Author – Sakshi Singh Tomar

Course – MBA (Business Analytics)

System ID – 2023338299

Co-Author – Dr. Aditi Chaturvedi (Assistant Professor)

TOPIC – DATA MINING IN SOCIAL MEDIA

DECLARATION

I, Sakshi Singh Tomar hereby declare that the research paper named " **DATA MINING IN SOCIAL MEDIA"** is entirely my own work and submitted towards partial fulfillment of requirement of my degree. All sources utilized in this paper have been properly credited, and I have not plagiarized any text. Sharda University's ethical and academic integrity guidelines were followed in the preparation of this study.

ACKNOWLEDGEMENT

I would like to sincerely thank my supervisor, **Dr. Aditi Chaturvedi (Assistant Prof.)**, for their valuable guidance and support throughout the development of this dissertation report. I am also grateful to all the **respondents** who generously shared their time and insights, which were essential to this study. Appreciation is extended to my **fellow students** for their constructive discussions, and to my **family and friends** for their continuous encouragement.

Place: **Signature of the candidate**

Date: **Name: Sakshi Singh Tomar**

TABLE OF CONTENT

S.no.	Content	Page no.
1.	Abstract	5-6
2.	Introduction	6-7
3.	Fundamentals of social media mining	7-8
4.	Techniques and algorithms in social media mining	8-11
5.	Literature Review	11-12
6.	Applications of data mining in social media	12-13
7.	Challenges and Ethical considerations	14-15
8.	Future trends in social media mining	15
9.	Conclusion	15
10.	References	15-17

ABSTRACT

Social media networks, which serve as the opinions, actions, and interactions of billions of users globally, have proliferated as sources of enormous and varied data in the modern digital era. There are possibilities and problems in trying to glean meaningful insights from this mountain of data. Through the use of data mining techniques, organizations, scholars, and legislators can make well-informed judgments, comprehend user behavior, and identify new trends by revealing hidden patterns, trends, and relationships within social media data. The use of data mining in social media analysis is examined in this abstract, which also highlights important approaches, difficulties, and possible advantages.

Finding patterns and gaining knowledge from huge datasets is the goal of data mining, which includes a variety of approaches. These methods can be used with text, photos, videos, and user interactions, among other forms of data in the social media context. Sentiment tracking, trend detection, and opinion mining are made possible by text mining techniques like sentiment analysis and topic modeling, which enable the extraction of insights from textual information. Tasks like object recognition, scene understanding, and visual sentiment analysis are made easier by image and video mining techniques, which allow the analysis of visual input. Through the identification of key users, groups, and information cascades, network analysis techniques reveal the structure and dynamics of social networks.

Nevertheless, there are difficulties in using data mining methods for social media analysis. Strong algorithms and infrastructure are needed for processing and analysis because of the massive volume and velocity of social media data, which creates issues for efficiency and scalability. Additionally, the noisy and unstructured nature of social media data makes feature extraction, data preparation, and interpretation more difficult. Aside from data protection laws, privacy and ethical concerns also need to be taken into account while managing sensitive user information.

The possible advantages of DM for SM analysis are significant, notwithstanding these difficulties. By gaining insightful knowledge on consumer preferences, market trends, and competitor activity, businesses may better inform their product development, marketing plans, and customer engagement initiatives. Social media data analysis allows researchers to better understand public opinion, societal trends, and group behavior. This research advances knowledge in political science, psychology, sociology, and other related subjects. By monitoring

public sentiment, spotting new challenges, and gathering information for policy decisions, policymakers can improve public service delivery and governance.

It also addresses privacy issues and ethical issues related to social media mining, emphasizing the significance of algorithmic openness, informed permission, and responsible data usage. To encourage moral behavior in social media research, techniques for reducing biases, guaranteeing data anonymization, and protecting user privacy are covered.
Ultimately, the study presents practical uses of DM in SM from a range of fields, such as social sciences, marketing, healthcare, and crisis management. Case studies are a good way to show how data-driven insights from social media data can influence society, allow for more focused interventions, and guide decision-making.

Conclusively, data mining techniques provide strong instruments for scrutinizing social media data, permitting the retrieval of significant perspectives and expertise from the extensive and ever-changing digital terrain. Businesses, researchers, and politicians can use data mining to transform how they use social media data for decision-making and society at large by addressing issues with scalability, data quality, and privacy.

INTRODUCTION

With the emergence of social media, the internet has transformed into a dynamic and alive social media landscape where billions of people engage, share, post, and carry out a wide range of daily tasks. Information is gathered, gathered, published, and shared by citizen journalists, and concurrently ingested by hundreds of others who provide impromptu feedback on the journalism. We can connect and communicate with one another anywhere, at any time, thanks to social media, which also gives us a fresh perspective on human behavior and allows us to see it on a never-before-seen scale. Through this social media lens, we have incredible potential to mine human behavioral patterns and comprehend individuals at scale that would otherwise be unachievable. As a result, we may create more individualized computer systems that benefit both people and society by better understanding them. There are no geographical restrictions in this new social media environment, and it is constantly produces vast amounts of data. As a result, we are dealing with an even more severe big data issue: "drowning in data, but thirsty for knowledge." Is it possible for data mining to help?

Regretfully, the traditional data that we are accustomed to in data mining is very different from social media data. The data is mostly user-generated, noisy, and unstructured, with a wealth of social relationships like friendships and followers-followees, in addition to its large quantity. This new kind of data necessitates the use of novel computational data analysis techniques. that may integrate data mining techniques, statistics, and social theories. Due to the urgent need for innovative methods, social media mining is a new interdisciplinary field.

The distinction between the real and virtual worlds is blurred by social media. Now, we can examine how communities (i.e., social molecules) arise and how individuals (i.e., social atoms)

interact by combining social theories with computer techniques. Social media data is unique, necessitating the development of innovative data mining methods that can manage user-generated material with complex social relationships. Under the general heading of data mining, social media mining is a new field that focuses on the investigation and creation of these novel methods. The technique of displaying, evaluating, and identifying useful patterns in social media data is known as social media mining.

A wide range of theories and methodologies are covered by social media mining, including those from computer science, data mining, machine learning, social network analysis, network science, sociology, ethnography, statistics, optimization, and mathematics. Additionally, it introduces basic concepts and important algorithms suitable for evaluating vast quantities of social media data. It comprises the tools needed to formally represent, measure, model, and identify important trends in the massive volumes of social media data.
A new kind of data scientist who is skilled in analyzing recalcitrant social media data and knowledgeable about social and computational theories is created by social media mining, which uses computational tools to help bridge the gap between what we already know (social and computational theories) and what we want to know about the vast world of social media.

FUNDAMENTALS OF SOCIAL MEDIA DATA MINING

According to its structure, social media data can be divided into three groups: semi-structured, unstructured, and structured. Because structured data is well-organized and adheres to a predetermined format, it is simple to analyze with conventional tools like Excel or SQL. User demographics (age, gender, geography), engagement metrics (likes, shares, comments), and ad performance statistics are a few examples. Tables with rows and columns are frequently used to hold this kind of data, making performance monitoring and reporting effective. Although it doesn't follow a strict format, semi-structured data does have some organization. It frequently arrives in JSON or XML forms with tags or metadata for simple extraction. This kind of information can be found in user profiles that mix free-form text with structured fields, or in responses to social media API requests. Last but not least, unstructured data is harder to examine without specialized tools like image recognition or natural language processing (NLP) since it has a predetermined framework. Text postings, comments, photos, and videos are a few examples of content that is free-form and uncategorized. To extract valuable insights from the examination of unstructured data, sophisticated methods like text mining and sentiment analysis are needed. When combined, these data types provide a thorough understanding of social media interactions and user behavior.

Data about social media comes from a variety of sources, each of which offers a distinct perspective on user involvement and activity. One of the main sources are posts, which are user-shared content that includes links, photographs, videos, and status updates. This data provides important insights into the interests and sharing behaviors of people. As direct comments or answers to posts, they offer valuable information on user attitude, public opinion, and content

interaction. Another important element that helps determine the popularity of material is likes, which show user approval or interest in a post. In a similar vein, shares enable users to disseminate material within their personal networks, increasing its visibility and revealing which pieces audiences respond to the best. By classifying or tagging content using hashtags, it becomes simpler to monitor discussions and trends related to particular subjects, occasions, or campaigns. The last important data source is clicks, which show how users interact with links in posts, advertisements, or profiles. Clicks are frequently used to gauge how well content drives engagement or traffic. When combined, these data sources provide a thorough understanding of user interaction with content, facilitating better social media strategy and more informed decision-making.

Each of the data collecting techniques used in social media analytics—web scraping, APIs, and data streaming—offers a different way to obtain information. By mimicking human browsing behavior, web scraping allows for the direct extraction of data from websites. Gathering vast volumes of publicly accessible data from websites or social media platforms, including posts, comments, and user interactions, is its usual purpose. In order to explore and extract structured data from web pages, this method necessitates specific tools. An other popular approach is the use of APIs (Application Programming Interfaces), which offer organized access to data from websites such as Facebook, Instagram, and Twitter. Within the constraints and usage limits of social media platforms, developers can utilize APIs to programmatically query for specific data, including user profiles, postings, or engagement metrics. Data streaming is the practice of gathering data in real time, usually using APIs or specialized tools, that records events and continuous user interactions as they occur. This technique is crucial for following real-time trends, keeping an eye on social media discussions, or collecting information for in-the-moment analysis. When used in tandem, these techniques make it possible to gather enormous, varied datasets from social media, which complement marketing plans, research, and consumer insights.

TECHNIQUES AND ALGORITHMS USED IN SOCIAL MEDIA DATA MINING

Getting valuable information out of vast amounts of unstructured text data is known as text mining. Text analysis and interpretation techniques are used to uncover patterns, trends, and insights from sources such as reviews, comments, and social media posts. Enabling computers to comprehend, interpret, and react to human language is the goal of the artificial intelligence field known as natural language processing, or NLP. Text categorization, sentiment analysis, and language translation are just a few of the tasks made easier by the widespread use of natural language processing (NLP) in social media analysis. NLP is used in social media to transform text-heavy data, like customer opinions or comments, into actionable information that may be used to extract important insights.

Sentiment analysis is an NLP technique that identifies the emotional tone of a string of words. It divides text data into sentiment categories like neutral, negative, and positive. Sentiment analysis is used in social media analysis to determine how the general public feels about events, goods, or

companies. Brands can enhance marketing strategy, customer service, and brand positioning by, for example, examining tweets or Facebook posts to learn how consumers feel about their products.

A method for locating themes or topics in a set of text data is called topic modeling. There are two popular techniques for topic modeling:

1. A probabilistic model known as Latent Dirichlet Allocation (LDA) makes the assumption that every text contains a variety of subjects and that each word in the document is related to a particular topic. It's frequently used to uncover latent theme organization in text corpora, like recognizing various subjects in a group of customer reviews or social media posts.

2. By reducing the dimensionality of the term-document matrix, Latent Semantic Analysis (LSA) identifies patterns in the relationships between terms in the text. Through term co-occurrence analysis, it is utilized to identify underlying themes or subjects in social media interactions.

Social media analysis is using artificial intelligence (AI) and machine learning (ML) more and more to improve and automate a variety of operations. ML algorithms aid in audience segmentation, trend identification, user behavior prediction, and social media strategy optimization. When AI and ML are used together, they can provide even more advanced analysis, including the ability to automatically extract insights from massive datasets, identify new patterns, and even recommend tailored content. Brands utilize AI-driven tools to forecast which content will most effectively connect with their audience, while social media platforms use AI to customize user feeds based on past interactions and preferences.

Supervised vs. Unsupervised Learning

Two main types of machine learning techniques are supervised learning and unsupervised learning.
1. Supervised Learning, which comprises both the input (features) and the right output (labels), is used to train models in supervised learning. This approach is useful for tasks like classification (e.g., categorizing tweets as positive or negative) or regression (predicting sales based on past social media performance). After learning from the labeled data, the model is able to predict new, unseen data.

2. Unsupervised Learning, includes using unlabeled data to train models. By clustering comparable items or lowering the dimensionality of big datasets, the model looks for patterns and structures in the data on its own. Unsupervised learning, for instance, can be applied to find user segments or new subjects in social media discussions without the need for pre-established classifications.

Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks are examples of deep learning models that are effective tools for analyzing social media data, especially when dealing with tasks that require unstructured input like text, photos, and videos.
CNNs, or convolutional neural networks, are commonly employed in image identification tasks,

but they can also be utilized for image-based analysis of social media data, such as identifying items or logos in messages. CNNs can be used to analyze photos in Facebook posts or Instagram posts, which helps organizations monitor trends in visual content or multimedia brand mentions.

Recurrent neural networks (RNNs): RNNs operate best with sequential data, like text or time series, where input order is important. RNNs can be used to forecast user behavior or evaluate time-dependent trends in user interactions on social media, such as tracking the evolution of engagement with a post over time.

Long Short-Term Memory (LSTM): An RNN type designed to retain long-term dependencies, LSTMs are perfect for processing lengthier text sequences. Even in cases where context or prior interactions are crucial, LSTMs can be utilized for sentiment analysis on vast amounts of text (such as tweets or Facebook posts) in social media to ascertain the emotional tone over time or across several posts.

Network Examination

The study of network analysis in social media looks at the connections and exchanges between entities, content, and users. The network's information flow and user connections are examined using graph theory. This approach is crucial for comprehending community structures, social impact, and the spread of information on social media sites like Facebook, Instagram, and Twitter. It makes it possible to identify content that quickly circulates through user networks, new trends, and important influencers.

Graphs of social networks

Network Social Visual depictions of the connections among people or entities in a social media ecosystem are called graphs. Nodes stand in for people or posts, and edges (links) for interactions like retweets, mentions, and friendships. These graphs are useful for locating key players in a network, locating remote communities or groups, and examining the structure of connections. For instance, examining a Twitter network graph can reveal important influencers or groups of people talking about particular subjects or hashtags.

Community detection is the process of locating user or post groups that are more closely related to one another than to other groups. This can be used in social media to identify interest-based groups, such as those talking about particular subjects, companies, or occasions. The Louvain method and modularity maximization are two examples of community discovery algorithms that can identify user groups with similar interests. This information can be used for audience behavior analysis or targeted marketing.

Influence analysis looks at how particular individuals or groups affect the network as a whole. This can entail tracking the dissemination of content, identifying influencers, or examining how particular people affect conversations. Influencer analysis on social media platforms can assist brands in identifying people who have the ability to shape trends and opinions.

Influence analysis examines the ways in which certain people or groups impact the network as a whole. This may involve monitoring content distribution, identifying influencers, or analyzing

the ways in which specific individuals impact discussions. Brands may find people who have the power to influence trends and attitudes by using influencer research on social media platforms.

Hashtag tracking is keeping an eye on how particular hashtags are used and interacted with on various social media sites. One important way to see what subjects or events people are talking about is by looking at hashtags. Brands, researchers, and social media analysts can assess the reach and popularity of specific campaigns or subjects by monitoring hashtags. Additionally, it aids in determining popular subjects and examining the attitude of the general population toward certain hashtags. Tools like Hootsuite, Twitter Analytics, and custom-built solutions with APIs can be used for hashtag tracking.

Finding odd or unexpected shifts in social media activity, like increases or decreases in engagement, sentiment, or topic conversations, is known as anomaly detection in trends. Anomaly detection methods can identify anomalous patterns (such as an abrupt spike in unfavorable sentiment toward a brand or product) that could need urgent attention by examining trends over time. By identifying these outliers, methods like as statistical modeling, machine learning, and time series analysis can facilitate prompt reactions to possible emergencies, viral occurrences, or new opportunities.

LITERATURE REVIEW

The authors of [17] have examined the methods currently employed in social media (SM) analysis. The efficacy of the SM data analysis presented in this paper is so due to data mining's ability to handle dynamic and unstructured data. According to the authors, research on both existing and untested data mining techniques will be conducted in the future in order to mine the data collected on SM.

The author provides an empirical overview of the data mining methods that are currently accessible for social network data mining in [13]. The author claims that future research will concentrate more on content mining, which can identify many patterns of human behavior by examining social network profile pages. A hybrid approach that combines content mining and social network analysis (web structure mining) would also be more beneficial.

The author of [18] has researched the data mining methods currently employed for social media data analysis. Analysis in this study has demonstrated that it is not reasonable to anticipate one system for mining various types of data. As a result, a variety of data mining approaches are accessible for use in many fields. Each data mining method or algorithm has advantages and disadvantages of its own.

After analyzing the survey paper data, the writers in [19] offered some helpful suggestions for managing the characteristics that primarily influence social media. The information a better facility for data control is offered by mining techniques. Data mining techniques facilitate the identification of patterns within the large data set and help identify commonalities between them.

The researchers create a new dimension for themselves to govern the uncontrollable data that exists in social media and social networks based on the results and outcomes they produce.

As per the authors of [2], the most popular social media mining algorithms are SVM, BN, and DT. They also mentioned that the most active domains that require social media data mining are social network data analysis, business, and management. Also the writers advise that the area of social media still calls for extra profound inquiry to house a twin-focus strategy which integrates proper.

A systematic data mining architecture for extracting intellectual content from social data was given by the authors in [20]. They used Facebook as the main source of data for this study, gathering various attributes such as comments, me, wall posts, and age as raw data. They then used sophisticated data mining techniques to extract intellectual knowledge, analyzed their findings, and proposed that social data mining is an intriguing and difficult research method to extract intellectual knowledge that can be applied to decision-making, pattern recognition, social mapping, job responsibility distribution, product promotion, and human behavior prediction.

An overview of support vector uncertainty in SVM was provided by the authors in [21]. The data point and support vector values were known in prior approaches. Assume, if it is uncertain, SVM gets more complicated when it comes to nonlinear kernel selection and object classification. They have therefore proposed that additional study be done to address the uncertainty of data points and nonlinear kernel selection.

APPLICATION OF DATA MINING IN SOCIAL MEDIA

Through the extraction of insightful information from massive volumes of data, data mining plays a critical role in social media. Companies utilize data mining in marketing and advertising to examine user behavior, preferences, and interactions in order to forecast new trends, improve ad placements, and provide tailored recommendations. Businesses may improve the effectiveness of their advertising campaigns by knowing the interests and interaction patterns of their customers.

Sentiment analysis tools use natural language processing (NLP) to categorize opinions as positive, negative, or neutral, allowing businesses to understand customer satisfaction, identify areas for improvement, and respond to feedback effectively. Data mining also helps businesses gauge public opinion by analyzing social media posts, reviews, and comments. This ties into brand reputation management, where businesses monitor online discussions, detect potential crises, and compare their brand perception against competitors. Finding key influencers and evaluating their impact on brand image further improves reputation management strategies.

To analyze online habits and divide audiences into groups according to demographics, interests, and past purchases, targeted advertisements and user profiling mostly rely on data mining. Businesses utilize this data to better target potential clients, optimize marketing expenditures,

and develop customized advertising campaigns. In order to improve engagement and conversion rates, social media networks use complex algorithms to deliver appropriate adverts.

To analyze online habits and divide audiences into groups according to demographics, interests, and past purchases, targeted advertisements and user profiling mostly rely on data mining. Businesses utilize this data to better target potential clients, optimize marketing expenditures, and develop customized advertising campaigns. In order to improve engagement and conversion rates, social media networks use complex algorithms to deliver appropriate adverts.

Detecting false information and fake news is one of the biggest problems in social media nowadays. Through content pattern analysis, source verification, and the use of AI-driven fact-checking models, data mining techniques assist in identifying bogus material. Platforms may stop the spread of false information by taking remedial action when they identify false narratives early. Data mining also helps with political and election analysis by tracking voter mood, assessing campaign effectiveness, and determining how automated bots affect public conversation. These insights are used by analysts and political parties to forecast election results and improve their strategy.

In social media influence and virality studies, it is essential to comprehend how material circulates on social media. Monitoring the spread of information across networks, identifying viral patterns, and identifying the main influencers influencing engagement are all made easier with data mining. Marketers and companies can use this information to improve interaction tactics and produce more powerful content.

Data mining techniques have a substantial positive impact on cybersecurity and fraud detection. Social media activity is analyzed by machine learning models to identify hate speech, phishing efforts, and fraud campaigns. Furthermore, verifying the legitimacy of online interactions requires the ability to spot phony accounts and bots. Social media platforms may preserve their reputation by using data mining technologies to identify odd engagement patterns, examine profile activities, and flag accounts that display bot-like traits.

Beyond corporate and security uses, data mining in social media also significantly improves public awareness and health informatics. Social media sites played a key role in monitoring disease outbreaks during the COVID-19 pandemic by examining posts, search patterns, and exchanges. Researchers evaluated public opinion regarding health policy, forecasted the spread of illnesses, and shared critical health information using this data. Additionally, data mining supports mental health analysis by tracking conversations about stress, anxiety, and depression, which helps medical organizations run focused awareness programs.

CHALLENGES AND ETHICAL CONSIDERATIONS

Data Security & Privacy

In social media analytics, protecting user data's security and privacy is a crucial concern. The California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR)

in the European Union both enforce stringent rules around the gathering, storing, and use of data. It is imperative for organizations to secure explicit agreement from users, guarantee data anonymization, and exhibit transparency about the handling of user information. Significant legal and financial ramifications may arise from breaking these rules.

Fairness and Prejudice in AI Algorithms

Unfair results may arise from AI algorithms employed in social media analytics that inherit biases from the data they are trained on. Inappropriate feature selection, systemic prejudice in the data sources, or skewed training data can all lead to biases. To address this problem and guarantee equal results, fairness-aware algorithms, frequent audits, and a variety of training datasets must be used.

Social Media Data Abuse for Manipulation

Misinformation campaigns, targeted political advertising, and other forms of manipulation can be carried out using data from social media. Analytics are used by bad actors to create misleading narratives that sway public opinion. The use of behavioral data by corporations to manipulate users without their knowledge raises ethical questions. Platforms need tighter regulations, improved fact-checking systems, and more openness in algorithmic decision-making to counter this.

Noise in Social Media Analytics and Data Quality

Unstructured and loud, social media data frequently contains redundant, inaccurate, or irrelevant information. Bots, spam, and phony accounts further deteriorate data quality, which impacts the precision of findings. Natural language processing (NLP) models to identify irregularities, sophisticated filtering techniques, and ongoing validation approaches to enhance data integrity are all necessary to guarantee data trustworthiness.

Problems with Real-time vs. Batch Processing

Real-time and batch processing are two methods for social media analytics, each with unique difficulties. In order to process large amounts of data in real time, high processing power, low latency designs, and effective streaming analytics are needed. On the other hand, batch processing is less resource-intensive but may result in out-of-date information, which makes it less useful for making decisions quickly. Businesses must balance these two strategies according to their infrastructure capabilities and analytical requirements.

For social media analytics to be morally sound, safe, and efficient, these issues must be addressed. In order to gain customers' confidence and get valuable insights from social media data, organizations must use responsible AI methods, adhere to privacy laws, and maintain transparency.

FUTURE TRENDS IN SOCIAL MEDIA MINING

The function of automation and artificial intelligence in social media analysis

Social media analytics are being revolutionized by automation and artificial intelligence (AI), which make sentiment analysis, predictive modeling, and real-time monitoring possible. Large

volumes of data may be processed quickly and effectively by AI-driven solutions, providing greater insights into user trends, behavior, and engagement patterns. AI's contribution to automating data processing will only grow as the technology develops.

The Effect of Blockchain Technology on Data Privacy
A potential way to improve data privacy in social media analytics is through blockchain technology. Blockchain can help consumers recover control over their personal data and guarantee transparency in the sharing of information by offering decentralized and unchangeable records. Moreover, this technology can help lessen illegal access and data breaches.

Advanced Natural Language Processing Models (BERT, GPT) for Social Media Analysis
The ability to evaluate and interpret social media information has greatly increased thanks to **Natural Language Processing** (NLP) models like GPT and BERT. Businesses can gain valuable insights from unstructured text data by using these models to improve sentiment analysis, topic modeling, and language comprehension. The capacity to mine social media will be significantly improved by the ongoing development of NLP technology.

Social Media Analytics Using Ethical AI
Maintaining ethical AI practices is essential as AI is increasingly included into social media analytics. This entails tackling algorithmic bias, preserving decision-making processes' transparency, and giving user consent first priority. The appropriate development and implementation of AI-driven analytics solutions will be greatly aided by ethical AI frameworks.

CONCLUSION

The history, parameters, process, system design, and various data collection methods are all covered in this essay's overview of data mining. system, data gathering techniques, and tactics. Every one of the tactics mentioned has pros and cons of its own. The primary focus of this study is social media mining. Popular social media mining algorithms including SVM, BN, and DT are discussed in this study.
The unique characteristics of social media data—velocity, scale, dynamic, noisy, unstructured, heterogeneous behavior, etc.—inspire researchers to do additional studies on both current and emerging technologies. Hopefully, future study will look at data mining methods, their effects, and emerging research questions.

References

1. Data mining in social media G Barbier, H Liu – Social network data analytics, 2011 - springer (Google scholar)
2. MohammadNoor Injadat, Fadi Salo, Ali Bou Nassif “Data Mining Techniques in Social Media: A Survey”, NEUCOM17295, Volume 214, PP:654-670, 2016
3. Mining social media : a brief introduction P Gundecha, H Liu – New directions in informatics, 2012 – pubsonline.informs.org

4. Overview on data mining in social mediaCA Pushpam, JG jayanthi – International journal of computer , 2017 researchgate.in
5. Network based modeling and intelligent data mining of social media analysis A Akay, Dragomir – IEEE journal of biomedical, 2014 ieeexplore.ieee.org
6. Hemlata Sahu, " A Brief Overview on Data Mining Survey" in International Journal of Computer Technology and Electronics Engineering (IJCTEE)
7. MohammadNoor Injadat, Fadi Salo, Ali Bou Nassif "Data Mining Techniques in Social Media: A Survey", NEUCOM17295 , 2016.
8. Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar" Data Mining Techniques & Distinct Applications: A Literature Review" in International Journal of Engineering Research & Technology (IJERT), 2012.
9. Dinesh Bhardwaj1, Sunil Mahajan2, "ANALYSIS OF DATA MINING TRENDS, APPLICATIONS, BENEFITS AND ISSUES", in International Journal of Computer Science and Communication Engineering, 2016.
10. Dr. Poonam Chaudhary , "Data Mining System, Functionalities and Applications: A Radical Review" in International Journal of Innovations in Engineering and Technology (IJIET), 2015
11. Neelamadhab Padhy1, Dr. Pragnyaban Mishra 2, and Rasmita Panigrahi3, "The Survey of Data Mining Applications And feature scope", in International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), 2012.
12. Smita1, Priti Sharma, " Use of Data Mining in Various Field: A Survey Paper", in IOSR Journal of Computer Engineering (IOSRJCE), 2014.
13. S.G.S Fernando et.al "Empirical Analysis of Data Mining Techniques for Social Network Websites" in COMPUSOFT, An international journal of advanced computer technology, VolumeIII, Issue-II PP:582-592, 2014.
14. Annan Naidu Paidi " Data Mining: Future Trends and Applications" in International Journal of Modern Engineering Research (IJMER)
15. N. Agarwal, H. Liu, S. Subramanya, J. Salerno, and P. Yu. Connecting sparsely distributed similar bloggers. pages 11 –20, dec. 2009.
16. E. Cox. Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration. Elsevier/Morgan Kaufmann, Amsterdam, 2005
17. Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F, "A Survey of Data Mining Techniques for Social Media Analysis" in Journal of Data Mining & Digital Humanities, PP:1-27, 2014.
18. Thabit Zatari, " Data Mining in Social Media " in International Journal of Scientific & Engineering Research, ISSN 2229-5518 Volume 6, Issue 7, PP:152-154, 2015.
19. Dr.B.Umadevi1, P.Surya2, "A Review on Various Data Mining Techniques in Social Media", in International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 4, PP: 8082-8086, 2017.
20. Rahman, M. M, "Mining Social Data to Extract Intellectual Knowledge", in International Journal of Intelligent Systems and Applications(IJISA), vol.4, no.10, PP:15-24, 2012.
21. XimingWang · Panos M. Pardalos," A Survey of Support Vector Machines with Uncertainties", © Springer-Verlag Berlin Heidelberg 2015, Ann. Data. Sci. (2014) 1(3–4) PP:293–309, 2014.
22. MISS. NAZNEENTARANNUM S. H. RIZVI, "A SYSTEMATIC OVERVIEW ON DATA MINING: CONCEPTS AND TECHNIQUES" in International Journal of Research in Computer & Information Technology (IJRCIT), Vol. 1, Special Issue 1, PP:136-139, 2016.
23. Anmol Kumar1, Amit Kumar Tyagi2, Surendra Kumar Tyagi3, "Data Mining: Various Issues and Challenges for Future :A Short discussion on Data Mining issues for future work", in International Journal of Emerging Technology and Advanced Engineering, Volume 4, Special Issue 1, PP:1-8, 2014.

24. Dipti Verma and Rakesh Nashine," Data Mining: Next Generation Challenges and Future Directions" in International Journal of Modeling and Optimization, Vol. 2, No. 5, PP: 603-608, 2012
25. Sagar S. Nikam, " A Comparative Study of Classification Techniques in Data Mining Algorithms" in ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY, Vol. 8, No. (1): PP: 13-19, 2015
26. B.R. Patel, "Comparative analysis of classification algorithm in EDM for improving student performance", International Journal of Computer Sciences and Engineering, Vol.5, Issue.10, pp.171- 175, 2017.
27. Nesma Settouti, Mohammed El Amine Bechar and Mohammed Amine Chikh, "Statistical Comparisons of the Top 10 Algorithms in Data Mining for Classification Task",in International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, No.1, PP:46-51, 2016
28. GemaBello-Orgaza,JasonJ.Jungb,∗,DavidCamacho, "Social bigdata: Recent achievements and new challenges", http://dx.doi.org/10.1016/ j.inffus.2015.08.005 1566-2535/© 2015 Elsevier.
29. Shweta Verma, Vivek Badhe, " Survey on Big Data and Mining Algorithm", IN IJSRSET, , Volume 2 | Issue 2 | , PP: 1338-1344, 2016.
30. Dr.M.Chidambaram, R.Umasundari, "A Survey on Feature Selection in Data Mining", in International Journal of Innovative Research in Computer Science & Technology (IJIRCST) Volume4, Issue-1, PP: 13 - 14, 2016
31. Parmeet Kaur, " AN OVERVIEW OF DATA MINING TOOLS", in International Journal of Engineering Applied Sciences and Technology, Vol. 1, Issue 6, PP: 41-46, 2016.
32. Wei Fan "Mining Big Data: Current Status, and Forecast to the Future" SIGKDD Explorations Volume 14, Issue 2, PP:1-5
33.] H. K. Chan1, E. Lacka2, R. W. Y. Yee3, M. K. Lim4, "A Case Study on Mining Social Media Data", in the Proceedings of the 2014 IEEE IEEM, 978-1-4799-6410-9/14/$31.00 ©2014 IEEE , PP: 593- 596
34.] David Jensen and Jennifer Neville, "Data Mining in Social Networks", Papers of the Symposium on Dynamic Social Network Modeling and Analysis. National Academy of Sciences,Washington, DC: National Academy Press. PP:1-13, 2002.