Developing a model of association rules with machine learning in determining user habits on social media

The object of the research is the habits of social media users. The problem in this research is that there is a lot of data spread across social media platforms ranging from image data, text data to audio data, making it difficult to identify starting from user behavior patterns, user interest in certain things and user habits. This research aims to explore scattered text data as text data has not been analyzed deeply in terms of word structure, thus concealing a lot of information and making it difficult to conduct analysis to determine user patterns and behavior on social media. The results obtained from this research are in the form of analysis and models that can identify and understand user patterns and behavior on social media using association rules and machine learning approaches. In applying the association technique, an a priori algorithm is used, which in the process determines all text data into item sets so that it can identify the habits of social media users and in the machine learning process there is a model formation process so that the results can be compared. There are several stages in the process of determining user habits, such as cleaning data, changing unstructured data to structured, then going through the rule association stage using an a priori algorithm applied to social media data so that the relationship can be seen between words. After that, machine learning is applied so that comparisons occur in seeing the relationships between words. This is new research in producing a model to identify user behavior using association rules and machine learning so that it can be used to map positive, negative or neutral user behavior

Keywords: habits, behavior, text data, associations, machine learning, social media

UDC 004

DOI: 10.15587/1729-4061.2024.305116

DEVELOPING A MODEL OF ASSOCIATION RULES WITH MACHINE LEARNING IN DETERMINING USER HABITS ON SOCIAL MEDIA

Antoni
Corresponding author
Master of Computer*
*Department of Engineering
Universitas Islam Sumatera Utara
Sisingamangaraja str., Teladan, Indonesia, 20217
E-mail: antonigtg@uisu.ft.ac.id
Mahrani Arfah
Master of Engineering*
Ferry Fachrizal
Master of Computer
Department of Computer Science
Politeknik Negeri Medan
Dr. T. Mansur, 9, Padang Bulan, Kec. Medan Baru, Kota Medan, Sumatera Utara, Indonesia, 20222
Okvi Nugroho
Master of Computer
Department of Information Technology
Universitas Muhammadiyah Sumatera Utara
Kapten Muchtar Basri str., 3, Glugur Darat II, Kec. Medan Tim., Kota Medan, Sumatera Utara, Indonesia, 20238

Received date 20.03.2024
Accepted date 27.05.2024
Published date 28.06.2024

***How to Cite**: Antoni, Arfah, M., Fachrizal, F., Nugroho, O. (2024). Developing a model of association rules with machine learning in determining user habits on social media. Eastern-European Journal of Enterprise Technologies, 3 (2 (129)), 55–61. https://doi.org/10.15587/1729-4061.2024.305116*

1. Introduction

The development of data in the digital era is increasingly rapid, especially in the utilization of social media, which has become a modern societal norm, leading to a surge in unstructured data available for extracting and identifying user habits. Within the context of user habits, several parameters such as activity, content, and interaction between users contribute to societal trends. Online media, particularly the social media platform X (Twitter), plays a significant role in facilitating communication and expression [1, 2]. X offers a microtext concept, requiring users to maximize character usage for communication. The volume of information circulating in tweet data escalates daily alongside the growing number of X social media users [3, 4]. This information surge results in vast amounts of data, garnering interest from numerous companies, research institutions, and governments [5–7]. X social media users exhibit diversity and possess unique motivations for engaging with the platform. Users often remain unaware that their activities on the X social network can shape their characteristics [8, 9].

There are problems in determining user habits such as the data available on social media platforms has never been extracted, the data is still unstructured so it is difficult to determine user habits manually with the various characteristics of users on social media such as user ID, user tweets. In determining user habits, data characteristics are needed so that we can obtain the required information. In this case, data mining techniques can be used to find out user habits, one of which is association rules. This technique will carry out analysis to see patterns in tweet data and find combinations of items that frequently appear as well as relationships between words in the data [10–12]. Then it will utilize machine learning techniques so that results can be compared to determine user habits.

Association rule learning is then applied using an a priori algorithm with the aim of finding association patterns

between words in the tweet data. The use of association rule techniques has been carried out by several researchers using social media data, such as research conducted by [13] concluding that the use of association rule learning to analyze action patterns on X accounts has produced a framework for conducting exploratory analysis of association rules in follower data on social media. Other research conducted [14] concluded that in carrying out associations, data needs to be processed first through text, which will produce links between data on social media. Meanwhile, other research conducted [15, 16] concluded that social media data can be used for association learning in detecting relationships between users. This research produces a link between user tweets and other user accounts [17]. This research will focus on the use of association rules and machine learning on social media data. Based on previous research, it can be concluded that social media data allows you to see relationships between users and determine follower patterns on social media. Therefore, research on determining user habits on social media becomes relevant by utilizing techniques from association rules with a priori algorithms and the use of machine learning.

2. Literature review and problem statement

The research [18] concluded that the use of social media has produced data that can be explored to find happiness from lifestyle, but there are several problems such as the difficulty of extracting social media data. These problems can be solved using machine learning. So this research can produce user relationships on social media with happiness using variables such as age, sleeping difficulty, frequency of activity and in this research we have not been able to find a model that can identify user habits starting from frequency of activity and frequency of communication interactions. All this makes it possible to carry out studies or research on association techniques and machine learning in forming models for identifying user habits.

In the research [19], X was used, which is a microtext social media that can use tweets with a total of 140 characters. There is a problem in determining user patterns with only 140-character word data, so in solving this problem, other parameters are used, such as time, location, and tweet data, to determine user patterns by utilizing artificial intelligence with the Recurrent Neural Network algorithm. All this makes it possible to carry out research in the application of association techniques and machine learning, making it easier for the model to identify user habits.

The research [20] used data from the social media X to carry out association learning and extract knowledge regarding people's attitudes towards crises around the world. The problem with this research is that it is difficult to determine the topic, so we intend to use the proposed methodology consisting of topic extraction and visualization techniques, such as WordClouds, to form opinion groups or themes. It then uses Association Rule Mining (ARM) to find sets of frequently used words and generate rules that infer user attitudes. Based on this, there are deficiencies in identifying user habits, it is necessary to apply research on association techniques and machine learning.

The research [21] concluded that Latent Dirichlet Allocation (LDA) can be used to process data and then classify the emotions contained in the dataset used. The problem is that tweet data is difficult to identify the emotions of social media users, so it requires an approach with the Latent Dirichlet Allocation (LDA) algorithm. The approach taken requires a large volume of data so it will use good accuracy. This research produces a model with 80 % accuracy in determining user emotions. However, the resulting model cannot determine user habits on social media, so it is necessary to approach and apply the model in applying association and machine learning techniques to identify user habits on social media.

The research [22] discusses data mining with users on social media with the theme of COVID-19. The problem is that there is a lot of data that contains the word COVID-19 so it needs to be classified. Therefore, this research will collect data, look at interactions between users on each topic related to COVID-19 and thirdly carry out sentiment analysis to find out the effect of COVID-19 on emotions by utilizing association techniques. However, the model that is formed can only see the sentiment that appears on social media, the model cannot identify user habits, so it is possible to carry out research to determine user habits on social media by applying association techniques and machine learning.

The research [23] carried out association mining, which uses a modified FP-Growth algorithm called FP-GCID (a new FP-Growth algorithm based on Cluster ID) to generate associations. This research carried out the rules for validation whether the two previously mentioned requirements are met.

The research [24] concluded that Latent semantic analysis (LSA) can be used to process data on social media so that behavioral patterns can be found for each user. However, there is a problem, namely that the data used has poor characteristics. So text processing needs to be applied to perfect the data and the resulting model still uses the highest frequency in the data, whereas to identify user habits, many variables or parameters are needed, such as communication trends, friendship trends and posting trends on social media. These parameters can be used to form models. All this makes it possible to carry out research by studying association techniques and machine learning from latent semantic analysis so that user behavior can be known.

3. The aim and the objectives of the study

The aim of this research is to create a model with association techniques to determine user habits using association rules and machine learning on social media X (Twitter), which in the process applies association rule techniques with machine learning to identify trends, patterns, and behavior manifested in interactions on social media platforms.

To achieve this aim, the following objectives are accomplished:

– to determine the parameters and variables used in implementing algorithms in machine learning-based association techniques;

– to determine model validation and evaluate model performance against machine learning-based association techniques.

4. Materials and methods

The object of this research is the habits of social media users.

The hypothesis in this research is that there is no difference in association techniques and machine learning in identifying user habits on social media, and there is a difference in applying association techniques and machine learning in identifying user habits, where the association technique connects user behavior items and then processes them by machine learning to potentially achieve better processing time. This research will use data from X social media taken from 2018 January to 2023 December. This research methodology consists of 3 processes such as scrapping data on X, text preprocessing and association rules as shown in Fig. 1 as follows.

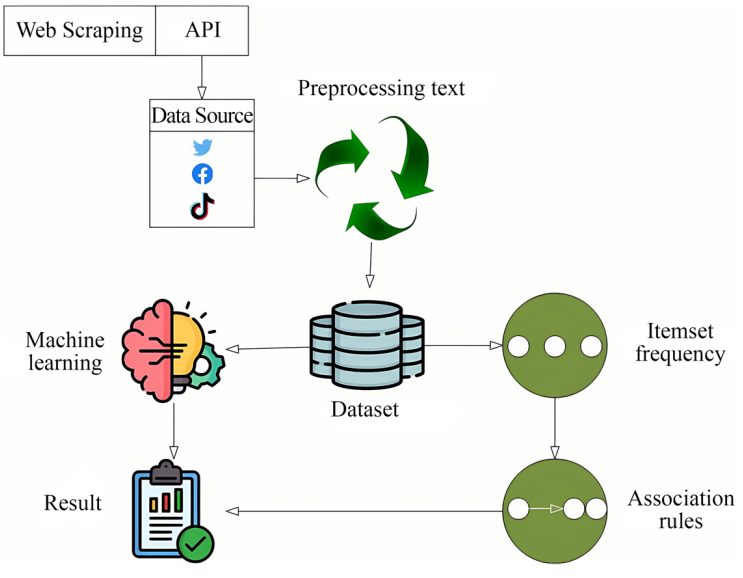

Fig. 1. Proposed architecture

The information from Fig. 1 explains that the research architecture will carry out data processing on X data resulting from data crawling. This data comes from social media X. The following is an explanation of the research architecture in searching for user habits with an association rule approach using an a priori algorithm:

1. In the research architecture, there is a data source that collects data on X, Facebook, TikTok social media using a scrapping process. Data obtained on X social media uses the Twint library, which collects 30,000 data from 2018 to 2023.

2. Preprocessing is the process of cleaning text or sentence es contained in documents, documents obtained from scrapping are still unstructured so it is necessary to preprocess the text with stages such as case folding, tokenize, stopword and stemming with the aim of obtaining data in a structured form.

After the tweet data has been preprocessed, the association rule process with the a priori algorithm is applied to what has been preprocessed. The following is the flow of the association rule with the a priori algorithm in Fig. 2 below.

Fig. 2 shows the process of association rules that determine the relationship between variables in a document and define the relationship between words in a tweet document on social media. In theory, what is often done in the association process is based on an item and a collection of items that make associations between them. In this research we will make adjustments by applying notation and terminology from association rule mining. This research will change the transaction variables with words in document tweets, items become words and groups of items become sets of words. This research will find words with the highest value. High level of association rules in a tweet document using support and confidence metrics. In association rules, there are several algorithms such as a priori and FP-Growth, which process the rules from the minimum value and then extract the set of words that appear most frequently. In this research, stages are utilized to observe the relationship between a collection of tweets, namely minimum support and confidence. In this research, it is assumed that $I=\{i1, i2, i3, i4.....in\}$ is a set and attribute that symbolizes a variable in a text document, then $P=\{t1, t2, t3....tn\}$ is denoted to be a collection of transactions. This research uses a collection of tweets. Each word in a tweet is symbolized by P, which contains words per word. So it produces a rule of the form $X \Rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$.

In selecting a rule from a set of rules for a variable from a collection of tweets, it must have a threshold. In applying the threshold, we will use minimum support and minimum confidence, each symbolized by X from a set of words. X is the proportion of tweets in a document. So the formula for this research is as follows:

$$\text{Conf}(X \rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}. \quad (1)$$

In the association rules there is a Lift, which will take measurements to determine the strength of the association rules that have been formed based on support and confidence values. The following is the formula for lift in the association rule:

$$\text{Lift}(X \rightarrow Y) = \frac{\text{Conf}(X \rightarrow Y)}{\text{Support}(Y)}. \quad (2)$$

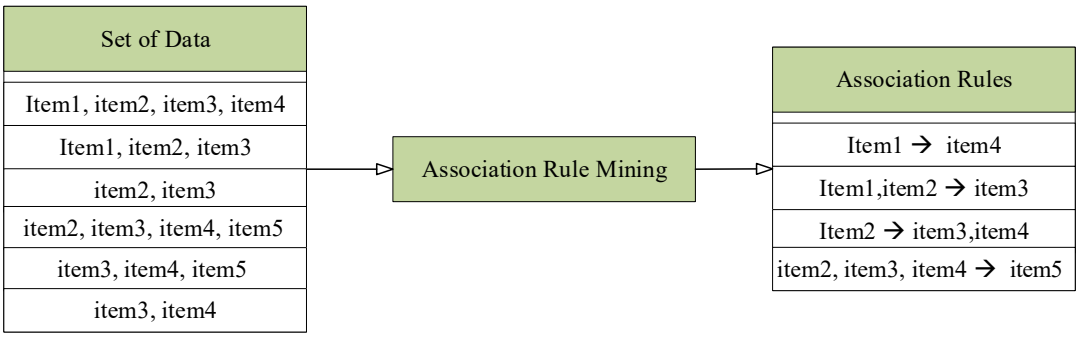

Fig. 2. Association rule

A priori will discover the words that appear most frequently in tweet documents. The structure of a priori is built based on the occurrence of words in the document, in the process all will have thresholds with minimum support and minimum confidence values.

5. Results of the association rule in the user habit model

5. 1. Determination of association rule parameters

In determining the association rule parameters, several variables are established to ascertain user habits on social media. One such variable is the word cloud, which displays words with the highest frequency in documents. For this research, tweets are managed using bilinear interpolation, and a word cloud plot is generated. This method illustrates the most significant words, showcasing a collection of important words with the highest frequency in tweets, as depicted in Fig. 3. By displaying a word cloud this research will represent the important words in the document. In Fig. 3, the size of the largest word illustrates the importance of the word in the tweet.

Fig. 3. Word cloud of tweet document data

Fig. 3 shows a word cloud, which is an interpretation of all data on each word that is often spoken by users on social media with the aim of describing each user's habits so that user behavior can be identified.

After that, a frequency of a collection of words, which will then be grouped into a list of words in a document. This research will display the words with the highest frequency in Table 1 such as "lari", "pergi", "berdiri", "tonton", "lucu".

Table 1

Highest word frequency

No.	Words	Word count
1	lari	1,551
2	pergi	1,191
3	berdiri	736
4	tonton	679
5	lucu	665
6	komika	600
7	hadir	568
8	komedi	521
9	cerita	489
10	tiket	456

Table 1 displays the results of the frequency of each word contained in the dataset obtained from social media platforms such as Facebook, X, and TikTok. For each word, a count of frequently appearing words is provided, allowing interpretation as the most commonly used words by users.

5. 2. Determination of association model performance

In this first stage of research, words are collected into a list within a document. The research displays the words with the highest frequency in Table 1. The top 10 words with the highest frequency will include examples such as 'lari', 'berdiri', 'cerita', 'lucu'. The association results are presented in Table 2 as follows.

Table 2

Highest word frequency

Antecedents	Consequents	Support	Confidence
{'lari'}	({'pergi'})	0.0548335	0.5279
{'berdiri'}	({'cerita'}	0.090438	0.2760
({'lari'}	({'lucu'}	0.0904384	0.1888

The rules contained in Table 2 relate to "lari" and "berdiri". The first rule gets a support value of 0.5279 or a percentage of 52 %, indicating that the text document data on X has the highest frequency of sentences with "lari" and "berdiri". The second rule displays the results for the words "berdiri" and "cerita" with a value of 0.2760 or a percentage of 27 %, and the third rule displays the results for the words "lari" and "lucu" with a value of 0.1888 or a percentage of 18 %. Then, the results of the association rules for the word "lari" are displayed in Table 3 as follows.

Table 3

Highest word frequency

No.	Antecedents	Consequents	Confidence	Lift
1	({'cerita'})	({'lucu'})	0.704	14.708
2	({'lari'})	({'lucu'})	0.598	12.503
3	({'berdiri'})	({'lucu'})	0.569	11.885
4	({'kriting'})	({'lucu'})	0.552	11.528
5	({'tonton'})	({'lucu'})	0.545	11.389
6	({'youtube'})	({'lucu'})	0.540	11.289
7	({'iklan'})	({'lucu'})	0.534	11.1
8	{'Indonesia'}	({'lucu'})	0.532	11.1:

Based on the data displayed in Tables 3, 4, the data set for the words "lari", "berdiri", "cerita" has a level of confidence with the respective values listed in Table 3. This research displays a visualization of the results of the association with tweet documents as shown in Fig. 4 as follows.

Fig. 4 shows the results of associations within the minimum threshold of 1 %. In these results, relationships between words in the document were found, such as "kriting, lucu" berdiri, tonton "berdiri, lucu". Then there is a visualization with a larger number of association rules. The following are the results of the association with a minimum support threshold of 0.5 % as shown in Fig. 5 below.

Fig. 5 shows a visualization of association rules with the level of words that precede each one based on the number of words in the sentence with the aim of getting a pattern of relationships between words so as to make it easier to apply a model that determines user habits.

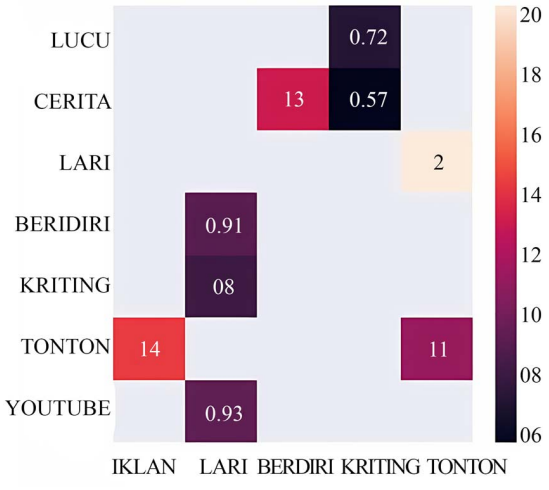

Fig. 4. Visualization of association rule results

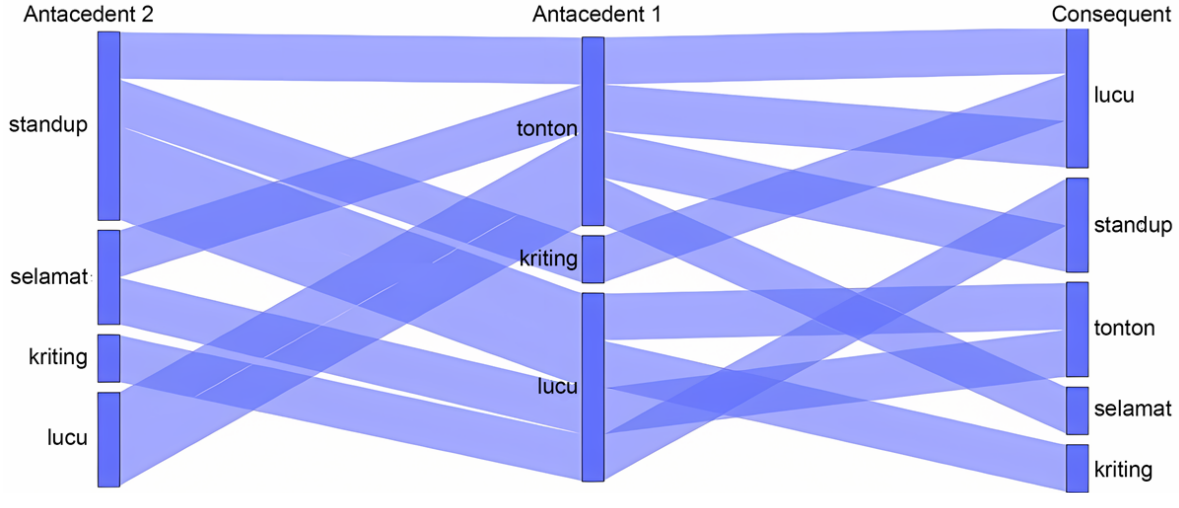

Fig. 5. Visualization of association rule results

6. Discussion of the results in determining the habits of social media users

The results obtained are models that utilize association rules and machine learning techniques in creating and designing models to determine user habits on social media. This process focuses on applying association techniques to examine user habits between items related to user behavior on social media, aiming to produce a model capable of identifying user habits on social media. In this model there is an object in applying the results that displays the entire analysis to obtain a model that can identify user habits on social media such as the analysis of identifying user habits on social media given in Table 2 and Table 3. In Tables 2, 3, patterns or the highest frequency of each word that is related or connected between items form an itemset pattern. Then, after the itemset pattern formation, an image object displays the entire process for finding the most important words in the social media data shown in Fig. 3, while Fig. 4 displays the evaluation model to assess the model's accuracy using confusion evaluation matrices. In Fig. 5, a model illustrates the relationship process between items such as communication trends, posting trends, and comment trends on social media data. In its application, the model is formed based on association rule techniques with a priori algorithms and techniques from machine learning. The data used were obtained from social media, then data preprocessing was carried out to obtain structured data, which was subsequently processed to form a model. This research then analyzes document data to determine user habits for each related word.

In implementing the proposed method, it is related to solutions that solve problems in identifying user habits. In the process of identifying user habits, a relationship between the items in Table 3 emerges because identifying user habits is highly effective using association rules and techniques. Then, after establishing the relationship between items, the machine learning technique executes a pattern recognition process between items to facilitate determining the habits of each user based on X's social media data. By employing the proposed method, this research yielded results with a level of novelty and uniqueness of technique or method that differs from previous research in determining user habits. Research conducted in [25] merely identified user habits using conventional methods combined with machine learning and data obtained through interview stages, while our research utilized social media data processed through natural language processing stages, including data annotation, application of text preprocessing, and processing using an

a priori algorithm and machine learning. These differences lead to variances in the accuracy obtained. With comprehensive research stages, good accuracy can be achieved in identifying user habits.

The weakness of this research lies in the data obtained from social media X, data from social media X must be cleaned repeatedly manually and data annotations applied to obtain good data in structured sentences. This research can develop towards latent semantic analysis and Latent Dirichlet Allocation techniques so that user behavior can be more accurate by combining machine learning techniques and association rules.

7. Conclusions

1. In determining user habits on social media, we employ verb parameters utilized in the association rule process, a data mining technique generating a set of rules for the entire itemset. This research proposes applying an a priori algorithm to social media data documents, offering a new perspective on user habits based on data. In the machine learning technique, each item set in the data undergoes training to identify user habits with only 5 habit items, whereas the association technique reveals 6 habit items interconnected with user activities. This fusion of association rules techniques in machine learning has a significant impact, leading to the effective determination of user habits on social media.

2. In the research results there is a relationship between the word "lari" and several words such as the word "pergi" and the word "berdiri", which produces a support value of 0.545 and a confidence value of 11.38. Association learning can be used to find out similarities between words so that the highest frequency of words greatly influences the results of the confidence value and support value and the results will show relationships between words that are very relevant to the data for each user on social media based on the activities carried out by the user.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. El Rahman, S. A., AlOtaibi, F. A., AlShehri, W. A. (2019). Sentiment Analysis of Twitter Data. 2019 International Conference on Computer and Information Sciences (ICCIS). https://doi.org/10.1109/iccisci.2019.8716464
2. Kumar, S., Khan, M. B., Hasanat, M. H. A., Saudagar, A. K. J., AlTameem, A., AlKhathami, M. (2022). An Anomaly Detection Framework for Twitter Data. Applied Sciences, 12 (21), 11059. https://doi.org/10.3390/app122111059
3. Lubis, A. R., Prayudani, S., Lubis, M., Nugroho, O. (2022). Sentiment Analysis on Online Learning During the Covid-19 Pandemic Based on Opinions on Twitter using KNN Method. 2022 1st International Conference on Information System & Information Technology (ICISIT). https://doi.org/10.1109/icisit54091.2022.9872926
4. Lubis, A. R., Nasution, M. K. M., Sitompul, O. S., Zamzami, E. M. (2023). A new approach to achieve the users' habitual opportunities on social media. IAES International Journal of Artificial Intelligence (IJ-AI), 12 (1), 41. https://doi.org/10.11591/ijai.v12.i1.pp41-47
5. Thakur, N. (2022). Twitter Big Data as a Resource for Exoskeleton Research: A Large-Scale Dataset of about 140,000 Tweets from 2017–2022 and 100 Research Questions. Analytics, 1 (2), 72–97. https://doi.org/10.3390/analytics1020007
6. Schmidt, S., Zorenb hmer, C., Arifi, D., Resch, B. (2023). Polarity-Based Sentiment Analysis of Georeferenced Tweets Related to the 2022 Twitter Acquisition. Information, 14 (2), 71. https://doi.org/10.3390/info14020071
7. Firdaniza, F., Ruchjana, B., Chaerani, D., Radianti, J. (2021). Information Diffusion Model in Twitter: A Systematic Literature Review. Information, 13 (1), 13. https://doi.org/10.3390/info13010013
8. Kwon, J.-H., Kim, S., Lee, Y.-K., Ryu, K. (2021). Characteristics of Social Media Content and Their Effects on Restaurant Patrons. Sustainability, 13 (2), 907. https://doi.org/10.3390/su13020907
9. Wibowo, A., Chen, S.-C., Wiangin, U., Ma, Y., Ruangkanjanases, A. (2020). Customer Behavior as an Outcome of Social Media Marketing: The Role of Social Media Marketing Activity and Customer Experience. Sustainability, 13 (1), 189. https://doi.org/10.3390/su13010189
10. Ali Hakami, N., Hosni Mahmoud, H. A. (2022). The Prediction of Consumer Behavior from Social Media Activities. Behavioral Sciences, 12 (8), 284. https://doi.org/10.3390/bs12080284
11. Gupta, V., Jung, K., Yoo, S.-C. (2020). Exploring the Power of Multimodal Features for Predicting the Popularity of Social Media Image in a Tourist Destination. Multimodal Technologies and Interaction, 4 (3), 64. https://doi.org/10.3390/mti4030064
12. Khafaga, D. S., Alharbi, A. H., Mohamed, I., Hosny, K. M. (2022). An Integrated Classification and Association Rule Technique for Early-Stage Diabetes Risk Prediction. Healthcare, 10 (10), 2070. https://doi.org/10.3390/healthcare10102070

13. Orama, J. A., Borr s, J., Moreno, A. (2021). Combining Cluster-Based Profiling Based on Social Media Features and Association Rule Mining for Personalised Recommendations of Touristic Activities. Applied Sciences, 11 (14), 6512. https://doi.org/10.3390/app11146512
14. Lee, S., Cha, Y., Han, S., Hyun, C. (2019). Application of Association Rule Mining and Social Network Analysis for Understanding Causality of Construction Defects. Sustainability, 11 (3), 618. https://doi.org/10.3390/su11030618
15. Kruse, R., Lokukatagoda, T., Alkhushayni, S. (2022). A framework for association rule learning with social media networks. IOP SciNotes, 3 (1), 015001. https://doi.org/10.1088/2633-1357/abe9be
16. Diaz-Garcia, J. A., Ruiz, M. D., Martin-Bautista, M. J. (2022). A survey on the use of association rules mining techniques in textual social media. Artificial Intelligence Review, 56 (2), 1175–1200. https://doi.org/10.1007/s10462-022-10196-3
17. Shazad, B., Ullah khan, H., Rehman, Z., Farooq, M., Mahmood, A., Mehmood, I. et al. (2019). Finding Temporal Influential Users in Social Media Using Association Rule Learning. Intelligent Automation and Soft Computing. https://doi.org/10.31209/2019.100000130
18. Zhang, J., Marino, C., Canale, N., Charrier, L., Lazzeri, G., Nardone, P., Vieno, A. (2022). The Effect of Problematic Social Media Use on Happiness among Adolescents: The Mediating Role of Lifestyle Habits. International Journal of Environmental Research and Public Health, 19 (5), 2576. https://doi.org/10.3390/ijerph19052576
19. Zingla, M. A., Ettaleb, M., Latiri, C. C., Slimani, Y. (2014). INEX2014: Tweet Contextualization Using Association Rules between Terms. Working Notes for CLEF 2014 Conference, 574–584. Available at: https://ceur-ws.org/Vol-1180/CLEF2014wn-Inex-ZinglaEt2014.pdf
20. Koukaras, P., Tjortjis, C., Rousidis, D. (2022). Mining association rules from COVID-19 related twitter data to discover word patterns, topics and inferences. Information Systems, 109, 102054. https://doi.org/10.1016/j.is.2022.102054
21. Güven, Z. A., Diri, B., Çakaloğlu, T. (2018). Classification of TurkishTweet emotions by n- stage Latent Dirichlet Allocation. 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT). https://doi.org/10.1109/ebbt.2018.8391454
22. Drias, Y., Drias, H. (2021). Sentiment Evolution Analysis and Association Rule Mining for COVID-19 Tweets. Journal of Digital Art & Humanities, 2 (2), 3–21. https://doi.org/10.33847/2712-8148.2.2_1
23. Lian, S., Gao, J., Li, H. (2018). A Method of Mining Association Rules for Geographical Points of Interest. ISPRS International Journal of Geo-Information, 7 (4), 146. https://doi.org/10.3390/ijgi7040146
24. Kusumaningrum, R., Wiedjayanto, M. I. A., Adhy, S., Suryono (2016). Classification of Indonesian news articles based on Latent Dirichlet Allocation. 2016 International Conference on Data and Software Engineering (ICoDSE). https://doi.org/10.1109/icodse.2016.7936106
25. Ranjan, R., Kumar, S. S. (2022). User behaviour analysis using data analytics and machine learning to predict malicious user versus legitimate user. High-Confidence Computing, 2 (1), 100034. https://doi.org/10.1016/j.hcc.2021.100034