

Concepts and applications of data mining and analysis of social networks

Azam hajiaghajani ¹

¹ Assistant Professor of economy, Department of Management, Chalous Branch, Islamic Azad University, Chalous, Iran;

ARTICLE INFO	ABSTRACT
Received: 11 December 2022 Reviewed: 8 December 2022 Revised: 20 January 2023 Accepted: 29 January 2023	Social media has become an important reference for information during the last few decades. They have been able to be effective in various fields such as business, entertainment, science, crisis management, politics, etc. For this reason, a social media analysis has become very important for researchers and large companies. The widespread use of social media leads to a complex problem called "accumulation of data". Many data science specialists seek to analyze this data in order to identify the behavioral characteristics of users, analyze interests and needs, and improve marketing processes. Different social media platforms have the ability to use all kinds of media, including text data, video, video, audio, and location information, etc. Therefore, data analysis in social networks is very important. In this research, the concepts and applications of data analysis in social networks will be investigated.
Keywords: <i>Data mining, social networks, big data analysis, data in social network</i>	

1. Introduction

The most important reason for the popularity of social media is the possibility of sharing public content at the lowest cost at any time and place. Today, almost everyone uses social media in their daily life. This issue has caused the production of massive amounts of data. Social media data mining is a process for using big data of content generated by users on social media sites and applications such as Instagram, Facebook, Twitter and many other programs. This work is done in order to extract patterns, draw conclusions about users, and also obtain information about them. The application of data mining in the analysis of social networks is actually the same as the mining process to obtain rare minerals (Nahr et al., 2021).

Social media data mining requires experts and data analysts as well as automated software programs to explore and collect raw social media data. Then identify the patterns and process related to the use of social media, online behaviors, content sharing, communication between people, online shopping behavior and many other things. Such models and processes are of interest to many companies, governments and many private organizations, and this helps organizations to design their strategies or introduce new programs, new products, processes or other services (Aliahmadi et al., 2016). Data mining

¹ Corresponding Author: hajiaghajani.azam@gmail.com

of social networks utilizes a wide range of fundamental concepts including computer science, data mining, machine learning, as well as statistics. This work is done by social media data miner by creating suitable algorithms to examine large amount of social network data files (Eisapour et al., 2013).

Data mining of social networks is based on theories and methods from social network analysis, network science, sociology, ethnography, optimization and mathematics. This science includes formal tools for displaying, measuring, and modeling meaningful patterns from large-scale social media data. Now that the speed of receiving this information has increased so much, its amount should be increased accordingly (Barzegar Keliji et al., 2021). For this reason, users usually expand their interests and become members of special groups and read their special and favorite content. This issue shows the need for a more active society in the social field, and the analysis of data collected from social networks can also show a more correct and transparent path to the audience (Ghahremani-Nahr et al., 2021).

Big data analysis from various social networks can be very effective in discovering the interaction pattern of users. Therefore, in this study, we have introduced big data and its impact in various fields, and then we have looked at its role in various social networks. In this article, our goal is to assess the value of big data in various social networks. By examining issues such as trend analysis and user behavior analysis during important events in social networks, the importance of this issue can be shown (Pishkar et al., 2021).

2. All kinds of information in social networks

Social media data mining is used in various industries including business development, social sciences, health care and educational purposes. Once the data obtained from social media analytics is data mined, it can be used in a variety of contexts. Often, companies use a communication model such as correlation to check the social similarity of users with other users (Shayan Nia & Mirataollahi Olya, 2021).

Also, organizations and businesses use user sentiment analysis to improve their satisfaction. Because users often publish positive or negative emotions in their posts. The owners of the networks, both virtual and real, carefully collect and keep this information. The methods of using and even generating income from this information are very wide and profitable. The most common types of these data are:

- ***Feedback***

This is the most common information among all social networks. Almost all social networks provide or find ways to collect opinions, criticisms and suggestions from their users. This information is so important for the continuation and expansion of the activity of all these networks that it is even protected in many cases (Nozari and sadeghi, 2021).

In many cases the network in question implements a system to collect this feedback to allow everyone to share their opinion directly with them. But this is not always possible. The most used to collect users' opinions is to use other social networks and collect their activities about their own network. This work usually has a higher implementation cost than the first method, but at the same time it is more useful and accurate. Many big and fundamental decisions of companies are made based on this information and feedback statistics (Bayanati et al., 2022).

Of course, all these feedbacks are not always only on paper or in virtual form. In large manufacturing companies, another method is also common, during which product sales information and the percentage of users' satisfaction with them are collected by statistical intermediary companies and provided to the manufacturing company (Nozari et al., 2023). This information, which includes even the customers'

conversations with the sellers and their mindset, up to the amount of sales of each product in each shop or region and comparing their performance with competing companies, is the key to the success and progress of each organization and company in its own product (Shayannia , 2022).

- ***Reaction***

The reactions of users in any social network include almost the largest part of the information of that network. These reactions can be about personal issues, the events, and happenings of the day, or even the background of the events that are going to happen in the future. Social networks carefully collect and classify this information. This information can be used to generate income, learn about people's behavior, and social psychology, prepare for changes, etc. In this section, social events play a bigger role and show organizations and even governments what the users and, as a result, people think (Nozari et al., 2021).

Of course, an important problem is the great variety of this type of data and its difficult classification. Social networks, especially virtual ones, usually think of solutions to extract this information more easily. In many cases, this information has even been observed to be sold to companies at high prices. Surveys are one of the simplest forms of extracting this information, which, depending on the number of participants and their formality, can contain very important information. Again, as before, networks that do not have direct access to their users extract and report this information from other networks (Aliahmadi et al., 2022).

- ***Trend***

This type of data is one of the newest types of information that is much easier to obtain than other information. In this type of data, social networks try to see what is the most popular topic or even make a specific topic popular. For this reason, in a short period of time, for example, a month or even a day, a so-called hot topic has become popular and a large number of network users are commenting on it. This information, which is considered a kind of public survey, can be very useful, even if the information itself is not used, it determines the interest and popularity of that topic, news, person or product (Tavakkoli-Moghaddam et al., 2022).

This type of data is especially useful for presenting companies' products or services and seeing people's reactions to them. This information has also played a very colorful role in the stock market and it has happened a lot that companies have gone bankrupt or reached the top because of this information (Salehi Koocheh Baghi et al., 2021).

Trends are not necessarily about products and news. Many times they even include a tourist city or even a controversial question. Twitter and Instagram microblog social networks use these types of trends a lot and try to encourage their users to be more and more active. Likewise, by building appropriate tools, in this way, they expand and improve the search in their network and the communication of users (Moeini et al., 2013).

- ***Personal Information***

This case, which is mostly used in virtual social networks, includes the collection of personal information of users such as name, job, place of residence, education, personal photos and even private relationships of users. In the simplest cases, this information is used only for statistics of network users and in the beyond society (Najafi et al., 2022).

Regarding the use of this information, one should be very careful not to violate people's privacy. Almost all social networks follow a specific agreement and rules for using this data. Most of these laws prohibit

them from using or selling a specific person's information. Of course, many cases have been observed that the social network has made itself the full owner of this information. But basically, this information is used anonymously so that it does not cause any problems for the user or the network (Aliahmadi et al., 2022).

The age group of users of a network, gender, hours and amount of activity of users and even their type of activity provide good information especially for companies, sociologists and psychologists. This information is usually used in advertising systems, or in most researches, it is published once in a while and the results are made known to the public (Nozari et al., 2021). Figure 1, Taxonomy of social media data management. is showing.

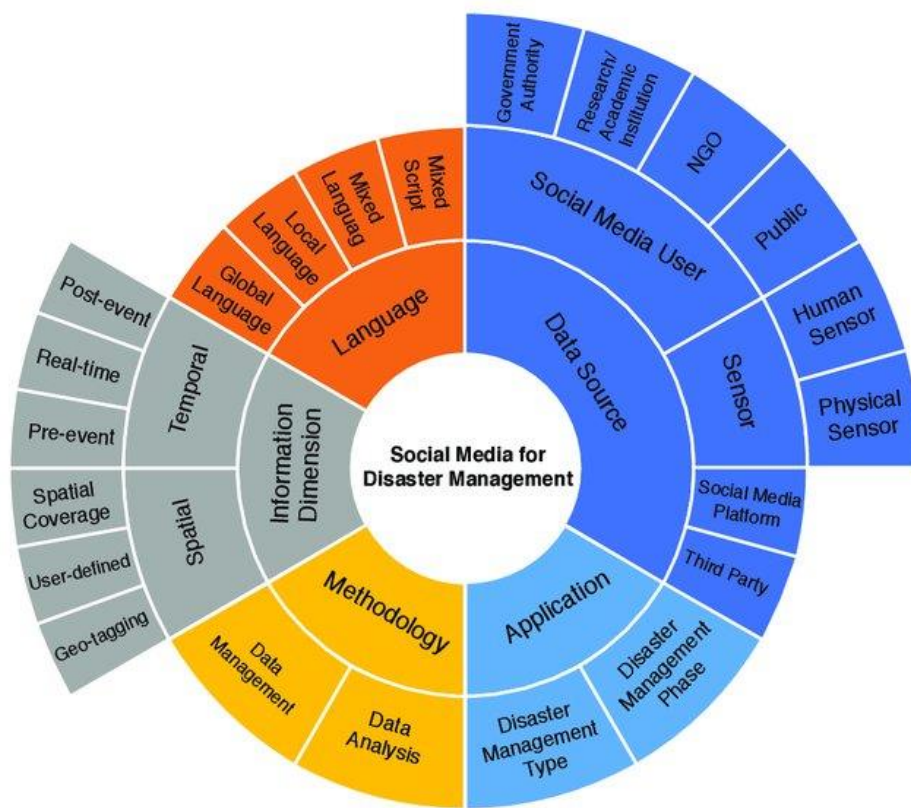


Fig. 1. Taxonomy of social media data management (Phengsuwan et al., 2021)

3. Social media analysis steps

For social media analysis, you should pay attention to what the purpose of an analysis is. Sufficient attention should also be paid to data sources, approaches, software architecture and data storage methods. For an effective and complete analysis, the following four steps should be considered:

- Discovery: In this step, we seek to discover hidden structures and patterns related to them.
- Tracking: At this stage, we decide on the data sources used.
- Preparation: In this step, we remove noisy data. Noisy data includes duplicate data, errors and fake data. In fact, we try to reduce the error rate as much as possible by cleaning the data.

- **Analysis:** In this step, according to the final goal, we choose a method of data analysis. Different methods are considered for data analysis, which are in the fields of Big Data and Machine Learning.

Social media data mining is used in various industries including business development, social sciences, health care and educational purposes. Once the data obtained from social media analytics is data mined, it can be used in a variety of contexts. Often, companies use a communication model such as correlation to check the social similarity of users with other users (Bayanati et al., 2023).

Also, organizations and businesses use user sentiment analysis to improve their satisfaction. Because users often publish positive or negative emotions in their posts.

In addition to the above, by analyzing the data of social networks, you can create attractive analytical reports of the trends of these media, trends in these networks, analysis of competitors, analysis of brand, service or product, analysis of public relations performance, analysis of face and personality, analysis of marketing campaigns and relationships. General and prepared other very diverse reports. These reports will greatly help businesses or individuals in advancing their goals (Chen et al., 2022).

4. Types of challenges in big data analysis

Social media data mining faces big challenges such as the big data paradox, the challenge of getting enough and complete samples, noise removal and evaluation dilemma. Social media data mining represents the virtual world of these media in a computational way, measures and designs models that help humans understand its interaction (Toloie-Eshlaghy et al., 2013).

In addition, this data mining provides the necessary tools to extract data for interesting patterns, analyze information dissemination, study impact and homophily, provide effective recommendations, and analyze new social behaviors in social media.

The important and basic issue that should be paid attention to is that social media data has the characteristics of massive data or Big Data. The data of social networks are obtained from various sources and the number of recorded records is such that the methods used in bulk data should be used to analyze them. The following key factors should be considered in this type of analysis:

- **Data volume:** Attention to the space required for data storage.
- **Speed:** Paying attention to the speed of creating data is very important. Below that one should be able to achieve the correct analysis in a reasonable period of time.
- **Data type:** It is necessary to pay attention to the fact that data will have different types. Many of these data are unstructured and some of them have their own structures.
- **Accuracy:** The quality of the data should be improved to the extent that the results of the data can be trusted.
- **Integrity and authenticity of data:** It is very important to pay attention to maintaining the security and compatibility of data, as well as the use of valuable data (Ghahremani-Nahr et al, 2022).

5. Conclusion

No one can deny the power of social networks in people's daily lives. Nowadays, people don't need to see each other closely to know each other's situation, everyone knows more about each other's condition

and what their friends and acquaintances are doing. Being constantly informed has been one of the main factors in the development of social networks in the first place. Now that the speed of receiving this information has increased so much, its amount should be increased accordingly. For this reason, users usually expand their interests and become members of special groups and read their special and favorite content. This issue shows the need for a more active society in the social field. These things have caused us to witness new and diverse information and news networks every day, or to witness new discussion groups with diverse topics every day. If people are going to spend more time on these networks than they will, there should be enough new and fresh content to satisfy everyone's needs.

With the expansion of social networks, especially on the virtual level, we see that people's privacy is becoming less and less important every day, and people themselves are also giving less importance to it. Twenty years ago, if someone published information and photos that are easily published today, it would not have a good effect and in many cases this information would be misused. But today, not only is this issue not so important, but the misuse of personal information has also become much less common. Of course, this process proceeds at a very low speed in our country, but it can still be felt and compared. This dimming of privacy also has a positive aspect, which is collective trust at the community level. With the dimming of this privacy, the number of accesses and misuses of this information will gradually decrease, and it will bring more peace and trust to the people. However, some are still strongly opposed to this trend and use various security measures to protect their information.

With the emergence and expansion of social networks, people's overall power and ability to access information increases. With this expansion, the power of social networks has also increased and they have presented themselves as an inseparable part of our lives. It is clear that their power includes having access to our information and much more.

As the methods of data analysis and evaluation progress, the methods of falsifying news and information also grow. Using complementary methods such as trust management and soft security can help us in more accurate data analysis. It should be noted that the analysis of social networks is a constant cycle that must be constantly improved and newer and more effective methods are replaced by older and ineffective methods. Online and offline companies measure their influence and influence in social networks through data. Kawi are able to provide specific products for specific groups of consumers. This measurement requires the collection of social network data, its complete analysis and correct and logical data-driven conclusions from it. For this set of work, you should go to specialists and experts in this field so that you can perform a correct and accurate data mining of social networks in the required topics.

References

Aliahmadi, A., Jafari-Eskandari, M., Mozafari, A., & Nozari, H. (2016). Comparing linear regression and artificial neural networks to forecast total productivity growth in Iran. *International Journal of Information, Business and Management*, 8(1), 93.

- Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). A framework for IoT and Blockchain Based on Marketing Systems with an Emphasis on Big Data Analysis. *International journal of Innovation in Marketing Elements*, 2(1), 25-34.
- Aliahmadi, A., Nozari, H., & Ghahremani-Nahr, J. (2022). AIoT-based sustainable smart supply chain framework. *International journal of innovation in management, economics and social sciences*, 2(2), 28-38.
- Barzegar Keliji, P., Aghajani, H. A., & Shayan Nia, S. A. (2021). Presenting a model based on artificial intelligence in the reverse supply chain of the home appliance industry in Tehran province by fuzzy genetic algorithm. *Journal of Industrial Strategic Management*, 6(1), 58-61.
- Bayanati, M. (2023). Business Model of Internet of Things and Blockchain Technology in Developing Countries. *International Journal of Innovation in Engineering*, 3(1), 13-22.
- Bayanati, M., Peivandizadeh, A., Heidari, M. R., Foroutan Mofrad, S., Sasouli, M. R., & Pourghader Chobar, A. (2022). Prioritize Strategies to Address the Sustainable Supply Chain Innovation Using Multicriteria Decision-Making Methods. *Complexity*, 2022.
- Chen, Y., Bayanati, M., Ebrahimi, M., & Khalijian, S. (2022). A Novel Optimization Approach for Educational Class Scheduling with considering the Students and Teachers' Preferences. *Discrete Dynamics in Nature and Society*, 2022.
- Eisapour, K., Bayanati, M., & Yousefpour, J. (2013). A mathematical model for ranking R&D organizations as a technology development factor. *Advances in Environmental Biology*, 717-721.
- Ghahremani-Nahr, J., & Nozari, H. (2021). A Survey for Investigating Key Performance Indicators in Digital Marketing. *International journal of Innovation in Marketing Elements*, 1(1), 1-6.
- Ghahremani-Nahr, J., Aliahmadi, A., & Nozari, H. (2022). An IoT-based sustainable supply chain framework and blockchain. *International Journal of Innovation in Engineering*, 2(1), 12-21.
- Moeini, E., Bayanati, M., Givi, S. B., & Soheili, S. (2013). Individual social responsibility a platform for intellectual property rights: Iranian users of CDs case. *Advances in Environmental Biology*, 709-717.
- Nahr, J. G., Nozari, H., & Sadeghi, M. E. (2021). Green supply chain based on artificial intelligence of things (AIoT). *International Journal of Innovation in Management, Economics and Social Sciences*, 1(2), 56-63.
- Najafi, S. E., Nozari, H., & Edalatpanah, S. A. (2022). Investigating the Key Parameters Affecting Sustainable IoT-Based Marketing. In *Computational Intelligence Methodologies Applied to Sustainable Development Goals* (pp. 51-61). Cham: Springer International Publishing.
- Nozari, H., & Sadeghi, M. E. (2021). Artificial intelligence and Machine Learning for Real-world problems (A survey). *International Journal of Innovation in Engineering*, 1(3), 38-47.
- Nozari, H., Fallah, M., Szmelter-Jarosz, A., & Krzemiński, M. (2021). Analysis of security criteria for IoT-based supply chain: a case study of FMCG industries. *Central European Management Journal*, 29(4).
- Nozari, H., Ghahremani-Nahr, J., & Szmelter-Jarosz, A. (2023). AI and machine learning for real-world problems. *Advances In Computers*, (online first).
- Nozari, H., Szmelter-Jarosz, A., & Ghahremani-Nahr, J. (2021). The Ideas of Sustainable and Green Marketing Based on the Internet of Everything—The Case of the Dairy Industry. *Future Internet*, 13(10), 266.
- Phengsuwan, J., Shah, T., Thekkummal, N. B., Wen, Z., Sun, R., Pullarkatt, D., ... & Ranjan, R. (2021). Use of social media data in disaster management: a survey. *Future Internet*, 13(2), 46.
- Pishkar, N., Nasimi, M. A., & RAHMATI, M. (2021). Developing a Conceptual Model of Green Supply Chain Antecedents and Consequences in the Qualification Approach.
- Rafierad, S., Aghajani, H. A., Agha Ahmadi, G., & Rahmaty, M. (2022). Construction and Validation of Dimensions and Components of the Organizational Anomie Scale in order to provide a Native Model in Government Hospitals. *Journal of System Management*, 8(2), 57-73.

Salehi Koocheh Baghi, S. A., Rahmaty, M., & Kia Kojouri, D. (2021). Presenting a Model of Organizational Insentience in the Red Crescent Society. *Quarterly Scientific Journal of Rescue and Relief*, 13(3), 228-236.

Shayan Nia, S. A., & Mirataollahi Olya, R. (2021). Provide a model for manpower scheduling using a hybrid meta-innovative algorithm in the Water and Sewerage Company. *Journal of Industrial Strategic Management*, 6(2), 18-32.

Shayannia, S. A. (2022). Designing a Multiobjective Human Resource Scheduling Model Using the Tabu Search Algorithm. *Discrete Dynamics in Nature and Society*, 2022.

Tavakkoli-Moghaddam, R., Ghahremani-Nahr, J., Samadi Parviznejad, P., Nozari, H., & Najafi, E. (2022). Application of internet of things in the food supply chain: a literature review. *Journal of applied research on industrial engineering*, 9(4), 475-492.

Toloie-Eshlaghy, A., & Bayanati, M. (2013). Ranking information system success factors in mobile banking systems with VIKOR. *Middle-East Journal of Scientific Research*, 13(11), 1515-25.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).