

Association Rule Analysis-Based Identification of Influential Users in the Social Media

Saqib Iqbal¹, Rehan Khan², Hikmat Ullah Khan^{2,*}, Fawaz Khaled Alarfaj⁴,
Abdullah Mohammed Alomair³ and Muzamil Ahmed²

¹College of Engineering, Al Ain University, Al Ain, United Arab Emirates

²Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt, 47040, Pakistan

³Department of Computer Science and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU),
Al-Houfuf, Saudi Arabia

⁴Department of Computer Science and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU),
Al-Houfuf, Saudi Arabia

*Corresponding Author: Hikmat Ullah Khan. Email: hikmat.ullah@ciitwah.edu.pk

Received: 04 April 2022; Accepted: 16 June 2022

Abstract: The exchange of information is an innate and natural process that assist in content dispersal. Social networking sites emerge to enrich their users by providing the facility for sharing information and social interaction. The extensive adoption of social networking sites also resulted in user content generation. There are diverse research areas explored by the researchers to investigate the influence of social media on users and confirmed that social media sites have a significant impact on markets, politics and social life. Facebook is extensively used platform to share information, thoughts and opinions through posts and comments. The identification of influential users on the social web has grown as hot research field because of vast applications in diverse areas for instance political campaigns marketing, e-commerce, commercial and, etc. Prior research studies either uses linguistic content or graph-based representation of social network for the detection of influential users. In this article, we incorporate association rule mining algorithms to identify the top influential users through frequent patterns. The association rules have been computed using the standard evaluation measures such as support, confidence, lift, and conviction. To verify the results, we also involve conventional metrics for example accuracy, precision, recall and F1-measure according to the association rules perspective. The detailed experiments are carried out using the benchmark College-Msg dataset extracted by Facebook. The obtained results validate the quality and visibility of the proposed approach. The outcome of propose model verify that the association rule mining is able to generate rules to identify the temporal influential users on Facebook who are consistent on regular basis. The preparation of rule set help to create knowledge-based systems which are efficient and widely used in recent era for decision making to solve real-world problems.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: Association rule mining; ranking; social web; influential users; social media

1 Introduction

Our everyday lives have become excessively reliant on social media as it provides space for people to express their feelings and opinion. The modern era of web, known as Web 2.0, has enabled the movement of common people from information buyers to information creators in a sort of popular news journalism. This sort of web based platforms constitute the social media, wikis, social clarification and labeling, and media sharing on social sites like Facebook and Twitter [1]. In social media sites, the users who make an impact on other users in the network have played a significant role are termed as *the influential users*. To predict users' influence, the users' behavior and cooperation with other users are analyzed. The level of activity is dependent on the success of the social media sites [2]. The identification of interesting topics and influential users on social media are considered an interesting and active research area. A few studies have attempted to distinguish influential users; nonetheless, To cope with influential users, most have utilized page rank centrality or degree centrality-based techniques [3].

In recent years, the social web attains more focus due to its nature as it facilitate users to create social relations all over the globe [4]. Social websites are intentionally developed in such manner to enhance the social connections as well as provide space for discussions, chatting, entertainment and academia networks [5]. Moreover social media provide space for homophily (user with similar opinions about a specific subject), as they can create groups and pages [6]. A huge amount of information and knowledge related to different topics are being generated through discussions and accessible for others [7]. The social media platforms, for instance Twitter, Facebook, WhatsApp, WeChat, Pinterest, etc. facilitate business organizations to make business to business and business to customer connection to market their products. Furthermore, users can also create virtual communities such as Quora, Stack-Over-Flow etc. on social networking sites for information seeking. In these communities, user posed a question or issue related to a specific topic and other community members including topic experts respond with relevant answers [8].

On the Facebook site alone, there are increasing billions of users every month as shown in Fig. 1. Specifically, Facebook separates itself through both ubiquity and effect on common people. In reverse sequential order, Facebook is an individual section that is distributed worldwide and provides content generation facilities on diverse topics. Social networks are generally the platforms that provide the linkage of solitary persons connecting on large scale [9]. As finding social media users' influence is important, the reward of recognizing influential users has been highly significant. As Facebook is very important for the marketing purposes, the influential users help companies to understand key concerns, identify trends, motivate to purchase product or use services. A large number of companies are moving their focus on influential users. Based on existing studies, finding influential users is dependent upon the location of users in the blogging network [10] social web activities like others and actions of users circulate to large audience so this is so interesting to find influential bloggers is an active research field in various domain [11]. To search out the influential users has a big relevance to online shopping and advertisements so identifying influential users using sentiment features has also been done [12]. On social network analysis, different kinds of research were carried out, and influential users recognized from the association rule mining has matched the outcome of page rank and degree centrality [13].

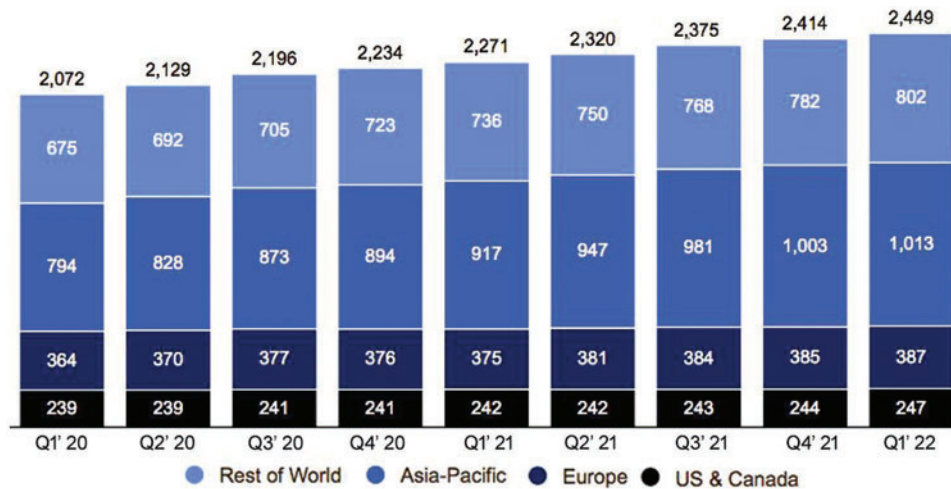


Figure 1: Monthly Active Users (MAUs) of facebook according to statistics 2022 [4]

Subsequently, users can associate themselves with a community anytime, anywhere. However, this involvement affects the order of existing influential users in the community. Therefore, temporally identification of the influential users assists organizations for business strategies and marketing, falsifying political campaigns and experts in academia network. For example, the influential users are repeatedly market-movers. Since these users have an impact within their community which allows them to motivate other users towards a specific point. Moreover, they can also influence decision making and product selection of their followers. According to study [14], about 64% of marketing enterprises admitted that blog advertising is the most effective way of marketing. The influential users are considered as blog representatives that impact opinions and thoughts of different users about a specific political party during election [15]. Furthermore, identification of influential users assists to comprehend the altering concerns and interests of users and anticipate potential hazards and rewards, and make timely and proactive changes to plans. The key contributions of this research study are as follows:

- Provision of the possibility to identify the influential users from social media by incorporating association rule learning algorithm.
- Analysis of how different influential users and their influence on the post remain consistent over a period of time or not.
- Building knowledge base based on association rules extracted using the Apriori algorithm which is employ to identify top influential users.
- Evaluation of the proposed approach based on detailed empirical analysis using College-Msg dataset using standard performance evaluation measures.

The reminder sections of this study are structured as follows: Section 2 explore the existing studies relevant to identification of influential users, feature-based models, network-based models, and finding influential users using association rule learning. Section 3 elaborate the proposed methodology in detail. Next, Section 4 presents the statistics of dataset and evaluation measures. Moreover, the Section 5 presents the analysis of experimental results and findings. Section 6 elaborates the conclusion of the proposed research study.

2 Related Work

Many researchers have worked on social media analysis in different ways and by using different kinds of techniques and tools for their work. Most of these works were performed on famous social media networks including Facebook, Twitter, Instagram, etc.

2.1 Influential Users and Social Networks

There are diverse methods and approaches applied for finding top influential users such as Features based approaches [16], graph-based approaches, etc. [17]. However, association rule mining has not been explored for this research purpose. The links that are direct on Facebook represent everything from common interests to friendly friendships, or people probably keen on celebrity gossip or breaking news. Social media investigation is the prevalent area within the investigation of arranging and information mining. Moreover, connect prediction is another active research area in the field of network and association rule mining [18]. The association forecast is additionally one of the areas where analysts have worked apart. Analysts around the world have moreover worked on other terms like user's identity from smaller-scale bloggers Churn forecast within the network, prediction from the expectation of advancement, and numerous more works within the field [19].

In all the other areas the user role is one of the most important factors, the user role includes finding the influence of users on social media and bloggers, in other areas of related work the ability to find influential users might also affect [10]. As finding social media users' influence is important, likewise it is also important to identify the potential bloggers. A model proposed by [20] introduces a model for influential users in blog sites to find some works employed to find influential users and data mining methods. Another framework proposed by [21] finds influential leaders using the propagation graph mining method. Their proposed framework will describe a transaction database of user actions from [9] a social graph. Hu and Lee, also find influential users to use frequent pattern mining. A popular platform to interact, communicate and collaborate between friends on social media platforms [22]. In this work, the sentiments extract through tweeter to operate the forecasting influence of social media [23]. Few of the active web communities are so vast and different that it will be challenging to successfully extract important knowledge using the web community. Mapping of an association rule can provide a new technique for personalizing various services market analysis. Twitter is used to advertise various campaigns, and sometimes advertisers appoint a few buzzers to make a campaign activity going more frequently [24]. Some of the researches have been done using algorithm of cluster, to find the top users base on influential aspect. Some of the effected algorithms are required to find influential technique, such as Eclat algorithm etc. [25].

A process that is meant to find frequent itemset is association rule learning. In relational databases, various kinds of correlations, associations, or structures from data sets are found in transactional databases and other forms of data sources. In association rule mining the basic purpose of extracting rule is to analyze the relationship of items in different transactions. Nowadays the social media sites have become a surprising influence on the social and the economic impact in the last few years. Nowadays Facebook is the most famous online social network which attracts billions of monthly active users all around the world. Although online social network attracts great attention from the practitioner and the research community. Social network plays a major role in marketing techniques, because of their publicity increase in the web search domain [2]. In social networks like Facebook, it is significant to recognize the nodes that affect their neighbors are known as influence node. In this paper, they show some different techniques based on the architectural attributes of different social platforms which might be published through users. In the social media sites, advertisements corresponding to

business are commonly placed between the media content [26]. However, the user can deny unwanted adds termed as adware by using different add blockers. The adware on some sites also steals users personal and sensitive information without user permission. However, the well-known platforms such as Facebook seek user permission before sharing sensitive information. The social media sites place advertisement between the actual content and force users to view advertisement. The placement of adware might be affecting the acceptance rate of users. Furthermore, the impact of adding replacement depends on the directly proportional to user involvement with advertised product [27].

2.2 Finding Top Influential Users Using Association Rule Mining

Erlandsson et al. [13] involve Aprori algorithm along with association rule learning to recognize the influential. They gathered a large dataset of the Facebook pages called “get together” and performed their analysis based on that page. During their experiments, they used to support, confidence, lift and conviction to find the potential users by employing association rule learning approach. The analysis social media platforms are well-known area under the umbrella of network science [28]. Moreover, the link prediction and link analysis are considers as primarily study of the social media analysis [29]. Some of the other common research fields focused on detection of popular user with respect to specific group on social media by incorporating post and comments [30]. Another study [31] introduce framework for recognition of information cascade with in social networks. Moreover, [32] presents an approach for identification of pattern involve in diffusion mechanism in social media. The social media networks have significant factors to determine the role of users. These factors assist in identification of influential users with in specific community. And then, with the help identified influential users, we can viral a certain information, market a specific product or motivate target audience towards a particular agenda. A study presented in [33] detect the influential user of Twitter social media site to examine whether a user who has large number of followers is the same as influential user.

Finding influencer users depends on the social networking site and flow of information on social media platform for example, we use tweets and mentions to find influential users on Twitter [9]. Literature shows that existing studies also involve clustering technique to determine potential users based on influential characteristics for example likes, comments and replies [8]. Moreover, a research study [34] introduce linear regression for recognition of influential users by involving categorical influential features. However, several research study incorporate learning-based approaches for instance Page Rank algorithm for identification of potential users within the network [35]. Jianqiang et al. [14] extract post, comments and likes data 100 educational institutes form Facebook and examine the relation between user and generate association rule to find out the association links between the opted courses and gender. Furthermore, few authors examine the personal hobbies on social media by involving association rule learning. De Salve et al. [1] address the issues with examination of folksonomies anatomy and involve association rule learning recognize the potential users with in a specific network.

3 Research Methodology

In order to carry out our research, we propose a framework that finds the time-based temporal users. The framework is shown in Fig. 2. In social network analysis, the impact of a user’s or node on other users or nodes in the network can be expressed using various centrality measures. For example, in the event that users P, Q, R and S share same interests, there’s a possibility that if P, Q, and R as of now have preferred on a point, S will too like it. This suggests that users can have a substantial influence on other users, or that influential levels can be determined from a few other users.

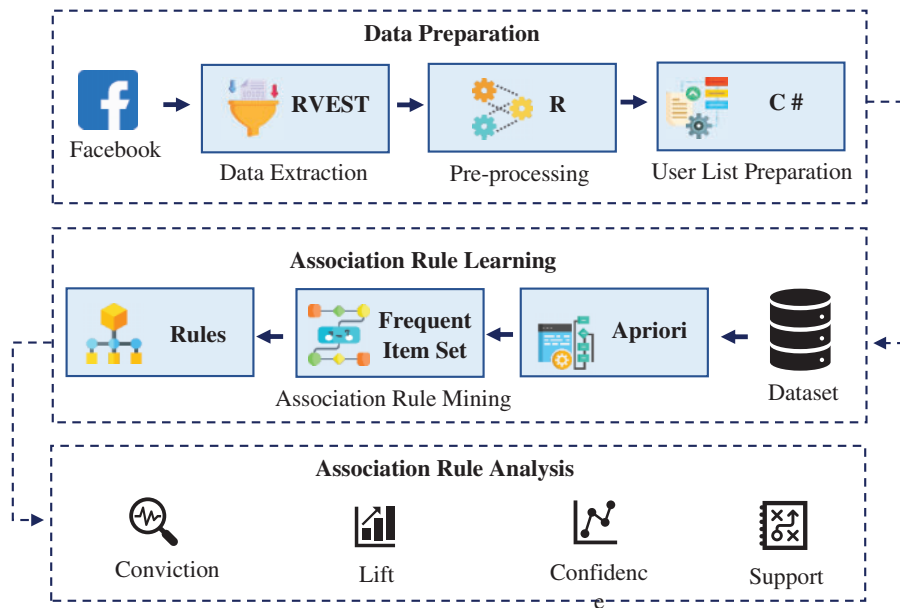


Figure 2: Flow diagram of the proposed work

3.1 Association Rule Learning

Association rule learning involves occurrence of different items to find frequent item sets. The frequent item sets are extracted based on two measures including support and confidence. The term support indicates how frequently a particular rule occur in the dataset. However, confidence is probability that a specific rule remains correct in the entire dataset. For example, the rule, “people who buy tires and auto accessories also acquire automotive services done”. By taking this example we will find out how often this rule occurs that when new tires or any other equipment being purchased then we also need the services. Moreover, this research study involves association rule learning approach to explore the influential user on social media. The association rule learning requires transactional data with occurrence of different items. Therefore, firstly modeled social media data according to transactional in which users are denotes as items and posts indicate transactions.

A frequent itemset is calculated with help of Apriori and the association rule learning over value-based databases. In Apriori algorithm employ “bottom-up” mechanism for candidate generation by extending the most frequent subset single item every time. These are intended to function a dataset involving different transactions. Association rule mining model gives set of frequent items along with rules generated based on given items. Let’s look at the mathematical formalism of the association rules problem for given items set presented as $U = \{u_1, u_2, u_3, \dots, u_n\}$ and transaction set as $T = \{t_1, t_2, t_3, \dots, t_n\}$ where n is size of attribute set. In T , each individual transaction t_i has a set of items that are presented with a unique transaction identifier, and items belong to U also called subset (\subset) of U itemset. Moreover, any generated rule can be specified as an implication, $X \cup Y$ where X and Y both are subset of U ($X, Y \subset U$) without having common values, i.e., $X \cap Y = \emptyset$. Furthermore, the term X and Y form the antecedent side of rule. Association rule learning Machine learning is considered as machine leaning strategy that points to discovering how one thing influences other with examine that how regularly these objects occur together in different transaction presents in dataset. In this regard, we applied support and confidence as criteria to extract the desired rules.

3.2 Association Rule Analysis

The primary objective of this study is to identify the temporal-based user influence by involving the Apriori algorithm mechanism of sliding window. The sliding window mechanism proved as feasible and efficient approach for current scenario as diverse domains. Several dimensions exist that will fruitful to know and learned association rules. It comprises of four evaluation metrics including support, confidence, lift and conviction.

- *Support*: The term support indicates the frequency of item sets that appeared most frequently in a dataset. Support can be defined by the number of existences of $\{U_1, U_2\}$ divided by the number of items in $\{W\}$ as given in Eq. (1).

$$Support(\{U_1, U_2\}) = \frac{U_1, U_2}{W} \quad (1)$$

- *Confidence*: This says how likely the item U_2 is purchased when an item U_1 is purchased, expressed as $\{U_1 > U_2\}$. The value of confidence for given items can be computed using Eq. (2).

$$Confidence(\{U_1, U_2\}) \rightarrow U_3 = \frac{Support(U_1, U_2, U_3)}{Support(U_1, U_2)} \quad (2)$$

- *Lift*: This says how likely the item U_2 is purchased when an item U_1 is purchased. Moreover, if items are not depended on one another then the value of lift is equal to 1, however if different items have dependency association with other items, then the value of lift must be greater than 1. The value of lift is computed using Eq. (3).

$$Lift(\{U_1, U_2\}) \rightarrow U_3 = \frac{Support(U_1, U_2, U_3)}{Support(U_1, U_2) \times Support(U_3)} \quad (3)$$

- *Conviction*: The probability of the item occurred without any other item if they were dependent is compared to the corresponding frequency of considered item appearing without other item. It is the fraction of the projected support that $\{U_1, U_2\}$ happens without C and computed using following Eq. (4).

$$Conviction(\{U_1, U_2\}) \rightarrow U_3 = \frac{1 - Support(U_1, U_2, U_3)}{1 - Conviction(U_1, U_2) \rightarrow U_3} \quad (4)$$

4 Experimental Setup

4.1 Dataset

Facebook is a popular social medium utilized by every category of people like celebrities, politicians and common people. Facebook users express their opinions by sharing comments, images in the form of sharing a post. We examine the pages of different topics and find their influence on users. R language-based (RVEST) package has been used for data extraction for the following pages. The data cleaning and preprocessing will be done using R language. The required format of the data has been prepared to contain the list of all the users who commented in the temporal order in a post has been prepared using Microsoft Visual Studio .NET. In this research study, the analysis has been carried out using the standard dataset of College-Msg. The detail of dataset statistics is presented in Tab. 1.

Table 1: Detail of dataset statistics

Parameter	Value
Nodes	32.4 K
Edges	54.4 K
Density	0.0001035
Maximum degree	623
Average degree	3
Average clustering coefficient	0.11268
Number of posts	18,994
Number of comments	7,46,561
Posts in 80% data	15,195
Posts in 20% data	3,799

In the dataset, the term node indicated as connection points with in a network that are connected through edges. For example, in the social media networks, the users of a particular platform are considered as nodes [16]. Similarly, the connection between more than one node is known as edge. In the network, edges join different nodes in a defined way also called connection type. For instance, a user U_1 as nodes comment on the post of another node U_2 then we join U_1 and U_2 together with comment as type connection. The degree of compactness of a substance. The density is a network analysis metric that portrays the probability of possible connections present in considered network [19]. Likewise, the value of degree can be defined as maximum degree of specific node in overall tree. For instance, the binary tree has maximum degree of a node is 2, similarly the ternary tree has maximum degree of a node is 3, respectively. In the graph network degree of a node can be calculated as the portion of edges that are incident on that node. However, the degree of nodes with edges as loop are counted as twice. The last measure is clustering coefficient, which shows the degree of a particular node in network tend to cluster together.

4.2 Performance Evaluation Measure

Let us discuss the metrics we used for evaluation of the research study. We share the logic behind proposal of using confusion matrix concept for introduction of metrics. For predicting user participation to measure how good association rule mining works, learn and test design have been utilized. To measure the comments, the active users are considered to be active by finding that they generate posts at regular intervals. For instance, if client S comments on a post and the existing rule say that P , Q & R influence user S , this rule is considered to be valid in case all of P , Q & R have made at least one comment each sometime recently S comments true positive is considered as a rule to predicts the effectiveness and activeness of a particular user. Likewise, false positive rate, the rule recognizes the activity of user, however, the particular user is inactive. The real adverse effect is that there are no rules if there are no active users. False negatives mean that the user is active but has no rules. Tab. 2 presents the method for construction of the confusion metric.

- *Accuracy*: The accuracy is widely used to evaluation measure to assess the performance of association rules learning models. The accuracy is fraction between the correctly predicted rules over all rules.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

- *Precision*: The evaluation metric precision defined as portion of positively predicted instances over positive class. The value of precision can be calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

- *Recall*: The evaluation metric recall examines the ability of model to predicts positive instances. The recall is the fraction between positive samples accurately predicted as positive over the positive samples. It can be computed as follows:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

- *F1-Score*: The F1-score considers both the recall and precision and computed the harmonic mean and provide a single value. The F1-score can be calculated as follows:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (8)$$

Table 2: A proposed method for confusion matrix construction

	Example rule: {P, Q, R} → S
D1	{P, Q, R, Z} → true positive
D2	{P, Q, R} → false positive
D3	{M, N, L} → true negative
D4	{S, Z} → false negative

5 Results and Discussion

The detailed experiments are carried out to verify proposed methodology. In experiments, 80% of the data is use for generating rules and 20% of the data is used to validate generated rules.

5.1 Overall Influential Users

We applied Apriori algorithm to generate rules to find out frequent itemset and against those items, we built association rules. Tab. 3 presents the frequency of rules according to their type extracted with 100 percent confidence using Apriori algorithm. Moreover, Tab. 4 shows a sample of generated rules sorted by the length in ascending order. In these rules, the users on the right-hand side are antecedents and influence the user on the left-hand side called the consequent. Social media users are represented in the rules by a unique identifier corresponding to anonymized user data. The rule 248,534→742 indicates that 248 and 534 are two users influencing another user 742. Moreover, in rules 4 and 5, the user 450 is participating with 65,915 to influence 687 and similarly, he is also adding two other groups of users 586 and 923 to influence the same user. Furthermore, rules 6, 7 and 8 show that the users 120 and 415 like to make a larger group to influence other users.

Table 3: Number of rules generated with apriori algorithm

Type of rules	No of values
Rules with length 1	80
Rules with length 2	1478
Rules with length 3	1967
Rules with length 4	2121
Rules with length 5	647
Rules with length 6	105
Total generated rules	6398

Table 4: Sample of generated rules (sorted by length ascending order)

Sr #	Rule	Length	Confidence
1	248, 534→742	2	100
2	159, 196→120	2	100
3	178, 783→502	2	100
4	450, 65, 915→687	3	100
5	450, 586, 923→687	3	100
6	120, 159, 334, 633→214	4	100
7	120, 214, 415, 74, 752→890	5	100
8	159, 214, 415, 607, 687→334	5	90
9	159, 334, 415, 65, 75, 890→120	6	100
10	120, 159, 334, 415, 75, 890, 90→214	7	100

The results show that our approach is feasible and achieves 80%, 87% and 45% confidence, accuracy and precision respectively. [Tab. 5](#) presents a list of top influential with the number of posts and comments. According to the results, the user 168 is the top influencer that influences other users with 447 posts and 4710 comments. Likewise, the user 120 has made 988 posts and made 1321 comments counted among the influential users.

Table 5: List of top influential users regarding posts and comments frequency

S. no	Users ID	No of posts	No of comments
1	168	447	4710
2	851	446	2723
3	530	670	2136
4	742	740	1657
5	159	543	1453
6	120	988	1321

(Continued)

Table 5: Continued

S. no	Users ID	No of posts	No of comments
7	90	566	1127
8	356	170	1085
9	75	310	872
10	65	151	567

5.2 Temporal Order of Influential Users

Finding the overall influential user from the whole data set, then we now move towards the temporal influential users. Firstly, we explore the influential users of the previous year's data set. The data statics are shown in [Tab. 6](#). For all these attributes of support, confidence and lift and in the data set we proceed with the length related to previous years. The minimum acceptable value of lift is 1, however, we have 99% lift values more than 1 and obtained an average value of lift as 7.534 is excellent. For all these experiments, we check the result of influential users compared with different sets of rules and compute the consequent and antecedents' names. We compute different values of confidence, support and lift against all of the influential users and find top users concerning each of the influential users.

Table 6: Data statics of temporal analysis

Attribute value	Previous years		
	Minimum	Maximum	Average
Support	0.003	0.135	0.004
Confidence	0.5	1	0.946
Lift	0.859	17	7.534
Rule length	1	4	3.414

According to [Tab. 7](#), we obtained a good confidence level among all the influential users. However, the support level is small due to the huge volume of data, Moreover, the value of lift is also good as we obtained a 9.05 maximum lift. After finding the top rules concerning confidence, then we find the top rules for the lift and support as presented in [Tabs. 8](#) and [9](#) respectively. According to the results, the value of lift is inversely proportional length rule. The maximum value of support is 0.06 for rule 356,948 → 416 with 0.65 and 5.953 confidence and lift respectively. Results also show that the value of support is inversely proportional as well as the lift is directly proportional to the confidence value. The user 120 frequently occurred in top rules regarding lift and attained 0.315, 1.0 and 4.90 maximum support, confidence and lift respectively. [Fig. 3](#) shows a different level of support and confidence for different influential users they many other users at different level.

Table 7: Top rules with respect to confidence

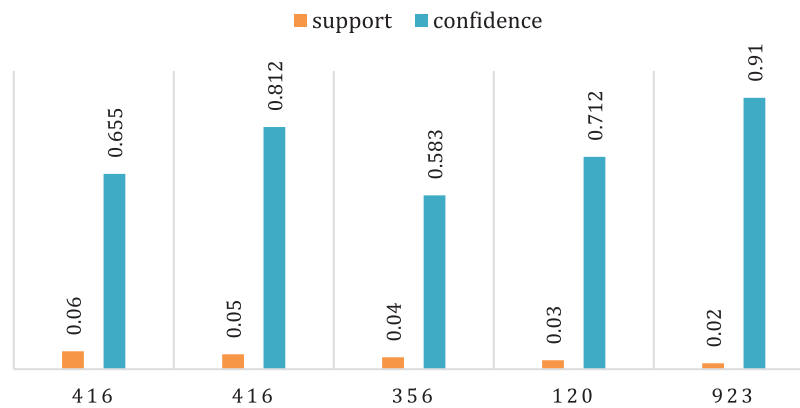
Antecedent	Consequent	Support	Confidence	Lift
159, 196	120	0.17	1.0	5.28
127, 147	629	0.02	1.000	8.24
415, 65, 75, 752	890	0.03	1.000	6.53
159, 415, 53, 75	120	0.019	1.000	5.28
502, 530, 687, 742, 851	450	0.13	1.000	9.88

Table 8: Top rules with respect to lift

Antecedent	Consequent	Support	Confidence	Lift
(248, 534)	742	0.0315	1.0	9.05
(189, 687, 920)	851	0.0315	1.0	8.67
(120, 334, 53, 890)	90	0.0300	1.0	3.90
(120, 214, 334, 633, 90)	159	0.0315	1.0	4.62
(120, 159, 334, 415, 890, 90)	214	0.0315	1.0	4.90

Table 9: Top rules with respect to support

Antecedent	Consequent	Support	Confidence	Lift
356, 948	416	0.06	0.655	5.953
313	416	0.05	0.881	6.280
765, 912	356	0.04	0.583	9.174
75, 752, 890	120	0.03	0.712	10.998
542, 629, 742	923	0.02	0.910	7.186

**Figure 3:** Top users with respect to support

5.3 Findings

While performing experiments, we explore the influential users using Apriori algorithm and association rule learning. In this study we recognize the influential users who has influence on other users based on their social media activities for example posts and comments, etc. For performing empirical analysis of proposed model, we incorporate benchmark dataset and evaluation metrics to assess the performance. We calculate the influential user with the help of association rule learning from the overall dataset of previous years. In this way, we calculate all influential users. When finding the influential users from all datasets. In the influence of user within same network changes with respect to time. To validate the phenomena that influence of users changes over time we employ previous dataset. The level of support is observed from 0.0135 to 0.17. However, there are so huge number of posts during the specific period the user cannot see other users on each post. Overall, we calculate support confidence and lift for the previous year and among all the time period we identify the influential users for their support confidence and lift and match all the result for better demonstration of influential users. We calculate the results of the most influential users over distinct time periods and compute the result variance. In this regard we have computed influential users all over the time according to their lift, confidence and support. During the empirical analysis, we notice that the influence of the particular user changes temporally.

When we find the result of the previous year the user (742, 851, 90) as the most influential user. When we compute results for confidence, the system shows the top influential users as user id (120, 629, 890). Now it shows the variation in influence regarding specific time period. As we calculate the outcomes of influential users over different time periods. The results depict different confidence value with respect to time interval as ranging from 0.6 to 1. The high confidence value clearly shows that our propose methodology outperform state of art methods. We identify the potential user in connection with different levels of association rule learning. Moreover, consider the length of rules ranging from 1 to 7 and the average of these rules are 3.5. We show the best result was gathered in previous year because their support and confidence are good and their overall rules are high length. The time is the key factor that impact the position of influence users. Furthermore, results also confirmed that influential users are changes over the time interval.

6 Conclusion

To explore the influential users of social media, association rule learning concept is explored and Apriori algorithm is applied on College-Msg users' dataset. The proposed methodology identified the influential users who influenced other users to generate activity in the social media in the form of comments and posts. For the empirical analysis, the standard dataset is used and standard performance metrics have been applied. The proposed methodology figured out the influential users from different perspectives. As time is as significant factor with respect to user influence. The empirical analysis also shows that influence of user may vary over different time intervals. The certain users may have higher influence and lower influence with time. We compute the influential user of the previous years and in order to change the evidence of change of influence of the different perspective, we find it in previous years and then we examine the results. During experiments, we explored the difference in their values of confidence, support and lift and the proposed system computed the values of support, confidence and lift against each of the influential users that we find. The results confirm that our approach is efficient and that we are able to get good confidence level with high number of rules shows that there are many influential users in the Facebook. While the primary objective is to find the temporal based influential

users and explored the lift and the confidence level of each of the user to support the findings of the research study.

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University for funding this work through Research Group No. RG-21-51-01.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] A. De Salve, P. Mori, B. Guidi, L. Ricci and R. Di Pietro, "Predicting influential users in online social network groups," *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 3, pp. 1–50, 2021.
- [2] S. A. Alsaif, A. Hidri and M. S. Hidri, "Towards inferring influential facebook users," *Computers*, vol. 10, no. 5, pp. 1–21, 2021.
- [3] F. Riquelme and P. González-Cantergiani, "Measuring user influence on twitter: A survey," *Information Processing and Management*, vol. 52, no. 5, pp. 949–975, 2016.
- [4] S. Jain and A. Sinha, "Discovering influential users in social network using weighted cumulative centrality," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 1, pp. e6521, 2022.
- [5] N. A. Ghani, S. Hamid, I. A. Targio Hashem and E. Ahmed, "Social media big data analytics: A survey," *Computers in Human Behavior*, vol. 101, no. 3, pp. 417–428, 2019.
- [6] C. Fan, Y. Jiang and A. Mostafavi, "The role of local influential users in spread of situational crisis information," *Journal of Computers-Mediated and Communication*, vol. 26, no. 2, pp. 108–127, 2021.
- [7] X. Zhang, G. Dai, P. Chang, S. Wei, X. Chen *et al.*, "A Multi-feature learning model with enhanced local attention for vehicle re-identification," *Computational Materials Science*, vol. 69, no. 3, pp. 3549–3561, 2021.
- [8] K. Ali, C. -T. Li and Y. -S. Chen, "Joint selection of influential users and locations under target region in location-based social networks," *Sensors*, vol. 21, no. 3, pp. 1–17, 2021.
- [9] I. Elkabani, L. A. Daher and R. Zantout, "Use of FP-growth algorithm in identifying influential users on twitter hashtags," in *Proc. of the 2020 the 4th Int. Conf. on Computer and Data Analysis*, New York, NY, USA, pp. 113–117, 2020.
- [10] J. Chen, Y. Deng, Z. Su, S. Wang, C. Gao *et al.*, "Identifying multiple influential users based on the overlapping influence in multiplex networks," *IEEE Access*, vol. 7, no. 1, pp. 156150–156159, 2019.
- [11] H. U. Khan and A. Daud, "Finding the top influential bloggers based on productivity and popularity features," *New Review of Hypermedia and Multimedia*, vol. 23, no. 3, pp. 189–206, 2017.
- [12] U. Ishfaq, H. U. Khan and K. Iqbal, "Modeling to find the top bloggers using sentiment features," in *2016 Int. Conf. on Computing, Electronic and Electrical Engineering (ICE Cube)*, Quetta, Pakistan, pp. 227–233, 2016.
- [13] F. Erlandsson, P. Bródka, A. Borg and H. Johnson, "Finding influential users in social media using association rule learning," *Entropy*, vol. 18, no. 5, pp. 1–15, 2016.
- [14] Z. Jianqiang, G. Xiaolin and T. Feng, "A new method of identifying influential users in the micro-blog networks," *IEEE Access*, vol. 5, no. 2, pp. 3008–3015, 2017.
- [15] S. Utz and J. Jankowski, "Making 'friends' in a virtual world: The role of preferential attachment, homophily, and status," *Social Science Computer Review*, vol. 34, no. 5, pp. 546–566, 2018.
- [16] Q. Ma, X. Luo and H. Zhuge, "Finding influential users of web event in social media," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 3, pp. e5029, 2019.
- [17] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu *et al.*, "Influence analysis in social networks: A survey," *Journal of Network and Computer Applications*, vol. 106, no. 2, pp. 17–32, 2018.

- [18] A. Zareie, A. Sheikhhahmadi and M. Jalili, "Identification of influential users in social networks based on users' interest," *Information Sciences*, vol. 493, no. 2, pp. 217–231, 2019.
- [19] G. Amati, S. Angelini, G. Gambosi, G. Rossi and P. Vocca, "Influential users in twitter: Detection and evolution analysis," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 3395–3407, 2019.
- [20] P. K. Novak, L. D. Amicis and I. Mozetič, "Impact investing market on twitter: Influential users and communities," *Applied Network Science*, vol. 3, no. 1, pp. 40, 2018.
- [21] R. E. L. Bacha and T. T. Zin, "Ranking of influential users based on user-tweet bipartite graph," in *2018 IEEE Int. Conf. on Service Operations and Logistics, and Informatics (SOLI)*, Singapore, pp. 97–101, 2018.
- [22] A. Mahmoudi, M. R. Yaakub and A. Abu Bakar, "New time-based model to identify the influential users in online social networks," *Data Technologies and Applications*, vol. 52, no. 2, pp. 278–290, 2018.
- [23] I. Nuzhdenko, A. Uteuov and K. Bochenina, "Detecting influential users in customer-oriented online communities," in *Int. Conf. on Computational Science*, Wuxi, China, pp. 832–838, 2018.
- [24] W. Mnasri, M. Azaouzz and L. Ben Romdhane, "Parallel social behavior-based algorithm for identification of influential users in social network," *Applied Intelligence*, vol. 51, no. 10, pp. 7365–7383, 2021.
- [25] W. Hasan Alwan, E. Fazl-Ersi and A. Vahedian, "Identifying influential users on instagram through visual content analysis," *IEEE Access*, vol. 8, no. 1, pp. 169594–169603, 2020.
- [26] A. Monteserin and M. G. Armentano, "Influence me! predicting links to influential users," *Information Retrieval*, vol. 22, no. 1, pp. 32–54, 2019.
- [27] M. Zhang, P. Yang, C. Tian and S. Tang, "You can act locally with efficiency: Influential user identification in mobile social networks," *IEEE Access*, vol. 5, no. 1, pp. 136–146, 2017.
- [28] G. Lingam, R. R. Rout and D. V. L. N. Somayajulu, "Adaptive deep Q-learning model for detecting social bots and influential users in online social networks," *Applied Intelligence*, vol. 49, no. 11, pp. 3947–3964, 2019.
- [29] Z. Zhao, H. Zhou, B. Zhang, F. Ji and C. Li, "Identifying high influential users in social media by analyzing users' behaviors," *Journal of Intelligent & Fuzzy Systems*, vol. 36, pp. 6207–6218, 2019.
- [30] Al-Garadi, M. Ali, V. K. Dewi, R. S. Devi, A. Ejaz *et al.*, "Analysis of online social network connections for identification of influential users: Survey and open research issues," *ACM Computing Surveys*, vol. 51, no. 1, pp. 1–20, 2018.
- [31] W. Sun, G. Zhang, X. Zhang, X. Zhang and N. Ge, "Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy," *Multimedia Tools and Applications*, vol. 80, no. 20, pp. 30803–30816, 2021.
- [32] A. Muruganantham and G. M. Gandhi, "Framework for social media analytics based on multi-criteria decision making (MCDM) model," *Multimedia Tools and Applications*, vol. 79, no. 5, pp. 3913–3927, 2020.
- [33] Q. Ma and J. Ma, "A robust method to discover influential users in social networks," *Soft Computing*, vol. 23, no. 4, pp. 1283–1295, 2019.
- [34] K. Saito, K. Ohara, M. Kimura and H. Motoda, "Which is more influential, 'who' or 'when' for a user to rate in online review site?," *Intelligent Data Analysis*, vol. 22, no. 2, pp. 639–657, 2018.
- [35] M. Alshahrani, F. Zhu, L. Zheng, S. Mekouar and S. Huang, "Selection of top-K influential users based on radius-neighborhood degree, multi-hops distance and selection threshold," *Journal of Big Data*, vol. 5, no. 1, pp. 28, 2018.