

Detection of Knowledge on Social Media Using Data Mining Techniques

Aseel Abdullah Alolayan, Ahmad A. Alhamed

Department of Information Systems, King Saud University, Riyadh, Saudi Arabia

Email: 445205825@student.ksu.edu.sa, aalhamed@ksu.edu.sa

How to cite this paper: Alolayan, A.A. and Alhamed, A.A. (2024) Detection of Knowledge on Social Media Using Data Mining Techniques. *Open Journal of Applied Sciences*, 14, 472-482.

<https://doi.org/10.4236/ojapps.2024.142034>

Received: November 18, 2023

Accepted: February 26, 2024

Published: February 29, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In light of the rapid growth and development of social media, it has become the focus of interest in many different scientific fields. They seek to extract useful information from it, and this is called (knowledge), such as extracting information related to people's behaviors and interactions to analyze feelings or understand the behavior of users or groups, and many others. This extracted knowledge has a very important role in decision-making, creating and improving marketing objectives and competitive advantage, monitoring events, whether political or economic, and development in all fields. Therefore, to extract this knowledge, we need to analyze the vast amount of data found within social media using the most popular data mining techniques and applications related to social media sites.

Keywords

Data Mining, Knowledge, Data Mining Techniques, Social Media

1. Introduction

Without a question, the tangible impact of social media is causing the globe to become a little village. It brings together individuals of all ages, races, and nationalities and enables them to communicate and share ideas, memories, and sentiments as well as photos, videos, and interests. This has made it possible for companies across all sectors to market, gain from, analyze, learn from, and improve their businesses using the information offered by social media.

The social media data is disorganized and presented in a variety of formats, including text, voice, photos, and videos. Additionally, the social media platforms offer a vast amount of continuous real-time data, rendering conventional statistical approaches inappropriate for analyzing this vast amount of data [1].

Social media become is extremely important to our daily lives. The most

popular social media platforms for connecting are websites like Facebook, Twitter, WhatsApp, and Instagram, many businesses Try to using amounts of social media data for capitalize on this social phenomenon to serve its interests [2].

Today's information and communication technology is evolving quickly. According to statista.com, the number of social media users worldwide reached 4.8 bn as of April 2023 [3]. This indicates the very rapid growth in the use of social media applications. Huge amounts of data are now available in various formats, including text, video, music, photos, and graphics, thanks to apps and social media users.

These data are kept in several kinds of repositories. The human race is afflicted by a new condition known as "Data Rich and Information Poor" as a result of this data influx. Data retrieval is not adequate to maximize the utilization of this resource. The actual difficulties faced by researchers are in summarizing, analyzing, extracting the information, and finding the pattern and link among the data. The answer to all of these problems is "Data Mining" [4].

Therefore studying the activities, interests, behaviors, and opinions of users by examining the posts published on social media platforms can be done to extract Knowledge. As a result, new machine learning applications are specified for extracting useful knowledge in different application fields in a variety of application areas, such as trend identification, social media analytics, pattern mining, sentiment analysis, and opinion mining [5].

Knowledge has evolved into a key source for many firms in today's competitive world [6].

Therefore, data mining techniques very useful in resolving the issues of discover valuable, accurate and useful knowledge from social media data for any sector needs to reach its goals, through the huge amounts of data around it.

2. Data Mining

Data mining is the use of algorithms to identify patterns in data and to provide knowledge that can be used to make decisions [7].

such as if a commercial company wants to advertise a specific product and increase its sales, it is possible to benefit from its customer data to find a pattern, classification, or behavior for them through the use of data mining..

Therefore, data mining can be used to predict this information [8].

Data mining refers to the discovery of knowledge in terms of patterns or rules from vast amounts of data, or the process of analyzing and extracting knowledge from data using computer learning techniques.

The process of extracting a handful of valuable nuggets from a huge amount of raw material is known as mining (**Figure 1**).

Many other phrases have meanings that are similar to or slightly different from data mining, such as "knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging".

Data mining is used as a synonym with another widely used term "Knowledge

Discovery from Data, or KDD”.

Others, on the other present, describe data mining as only a necessary stage in the knowledge discovery process.

Figure 2 shows the process of knowledge discovery, and includes an iterative series of the following steps:

- 1) Data cleaning (To eliminate confusing and noisy data);
- 2) Data integration (allowing the combining of data from many sources);
- 3) Data selection (where the database is searched for data related to the analyzing task);
- 4) Data transformation (where data are consolidated or turned into mining-ready formats by executing summary or aggregation procedures, for example);
- 5) Data mining (a basic step in which intelligent techniques are used to extract data patterns);
- 6) Pattern evaluation (based on various criteria, to discover the truly interesting patterns that represent knowledge);
- 7) Knowledge presentation (Here, the user is given the mined knowledge using visualization and knowledge representation techniques).

This viewpoint is that although data mining is an important phase in the process because it reveals hidden patterns for review, it is only one step in the whole process. We concur that one step in the knowledge discovery process is data mining. The longer-term of knowledge discovery from data is losing ground to the term “data mining” in the industry, the media, and the database research environment [9].

The useful information extracted from using data mining techniques, can be applied in different fields appropriate decisions. In terms of technology, discovering

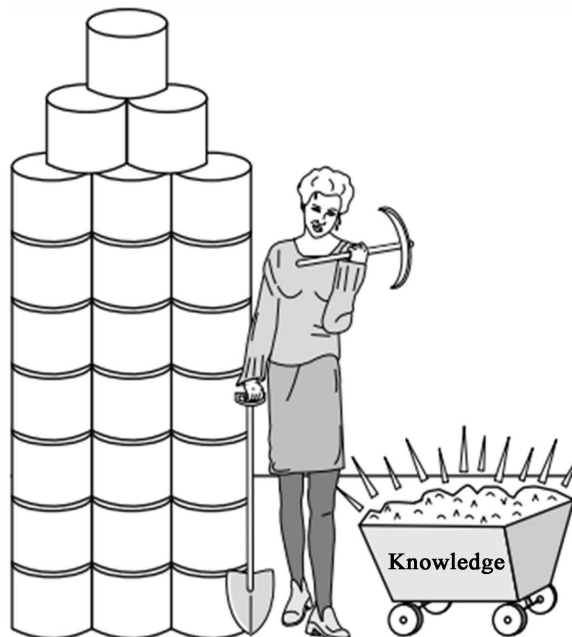


Figure 1. Data mining - looking for knowledge (interesting patterns) in data.

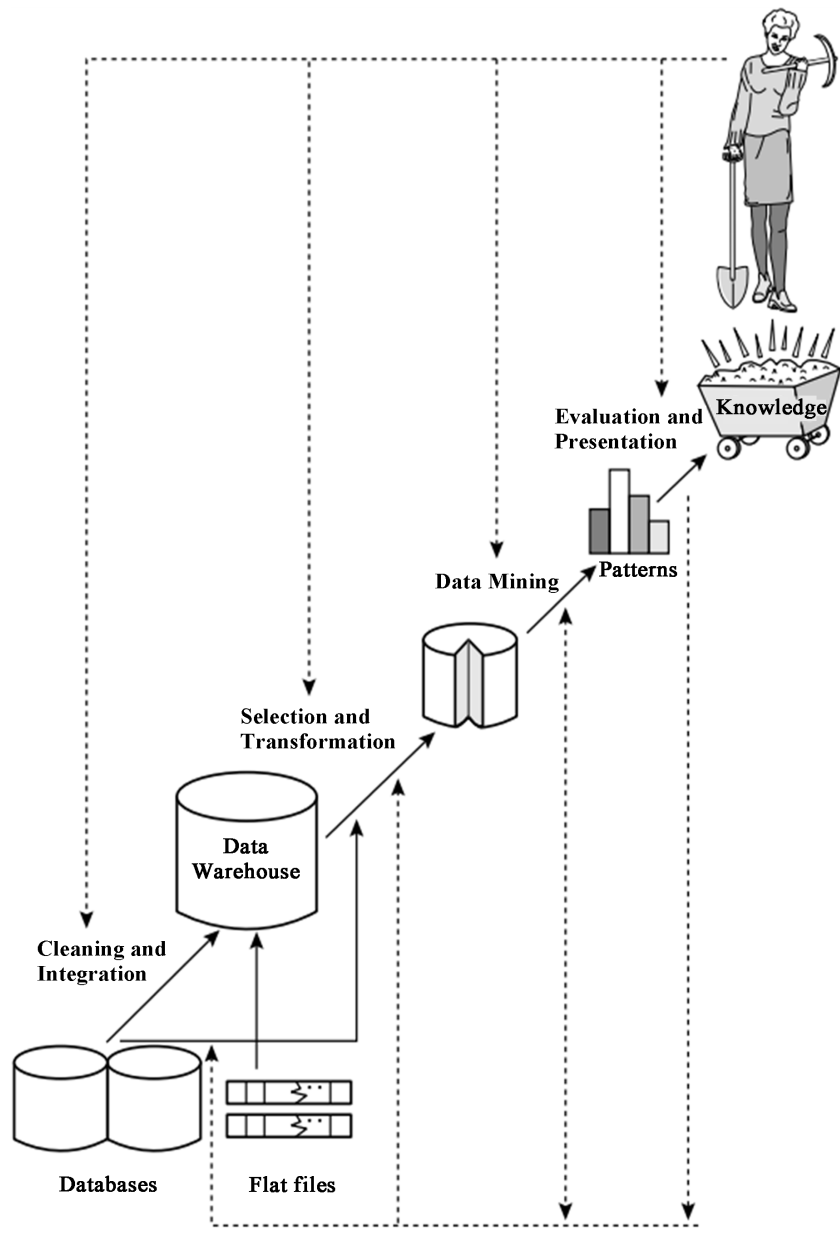


Figure 2. Data mining as a step in the process of knowledge discovery.

patterns or correlations in sizable relational databases is known as data mining, using approaches intersection of artificial intelligence, statistics, machine learning and database systems [4].

3. Data Mining on Social Media

What are social media?

Social media are applications that require an Internet connection to access and allow users to communicate and interact with each other through user-generated content [10]. Communication is important in our daily lives. Communicate has developed and changed in this era and has extended to digital communication,

social media platforms dominate the digital space.

The important social media development is the widespread adoption of social networking platforms that are focused on commerce such as Facebook, Instagram, LinkedIn, Pinterest, Flickr Tumblr, Twitter and many more [11].

Data mining in Social Media, It uses raw social media data to analyze and extracts patterns, correlations, and trends from it.

For example, if a specific entity, whether a company private or Public wants to design its strategies or provide new products or services, it uses data mining technology on social media to access online behaviors', content sharing, interpersonal communication, online purchasing behavior, etc [4].

Users of social networking sites can communicate with people in their network by sharing thoughts, digital images, videos, posts, and information on activities and events taking place online or in the real world. Members might be able to get in touch with any other members, depending on the social networking platform. In other instances, members can get in touch with anyone they are connected to, and then anyone that connection is connected to, and so forth [2].

Figure 3 shows on depicting the top 15 social networking sites and apps as calculated by industry-leading provider of business data, Statista [12].

Social networking apps are going to grow rapidly even bigger as people adopt them into their everyday lives.

Large social media data sets can be mined using data mining techniques, which has the potential to further enhance search results for common search engines, enable targeted marketing for businesses, aid psychologists in their study of behavior, give sociologists new insights into social structure, personalize web services for users, and even assist all of us in identifying and preventing spam. Additionally, the open access to data offers researchers a wealth of knowledge never before available to optimize data mining methods and enhance performance. Social media is an attractive data source for developing and testing new

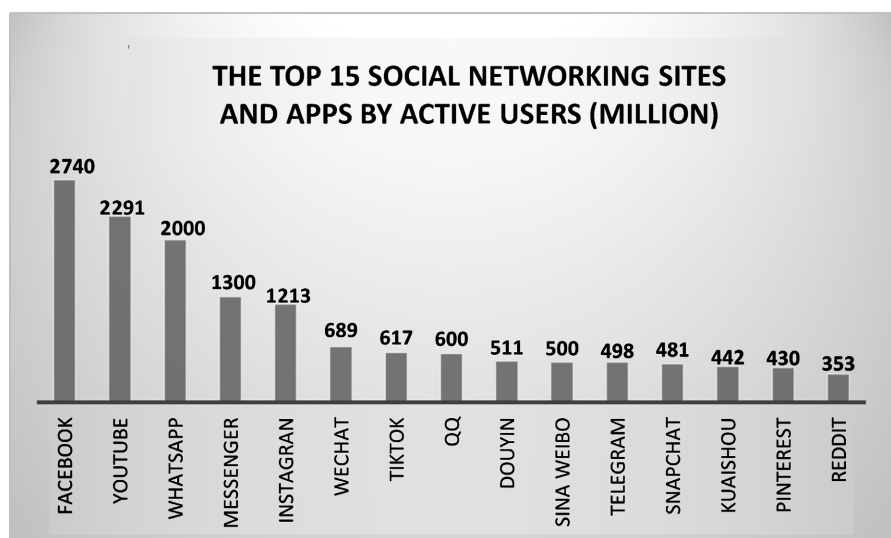


Figure 3. The 15 Biggest Social Media Sites and Apps [2023].

data mining techniques, the advancement of the data mining field itself depends on large data sets, the opportunity to learn how a person's position in the network affects everything from their tastes to their moods to their health is one of the motivating elements for data mining social networking sites [2].

Extracting Knowledge from Facebook

The 15 Biggest Social Media Sites and Apps in 2023 are shown in (Figure 3), it has been shown that among other applications, the Facebook application has the highest usage. Therefore, I would like to talk in this paper about extracting knowledge from this application.

Users of social media platforms can engage with one another using a variety of methods, including chats, forums, comments, and more. Important knowledge is shared and learned by users as a result. The material on these social networking sites can simply be characterized as hazy and unstructured. Spelling, grammar, and sentence structure are typically disregarded in everyday conversations. This could lead to a variety of ambiguities, which makes analyzing and extracting data patterns from big datasets challenging. Therefore, should analyze Facebook text data in an effort to uncover valuable information from it and present it in various ways. An example of this is in one of the studies, 3815 posts were taken and examined from 16 news channels' Facebook pages. On the data gathered, various text mining techniques were used. Findings showed that Fox News is the news outlet with the most Facebook postings shared, followed by CNN and ABC News, in that order [13].

In another study to the detection of fake Facebook profiles using data-mining techniques. It is necessary to make a few essential points clear. In order for the model to decide, it first relies on the informative attributes.

In Figure 4, these attributes are depicted. As can be observed, the "mutual friends" attribute is the most informative while the "introduction" attribute is the least informative. The second thing we observe is that several properties were

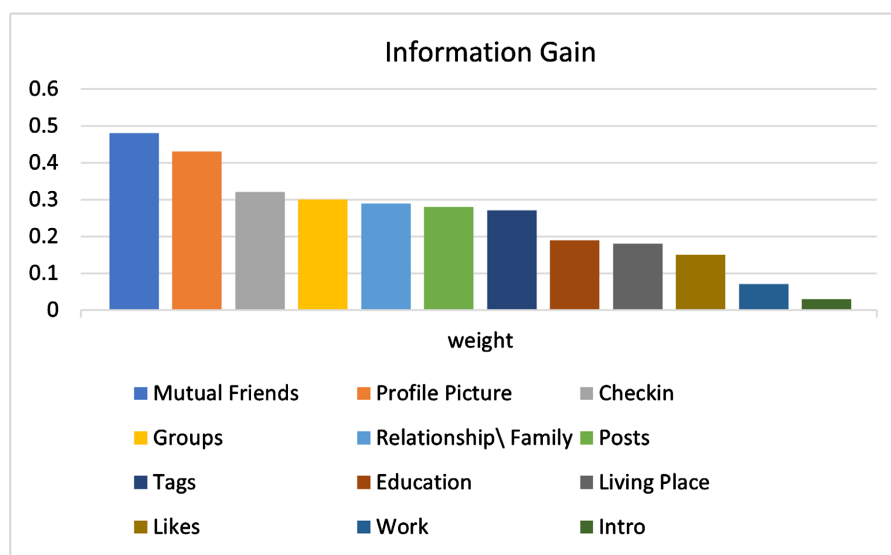


Figure 4. Information gain of the attributes.

the same in both authentic and fraudulent profiles.

For example, fake profiles frequently have no tags, no posts, and a lot of liking activity. Unfortunately, a large number of real profiles have the same values, which confuses the classification methods.

Rapid Miner Studio 8.0.1 was used to run an experiment to assess the ability to detect phony Facebook profiles. Based on a dataset of 982 profiles (781 true and 201 fraudulent), the model's accuracy is determined. In every trial, the supervised algorithms outperformed better than the unsupervised algorithms, More Particularly, the ID3 decision tree showed the most accuracy among all methods.

Shows the histogram chart for the interfering attributes in relation to the two classes (fake and real).The supervised algorithms use the k-NN estimator, which results in a high accuracy rate.

In **Figure 5** we observe that the “groups” and “likes” attributes, which had the highest interference factor, contained the majority of the missing information. By excluding these attributes from the calculation, k-NN increases accuracy rate [14].

The use of data mining techniques, such as visualization, classification, clustering, word clouds, and information retrieval, using data taken from newspaper

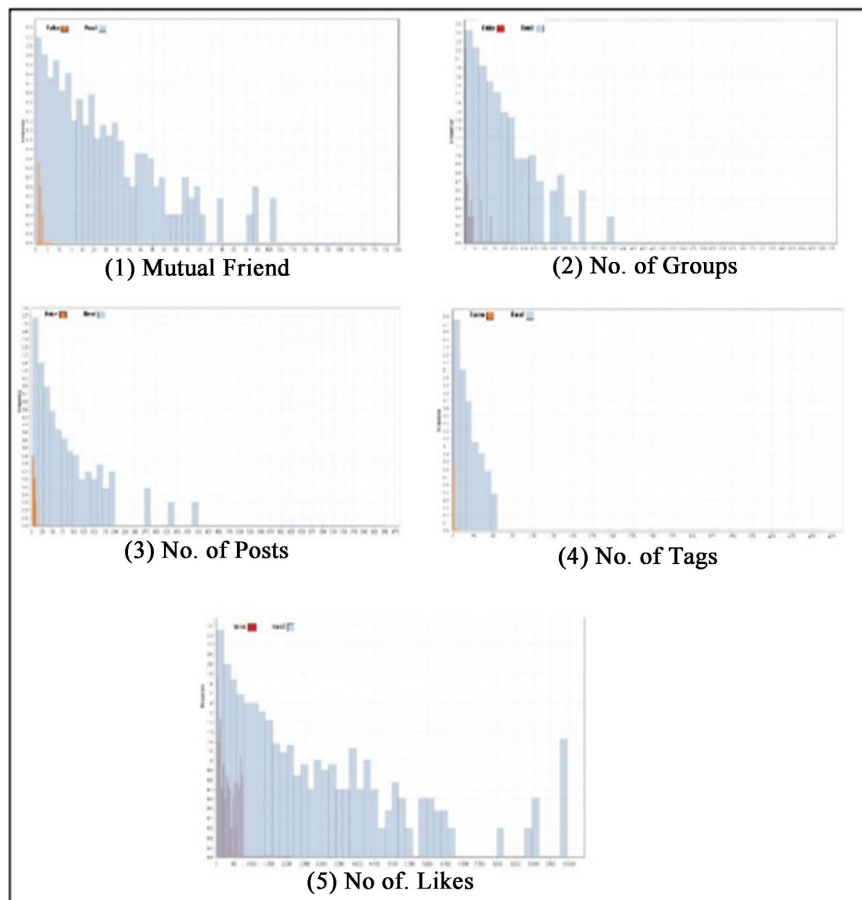


Figure 5. Attributes distribution.

Facebook pages is also presented in one of the other research.

The relationship between the vast amounts of unstructured text that is currently available and the analysis and interpretation of human languages are two of the hottest subjects in NLP.

Despite the growth of Internet Arabic users, systems for naturally evaluating Arabic digital assets are not as readily available as they are for analyzing English. There are technologies that aim to find fascinating knowledge and offer it in various forms.

62327 postings from 24 Arab Gulf newspapers' Facebook pages were looked into and evaluated. The results showed that the terms that are most frequently associated across all of the newspapers are "Allah الله" followed by "Emirates امارات", "Year عام", "Good خي", "Save حفظ", "Blessed مبارك", and "Graces نعم", "Happy سعيد", "Peace سلام", and "live عاشت".

Additionally, Albayan News (UAE) is the newspaper that shares Facebook posts the most, followed by Alshabiba (Oman), Alkhaleej (UAE), and Emarat Alyoum (UAE). The UAE is the nation that shares posts on Facebook the most, followed by Oman and KSA.

These results appeared using data mining techniques text mining continues to receive more research attention [15].

4. Data Mining Techniques

Several data mining techniques, including classification, pattern discovery, summarization and rule discovery can be used to extract knowledge [16].

There are various types of data Mining techniques:

Characterization, Classification, Regression, Association, Clustering, Change Detection, Deviation Detection, Link Analysis and Sequential Pattern Mining.

The use of social networks in business can be seen in a variety of areas, including co-innovation, customer service, general advertising, expanding spoken advertising, marketing research, plan generation and new development, publicity, employee communication, and reputation management [2].

On the (Figure 6) shows some of the data mining techniques that are used for social media by researchers, according to this article the SVM, BN, and DT are the most applied algorithm in the area of social media [1].

Table 1 presents a summary of the advantages and limitations of data mining algorithms [4].

The most common data mining applications related to social networking sites include:

- **Group detection** - Finding and recognizing a group is one of the most well-liked data mining uses for social networking sites. In general, group detection used on social networking sites is based on examining the network's structure and identifying people who interact with one another more frequently than they do with other users. Knowing which organizations a person is a part of might provide insights into that person, such as what kinds of

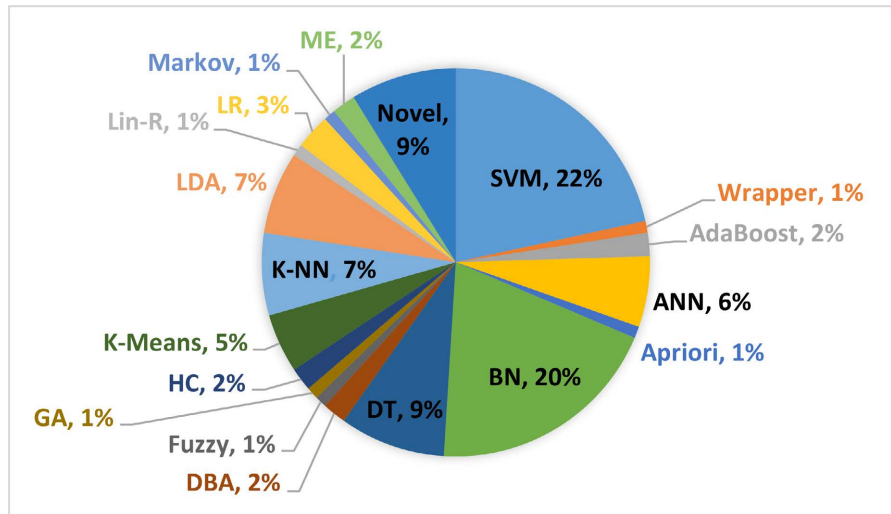


Figure 6. Data mining algorithms in social media.

Table 1. Summary of the advantages and limitations of data mining algorithms.

| S. No | DM Algorithm | Technique used | No. of Papers Implemented | Merits | Limitations |
|-------|-------------------------|----------------|---------------------------|---|---|
| 1 | Support Vector Machines | Classification | 2, 15, 16, 18, 21, 25 | Very accurate, less over fitting, robust to noise. | Binary classifier, In multi-class classification, kernel selection and interpretability are some weaknesses of SVM, Computationally expensive, runs slow. |
| 2 | Bayesian Networks | Classification | 2, 15, 16 | Missing data entries can be handled successfully, Over-fitting of data is avoidable | Quality and extent of prior knowledge play an important role, Significant computational cost |
| 3 | Decision Tree | Classification | 2, 15, 16, 18 | Easy to understand, Easy to generate rules | Over fitting, does not handle easily non numeric data, can be quite large pruning is necessary |
| 4 | C4.5 | Classification | 15, 16, 25 | Quite fast, Output is human readable. | Small variation in data can lead to different decision tree, does not work very well on small training data set, Over fitting |
| 5 | K-Nearest Neighbor | Classification | 2, 15, 16, 18, 25 | Ease of understanding and implementation, depending on the distance metric, KNN can be quite accurate | Computationally expensive Noisy data can throw off kNN classifications. kNN generally requires greater storage requirements than eager classifiers, Selecting a good distance metric is crucial to kNN's accuracy |
| 6 | K-Means | Clustering | 2, 15, 16, 18 | Faster and more efficient especially over large datasets | Sensitive to outliers and the initial choice of centroids, It is designed to operate on continuous data—extra tricks are needed to work on discrete data |

activities, products, and services they could find interesting.

- **Group profiling** - Since there are millions of groups on social media websites, it is not realistic to try to hand respond to each group's question manually. It is helpful to be able to automatically profile a group for a variety of reasons, from simply scientific ones to targeted marketing of products, ser-

vices, and concepts.

Recommendation systems - A recommendation system analyzes social networking data and makes suggestions for users to join new groups or make new friends. The ability to recommend group membership to an individual is advantageous for a group that would like to have additional members and can be helpful to an individual who is looking to find other individuals or a group of people with similar interests or goals. Without an automated system, dealing with such a big number of people and groups would be nearly impossible. Also, group characteristics change over time. For those reasons, data mining algorithms drive the inherent recommendations made to users. The automated recommendations on social networking sites, which enable users to quickly build and grow an online social network with little effort on their side, are directly responsible for a large portion of its attractiveness [2].

5. Conclusion

This paper discussed the great need for the science of data mining, its importance, and the use and development of its techniques to benefit as much as possible from the huge amount of data that exists around us in social media by extracting knowledge that contributes to the development process in all fields, Data mining extracts knowledge and analyzes data patterns. The paper described that data mining is only a necessary stage in the knowledge discovery process, and touched on the types of techniques used in data mining and social media mining techniques. Through this research, it became clear that the Facebook application has the highest usage rate among other applications during the year 2023, and this paper showed the techniques and algorithms used to extract knowledge from this application and some examples of them.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Injadat, M., Salo, F. and Nassif, A.B. (2016) Data Mining Techniques in Social Media: A Survey. *Neurocomputing*, **214**, 654-670. <https://doi.org/10.1016/j.neucom.2016.06.045>
- [2] Martyniuk, H., Kozlovskiy, V., Lazarenko, S. and Balanyuk, Y. (2021) Data Mining Technics and Cyber Hygiene Behaviors in Social Media. *South Florida Journal of Development*, **2**, 2503-2515. <https://doi.org/10.46932/sfjdv2n2-108>
- [3] Number of Internet and Social Media Users Worldwide. <https://www.statista.com/statistics/617136/digital-population-worldwide/>
- [4] Pushpam, C.A. and Jayanthi, J.G. (2017) Overview on Data Mining in Social Media. *International Journal of Computer Sciences and Engineering*, **5**, 147-157. <https://doi.org/10.26438/ijcse/v5i11.147157>
- [5] Belcastro, L., Cantini, R. and Marozzo, F. (2022) Knowledge Discovery from Large

- Amounts of Social Media Data. *Applied Sciences*, **12**, Article 1209.
<https://doi.org/10.3390/app12031209>
- [6] Zarei, E. and Jabbarzadeh, A. (2019) Knowledge Management and Social Media: A Scientometrics Survey. *International Journal of Data and Network Science*, **3**, 359-378.
<https://doi.org/10.5267/j.ijdns.2019.2.008>
- [7] Riadi, I. (2017) Detection of Cyberbullying on Social Media Using Data Mining Techniques. *International Journal of Computer Science and Information Security (IJCSIS)*, **15**, 2, 3, 5, 15.
- [8] Prasdika, P. and Sugiantoro, B. (2018) A Review Paper on Big Data and Data Mining Concepts and Techniques. *International Journal on Informatics for Development*, **7**, 36-38. <https://doi.org/10.14421/ijid.2018.07107>
- [9] Han, J., Pei, J. and Tong, H. (2022) Data Mining: Concepts and Techniques. Elsevier, Amsterdam.
https://books.google.com.eg/books?hl=ar&lr=&id=NR1oEAAAQBAI&oi=fnd&pg=PP1&dq=info:TCHftT_XL_UJ:scholar.google.com/&ots=N1KTjugqY&sig=oxGld6EKTh2P7kFeCISU82h1rtk&redir_esc=y#v=onepage&q&f=false
- [10] Carr, C.T. and Hayes, R.A. (2015) Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication*, **23**, 46-65.
<https://doi.org/10.1080/15456870.2015.972282>
- [11] Niguidula, J.D. and Lacasandile, A.D. (2017) Social Trends: The Theory, Research and Sociality Discerned through Twitter. *International Journal of Computer Science and Information Security (IJCSIS)*, **15**, 14-15.
- [12] Kallas (2023) The 15 Biggest Social Media Sites and Apps in 2024.
<https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/>
- [13] Salloum, S.A., Al-Emran, M. and Shaalan, K. (2017) Mining Social Media Text: Extracting Knowledge from Facebook. *International Journal of Computing and Digital Systems*, **6**, 73-81. <https://doi.org/10.12785/IJCDS/060203>
- [14] Albayati, M.B. and Altamimi, A.M. (2019) Identifying Fake Facebook Profiles Using Data Mining Techniques. *Journal of ICT Research & Applications*, **13**, 2, 3, 5, 13, 15.
<https://doi.org/10.5614/itbj.ict.res.appl.2019.13.2.2>
- [15] Salloum, S.A., Mhamdi, C., Al-Emran, M. and Shaalan, K. (2017) Analysis and Classification of Arabic Newspapers' Facebook Pages Using Text Mining Techniques. *International Journal of Information Technology and Language Studies*, **1**, 8-17.
- [16] Bello-Orgaz, G., Jung, J.J. and Camacho, D. (2016) Social Big Data: Recent Achievements and New Challenges. *Information Fusion*, **28**, 45-59.
<https://doi.org/10.1016/j.inffus.2015.08.005>