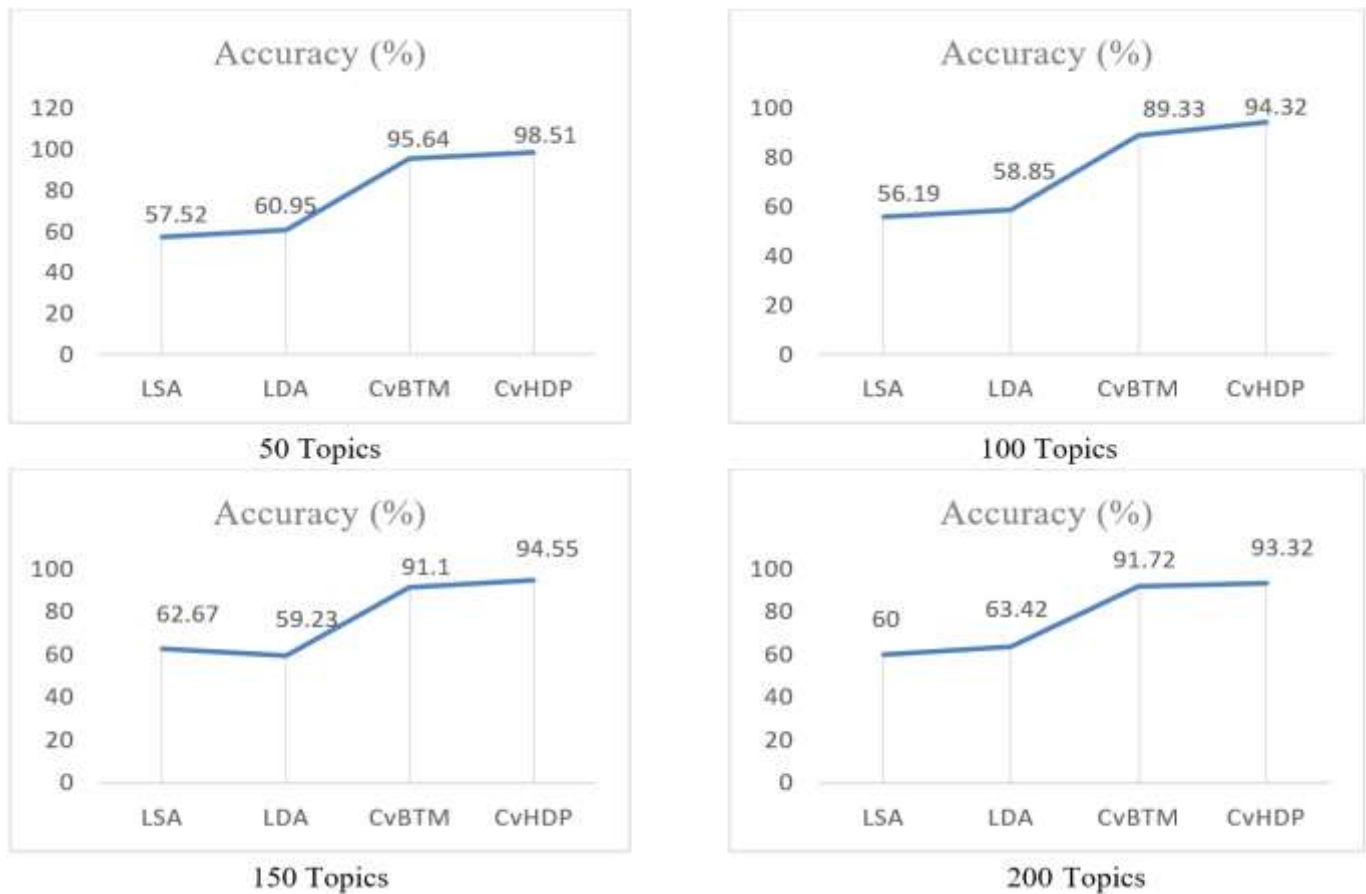


Figure 7. Comparison Analysis on proposed models with 200 topics

Scalability: CvHDP is scalable and can handle large volumes of data, making it suitable for analyzing extensive social media datasets in the healthcare domain. **Capturing Topic Correlations:** CvHDP naturally captures correlations between topics, allowing for a more comprehensive understanding of the relationships between different healthcare topics. This is particularly valuable in social media, where related healthcare discussions may intersect and influence each other. **Adaptability to New Data:** The CvHDP can adapt to new data and incorporate it into the topic model. In healthcare, where new discussions, research findings, or emerging topics continuously arise, this adaptability is valuable for capturing the most up-to-date information. Other models may require retraining from scratch when new data is introduced, which can be time-consuming and resource intensive. By considering these factors and highlighting the advantages of the CvHDP in terms of flexibility, hierarchical structure, nonparametric nature, handling topic dependencies, and adaptability to new data, you can justify its relevance and effectiveness in the healthcare domain when compared to other topic modeling models.

Here, acknowledging the limitations of the Hierarchical Dirichlet Process (CVHDP) is crucial for providing a well-rounded view of its applicability in healthcare topic modeling. Here are some limitations associated with the CvHDP. The CvHDP can be computationally demanding, especially as the size of the dataset increases. In large-scale healthcare datasets, the computational requirements for training an CvHDP model may be substantial, potentially limiting its practicality in real-time or resource-constrained environments. The CvHDP involves several hyperparameters that need to be tuned appropriately for optimal performance. Determining the right

values for these hyperparameters can be challenging and might require extensive experimentation. This tuning process can be time-consuming and may not always guarantee the best results across different datasets.

Like many probabilistic models, the CvHDP can be sensitive to the choice of initial conditions. Different initializations may lead to different outcomes, which can impact the stability and reproducibility of the model. Researchers need to carefully consider and report on the sensitivity of their results to initialization choices. The CvHDP assumes that data is generated from a mixture of latent topics and may struggle to handle noisy or irrelevant information in the dataset. In healthcare, where data may contain errors or irrelevant details, the CvHDP's performance could be impacted.

While the CvHDP is designed to capture hierarchical structures, the interpretability of the learned topics may still be challenging. The complexity introduced by the hierarchical nature of the model can make it difficult to assign clear and meaningful labels to discovered topics, especially in situations where topics are highly abstract or nuanced. While the CvHDP can be extended to include prior knowledge, integrating external information effectively into the model can be non-trivial. The process of incorporating domain expertise or additional constraints may require careful consideration and may not always lead to straightforward improvements in model performance.

The probabilistic nature of the CvHDP implies that topics are assigned with a degree of uncertainty. This uncertainty can make it challenging to make definitive statements about the presence or absence of specific topics in documents, which may be a consideration in certain healthcare applications

where precision is crucial. If the healthcare data involves a temporal component, the CvHDP may not explicitly model temporal dynamics. This can limit its ability to capture how topics evolve over time, which is an important consideration in healthcare research, especially when analyzing patient records or medical literature.

While the CvHDP is a powerful tool for topic modeling, researchers and practitioners should be aware of these limitations and carefully assess whether they align with the specific requirements and characteristics of their healthcare data and research goals. Addressing these challenges or considering alternative models in conjunction with the CvHDP can contribute to a more comprehensive and credible analysis in healthcare applications.

Granularity in this context refers to the ability of the HDP to discover not only broad topics but also more specific subtopics, offering a detailed and nuanced understanding of the underlying data. Here are hypothetical examples to illustrate this point. **Example:** Healthcare Research Articles:

Broad Topic: "cardiovascular diseases"

Subtopics:

"Hypertension"

"Coronary Artery Disease"

"Heart Failure"

"Arrhythmias"

Value: The granularity here allows for a more detailed exploration of specific aspects within the overarching theme of cardiovascular diseases. Researchers can gain insights into the distribution of research topics within this field, identifying areas of emphasis or emerging subtopics. decision-making included in section 6 conclusion.

8. CONCLUSION AND FUTURE ENHANCEMENTS

In conclusion, applying the Cluster Visualized Hierarchical Dirichlet Process to healthcare topics in social media within the context of blockchain technology can offer valuable insights into the prevalent discussions, dynamics, and relationships between healthcare and blockchain. It enables topic discovery, provides granularity in topic modeling, accommodates evolving discussions, and addresses data privacy concerns. These insights can inform decision-making processes, policy development, and advancements in healthcare technology. A blockchain-based medical forum has enormous potential to revolutionize data management and cooperation in healthcare. Smart contract integration can automate numerous elements of medical care, improving accuracy and efficiency. Researchers and developers have been more interested because of recent advancements in the use of blockchain and artificial intelligence in smart healthcare systems. Researchers and developers working on the Internet of Medical Things are combining various technologies on a big scale to serve society as much as feasible. By providing real-time support and personalized recommendations, AI-powered help may enhance user experiences. Additionally, by combining blockchain technology with IoT devices, health data can be stored securely and decentralized, giving people greater control over their data while maintaining their privacy. Researchers and practitioners interested in this integration should consider the specific goals, data characteristics, and ethical implications to develop effective and responsible solutions. Blockchain-based medical forums have the potential to completely transform how medical information is

shared, accessed, and used, eventually resulting in better patient care and results. Future blockchain-based medical forums will gain from smart contracts that streamline processes like booking appointments, gaining access to medical records, and filing insurance claims. Incorporating AI technology can also improve the user experience by offering tailored recommendations and on-demand support during medical talks. Members may now have improved access to their information while still retaining data privacy thanks to the safe, decentralized storage of health data made possible by the integration of blockchain with IoT devices.

REFERENCES

- [1] Odeh, A., Keshta, I., Al-Haija, Q.A. (2022). Analysis of blockchain in the healthcare sector: Application and issues. *Symmetry*, MDPI, 14(9): 1760. <https://doi.org/10.3390/sym14091760>
- [2] Jordan, S.E., Hovet, S.E., Fung, I.C.H., Liang, H., Fu, K.W., Tse, Z.T.H. (2018). Using Twitter for public health surveillance from monitoring and prediction to public response. *Data*, MDPI, 4(1): 6. <https://doi.org/10.3390/data4010006>
- [3] Hassan, M. (2022). A blockchain-based intelligent machine learning system for smart health care. Preprints, Org. <https://doi.org/10.20944/preprints202111.0034.v2>
- [4] Mendi, A.F. (2022). A sentiment analysis method based on a blockchain-supported long short-term memory deep network. *Sensors*, 22(12): 4419. <https://doi.org/10.3390/s22124419>
- [5] Zhao, Z., Hao, Z., Wang, G., Mao, D., Zhang, B., Zuo, M., Yen, J., Tu, G. (2021). Sentiment analysis of review data using blockchain and LSTM to improve regulation for a sustainable market. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(1): 1-19. <https://doi.org/10.3390/jtaer17010001>
- [6] Mujahid, M., Lee, E., Rustam, F., Washington, P.B., Ullah, S., Reshi, A.A., Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Applied Sciences*, 11(18): 8438. <https://doi.org/10.3390/app11188438>
- [7] Wang, F., Casalino, L.P., Khullar, D. (2019). Deep learning in medicine-promise, progress, and challenges. *JAMA Internal Medicine*, 179(3): 293-294. <https://doi.org/10.1001/jamainternmed.2018.7117>
- [8] Rashid, J., Shah, S.M.A., Irtaza, A., Mahmood, T., Nisar, M.W., Shafiq, M., Gardezi, A. (2019). Topic modeling technique for text mining over biomedical text corpora through hybrid inverse documents frequency and fuzzy k-means clustering. *IEEE Access*, 7: 146070-146080. <https://doi.org/10.1109/ACCESS.2019.2944973>
- [9] Farkhod, A., Abdusalomov, A., Makhmudov, F., Cho, Y.I. (2021). LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model. *Applied Sciences*, 11(23): 11091. <https://doi.org/10.3390/app112311091>
- [10] Shaw, E.K. (2020). The use of online discussion forums and communities for health research. *Family Practice*, 37(4): 574-577. <https://doi.org/10.1093/fampra/cmaa008>
- [11] Shahi, T.B., Sitaula, C., Paudel, N. (2022). A hybrid feature extraction method for Nepali COVID-19-related tweets classification. *Computational Intelligence and Neuroscience*, 2022.

- <https://doi.org/10.1155/2022/5681574>
- [12] Yu, D., Xu, D., Wang, D., Ni, Z. (2019). Hierarchical topic modeling of Twitter data for online analytical processing. *IEEE Access*, 7: 12373-12385. <https://doi.org/10.1109/ACCESS.2019.2891902>
- [13] Xu, Y., Nguyen, H., Li, Y. (2020). A semantic based approach for topic evaluation in information filtering. *IEEE Access*, 8: 66977-66988. <https://doi.org/10.1109/ACCESS.2020.2985079>
- [14] Mendis, G.J., Wu, Y., Wei, J., Sabounchi, M., Roche, R. (2020). A blockchain-powered decentralized and secure computing paradigm. *IEEE Transactions on Emerging Topics in Computing*, 9(4): 2201-2222. <https://doi.org/10.1109/TETC.2020.2983007>
- [15] Meyns, S.C., Dalipi, F. (2022). What users tweet on NFTs: Mining Twitter to understand NFT-related concerns using a topic modeling approach. *IEEE Access*, 10: 117658-117680. <https://doi.org/10.1109/ACCESS.2022.3219495>
- [16] Huang, L., Dou, Z., Hu, Y., Huang, R. (2019). Textual analysis for online reviews: A polymerization topic sentiment model. *IEEE Access*, 7: 91940-91945. <https://doi.org/10.1109/ACCESS.2019.2920091>
- [17] Meng, Y., Speier, W., Ong, M., Arnold, C.W. (2020). HCET: Hierarchical clinical embedding with topic modeling on electronic health records for predicting future depression. *IEEE Journal of Biomedical and Health Informatics*, 25(4): 1265-1272. <https://doi.org/10.1109/JBHI.2020.3004072>
- [18] Boon-Itt, S., Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on Twitter: Sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance*, 6(4): e21978. <https://doi.org/10.2196/21978>
- [19] Zucco, C., Calabrese, B., Agapito, G., Guzzi, P.H., Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(1): e1333. <https://doi.org/10.1002/widm.1333>
- [20] Jelodar, H., Wang, Y., Orji, R., Huang, S. (2020). Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE Journal of Biomedical and Health Informatics*, 24(10): 2733-2742. <https://doi.org/10.1109/JBHI.2020.3001216>
- [21] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv Preprint arXiv: 2203.05794*. <https://doi.org/10.48550/arXiv.2203.05794>
- [22] Lee, J., Kim, Y., Kwak, E., Park, S. (2021). A study on research trends for gestational diabetes mellitus and breastfeeding: Focusing on text network analysis and topic modeling. *The Journal of Korean Academic Society of Nursing Education*, 27(2): 175-185. <https://doi.org/10.5977/jkasne.2021.27.2.175>
- [23] Weking, J., Mandalenakis, M., Hein, A., Hermes, S., Böhm, M., Krčmar, H. (2020). The impact of blockchain technology on business models-a taxonomy and archetypal patterns. *Electronic Markets*, 30: 285-305. <https://doi.org/10.1007/s12525-019-00386-3>
- [24] Li, G., Xue, J., Li, N., Ivanov, D. (2022). Blockchain-supported business model design, supply chain resilience, and firm performance. *Transportation Research Part E: Logistics and Transportation Review*, 163: 102773. <https://doi.org/10.1016/j.tre.2022.102773>
- [25] Alattar, F., Shaalan, K. (2021). Emerging research topic detection using filtered-lda. *AI*, 2(4): 578-599. <https://doi.org/10.3390/ai2040035>
- [26] Alattar, F., Shaalan, K. (2021). A survey on opinion reason mining and interpreting sentiment variations. *IEEE Access*, 9: 39636-39655. <https://doi.org/10.1109/ACCESS.2021.3063921>
- [27] Shen, X., Wang, L. (2020). Topic evolution and emerging topic analysis based on open source software. *Journal of Data and Information Science*, 5(4): 126-136. <https://doi.org/10.2478/jdis-2020-0033>
- [28] Behpour, S., Mohammadi, M., Albert, M.V., Alam, Z.S., Wang, L., Xiao, T. (2021). Automatic trend detection: Time-biased document clustering. *Knowledge-Based Systems*, 220: 106907. <https://doi.org/10.1016/j.knsys.2021.106907>
- [29] Li, H., Qian, Y., Jiang, Y., Liu, Y., Zhou, F. (2023). A novel label-based multimodal topic model for social media analysis. *Decision Support Systems*, 164: 113863. <https://doi.org/10.1016/j.dss.2022.113863>
- [30] Daud, S., Ullah, M., Rehman, A., Saba, T., Damaševičius, R., Sattar, A. (2023). Topic classification of online news articles using optimized machine learning models. *Computers*, 12(1): 16. <https://doi.org/10.3390/computers12010016>
- [31] Bianchi, F., Terragni, S., Hovy, D. (2020). Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *arXiv Preprint arXiv: 2004.03974*. <https://doi.org/10.48550/arXiv.2004.03974>
- [32] Subbarayudu, Y., Sureshbabu, A. (2024). The detection of community health surveillance using distributed semantic assisted non-negative matrix factorization on topic modeling through sentiment analysis. *Multimedia Tools and Applications*, 1-9. <https://doi.org/10.1007/s11042-024-18321-w>
- [33] Wang, R., Hu, X., Zhou, D., He, Y., Xiong, Y., Ye, C., Xu, H. (2020). Neural topic modeling with bidirectional adversarial training. *arXiv Preprint arXiv: 2004.12331*. <https://doi.org/10.48550/arXiv.2004.12331>
- [34] Subbarayudu, Y., Sureshbabu, A. (2023). The Evaluation of Distributed Topic Models for Recognition of Health-Related Topics in Social Media Through Machine Learning Paradigms. *International Journal of Intelligent Systems and Applications in Engineering*, 11(4), 511-534.
- [35] Rashid, J., Shah, S.M.A., Irtaza, A., Mahmood, T., Nisar, M.W., Shafiq, M., Gardezi, A. (2019). Topic modeling technique for text mining over biomedical text corpora through hybrid inverse documents frequency and fuzzy k-means clustering. *IEEE Access*, 7: 146070-146080. <https://doi.org/10.1109/ACCESS.2019.2944973>
- [36] Rosario, B., Hearst, M.A. (2004). Classifying semantic relations in bioscience texts. In *Proceedings of the 42nd Annual Meeting of The Association for Computational Linguistics (ACL-04)*, pp. 430-437.