ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A NOVEL AUTHENTICATION SYSTEM FOR LIMITED USER GROUP ON SOCIAL NETWORKS

J.Adamkani*, Dr.K.Nirmala
Research Scholar, Quaid-E-Millath Govt College for Women, Chennai, India,
Research Supervisor, Quaid-E-Millath Govt College for Women, Chennai, India.

DOI: 10.5281/zenodo.49785

ABSTRACT

Social networks are a characteristic role of today's Internet user and it is utilized by more than a billion peoples around the world. Social networks permit individuals to impart thoughts and their substance to their friends and others. This interaction and correspondence uncovers a considerable measure of data, frequently including individual data unmistakable to any individual who needs to view it. Once in a while social network users also suffer from some security issues data annoying and pernicious actives over social groups. So as to give security in social locales with nature, a novel security algorithm has been proposed. Block unwanted users from joining in the LUG based on the profile of the user. This will block unwanted contents and users in the LUG; access the user profile, social behaviors and their friend's information before allowing them. The exploration of this work endeavors to incorporate a combinatory methodology in a parametric study on the security angles. The primary element taken for the exploration is the sorts of limited user forums, for example, adult users, minor users, business users etc. This factor will be combined with the type of security measure known as token based. With this combination a number of parameters will be obtained in SNS. The exploration work will endeavor at improving the security with ease of use in SNS. In the event that the user from the group posts any undesirable information, then the user will be blacklisted. For each user in the group reputation score will be ascertained. In light of the score, they will permit to get to the group content.

KEYWORDS: SNS, Data mining, privacy policies, Feature extraction.

INTRODUCTION

Social Networking is gaining popularity tremendously. Social Network Sites (SNS) [1] provide an online opening for concurring people to meet together and share their data online. It also provides an effective business platform for entrepreneurs, to promote their business. It can also facilitate open online forums for like-minded amateur or even professionals such as writers, actors, musicians and the like. But novice users of social Network Sites (SNS) still do not realize the potential and danger of SNS. [2] The procedure to become a member of any SNS is very simple, that anyone can enroll by answering a few basic profile based questions. The danger starts with revealing his/her personal information to the rest of the world. The easy way to understand a social network is to compare it with a circle of friends in schools days or traditional groups formed within a community.

There are several social network systems have different groups and different predefined and customized policy setting options. Due to the tremendous growth of social network sites, handling security and privacy issues are very important. Limited User group has different rules to verify and restrict users, but in several SNS, the users may specify different privacy policies, so verifying the data without affecting the privacy of users is more important. So, when one joined a group, then the user may be able to influence their friends to join. This type of influence between users in SNSs is called as *social influence*, which also used by several business agencies, which is due to the large population and the strong social influence in SNSs [3] [4].

On the other hand, the chance of doing target oriented advertisement in SNSs also opens a door for malicious actions. Specifically,[5] dishonest users in an SNS may intentionally give misleading recommendations to their

friends and contacts. To take advantage of the social influence, several users disseminate their advertisements and recommendations via SNS. In fact, this type of advertisement becomes very common in Facebook and Twitter. Due to the misleading recommendation and fake friends suggestions and given by dishonest members, even if a group and members in the groups are may still be misled to join in it [6]. Sometimes the user needs to specify different LUG (limited User Group) policies. As like the user policies, some LUG has some policies to restrict user from joining and commenting on their groups. To perform the above operations, the data mining techniques are integrated.

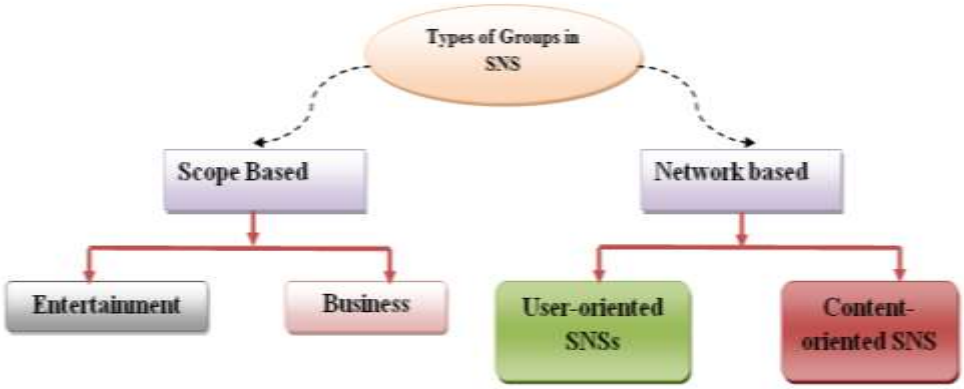
TYPES OF GROUPS IN SNS

Scope based:

In terms of their scope model, SNSs can be classified into the following two types [15]:

Entertainment: Most SNSs are dedicated to entertainment. Their focus is on delivering fun and interactive social experience online to registered users. Popular SNS sites that are mainly entertainment-oriented are Facebook, Myspace, Hi5 and Flickr ETC.,

Business: In this category, the focus of SNSs is to connect the world's professionals to make them more productive and successful. Through business SNSs, registered users create profiles that summarize their professional expertise and accomplishments. Indicative SNS sites in this category are LinkedIn and Xing.



Network based

The taxonomy addresses the network model of SNSs that can be classified into the following two types:

User-oriented SNSs: The SNSs of this category emphasize on the social relationships. In such SNSs, content sharing is mainly among users who belong to the same community. Indicative SNSs of this category are the Facebook, MySpace and LinkedIn.

Content-oriented SNSs: The users' network is not determined by the underlying social relationships but by their common interests. Indicative SNSs of this category are blog networks, question answering networks and video networks (e.g., YouTube).

SOCIAL NETWORK MINING

Vast amounts of user-generated content are created on social media sites every day. [16] This trend is likely to continue with exponentially more content in the future. Hence, it is critical for consumers, service providers and producers, to outline the management and utilization of enormous user-generated data. Social media growth is motivated by these challenges: (1) how can a user be heard? (2) Which source of information would a user use? (3) How can user experience be improved? Answers to these questions are secreted in the social media data. These challenges present sufficient opportunities for data miners developing new algorithms and methods for social media. Data generated on social media sites are different from usual attribute-value data for classic data mining. Social media data are highly user-generated content on social media sites. Social media data are huge, unstructured, noisy,

dynamic and distributed. These distinguishing data pose challenges to data mining tasks to invent new, efficient techniques and algorithms. For example, Facebook and Twitter report Web traffic data from approximately 149 million and 90 million unique U.S. visitors per month, respectively. On the video sharing site YouTube, more than 4 billion videos are sighted per day, and 60 hours of videos are uploaded every minute. The picture sharing site Flickr, as of August 2011, hosts more than 6 billion photo images. Depending on social media platforms, social media data can frequently be very noisy[17]. Removing the noise from the data is necessary before performing effective mining.

RELATED WORK

This review outlines the theoretical [7] frameworks researchers have used to understand adolescents and SNS. It brings together work from disparate fields that examine the relationship between SNS and social capital, privacy, youth safety, psychological well-being, and educational achievement.These research strands speak to high-profile concerns and controversies that surround youth participation in these online communities, and offer ripe areas for future research [14].This work [8] Beato, Filipe, Markulf Kohlweiss, and Karel Wouters designed and implemented a system that allows users to define and enforce selective access control policies for data published on social network sites. By using a PKI encryption scheme, such as Open PGP they were able to keep users' data confidential, even towards the SNS operator, by means of encryption. Through the integration into a Firefox extension encrypted content is automatically decrypted by the browser of authorized users. They tested the extension with the own social network site tesbed and in other social network sites, like Facebook and MySpace.[9] authors present a horizontal view of social influence, more specifically a quantitative study of the influence of neighbours on the probability of a particular node to join a group, on four popular Online Social Networks (OSNs), namely Orkut, YouTube, LiveJournal, and Flickr. Some key points about the social networking issues such as social Media platforms improve reach and promote campaign messages and organizations' activities. [10] They can simultaneously enable: i) rapid and ongoing capturing of public mood, sentiment and knowledge about health issues; ii) free or extremely inexpensive amplification of broadcast messages; iii) a range of opportunities to tailor messages and engage the public in a conversation about health promotion and health protection; iv) user-generated content and feedback systems which improve loyalty and trust in organizations and confidence in information [13]. Open, evolving SNSs represent remarkable new research opportunities are discussed in [11] [12]. These sites provide users with templates that, while intended for recreational purposes and organized presentation, are ideally suited for data collection and analysis. As Rogers foreshadowed 20 years ago, when "videotext, electronic messaging systems, and computer bulletin boards" were just gaining popularity: The new interactive media of the 1980s offer potential means to deal with certain of the epistemological problems of network research.

PROBLEM DEFINITION

Finding and restricting the users in the social groups according to the different policies and rules, this aims to identify and classify the users according to their behavior, and performing best decision on SNSs. Handling the privacy policies on Group Recommendation System (GRS) and Social Network sites to restrict and allow irrelevant users. Even though groups consist of members with different characteristics and behaviors, which can be defined by their profile features, as their group size, grow, they tend to attract people with similar characteristics, but without affecting the privacy, the system finds optimal way to solve by using Data Mining techniques.

Steps for solving issues in SNS:

1. To rank order security aspects of selected network security mechanisms as one factor on their effectiveness in SNS for various forums.
2. To analyze personalized social authentications for restricted groups of SNS.
3. To evolve optimum authentication mechanism through algorithms for combinations of the two factors.
4. To propose a reliable authentication model for combinations through experiment and social validation.

PROPOSED METHODOLOGY

Success percentage rate of any SNS can be improved by performing high security and finding the best feature for group recommendation and restriction. The information on a numerous parameters has collected through a social network sites such as FaceBook on personal account basis. It helps to identify the feature and the policy which involved in that. The present study is in relation to classify the various features to study the causes of such privacy violations which belongs to the process of knowledge discovery and data mining. This information will be helpful

for the academic students and other SNS users to reduce the unwanted users joining in their groups. To perform the above, an improved decision tree algorithm has been proposed. EDT involves incrementally building an ensemble by training each new model instance to emphasize the training instances that previous models mis-classified and omitted. The proposed EDT often leads to a dramatic improvement in predictive accuracy by providing an unordered rule set, when a case is classified; all applicable rules are found and help to calculate the priority. This finally improves both the interpretability of rule sets and their predictive accuracy.

To address the problem of identifying dishonest groups and users in SNSs, the proposed work makes the following contributions:

- The research proposes a new framework named as P-LUG (privacy on Limited User Group) for access restriction process based on the user and profile details.
- This proposes a new privacy and restriction policy specification algorithm to detect dishonest users and groups in SNSs. In specific, users in a SNS can independently set their policies by applying the algorithm; this will find and restrict users in the SNS.
- Using the feature selection process, the algorithm finds the best feature, to rapidly identify the dishonest users and groups; even every group can restrict users from joining by applying the rules and policies.
- Individual policy based decision making evolves the user and user groups in SNS.
- The types of restricted user groups such as adult users, minor users, business users etc will allow users according to their group descriptions. To perform the same, the system uses enhanced decision tree algorithm.
- Finally the system finds the reputation score for every user according to their profile and behavior.
- This provides an analytical study on quantifying the performance of the algorithm, this includes the probability of false positive, probability of false negative the total time taken for calculating reputation score.

Advantages:

- EDT has been shown to yield better accuracy than existing techniques.
- Overcomes the problem of over-fit the training data.
- Overcomes the problem which arises due to the missing values.
- Performs priority based rule selection.

Decision Tree

The nature of group recommendation system (GRS) in LUG is a classification problem. Based on a user's profile features, LUG in SNS finds the most suitable groups for a user and appropriate user for every group. One solution to classification type problem is decision tree algorithm, which is based on binary recursive partitioning. To find better result the proposed work integrated each of splitting rule to LUG based on different features.

This section first presents the model of SNSs and gives the formal definitions on different types of process involved in the proposed system, after that this formalizes the behaviors of users in SNSs on how to provide suggestions and restrictions. Initially the system performs data collection process, feature selection, and created enhanced decision tree.

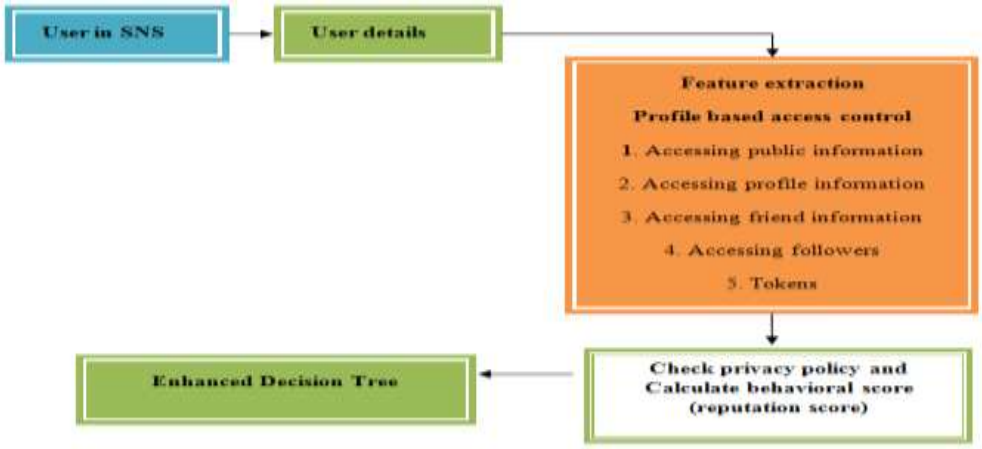


Fig 2 the overall architecture of the proposed system

Figure 2 shows basic architecture of the LUG in SNS using the proposed algorithm. It consists of three components:

i) profile feature extraction,
ii) rule matching
iii) Reputation based decision model.
iv) The study is limited to the student related groups data. Finally, the pre-processed data were transformed into a suitable format to apply data mining techniques. The following fig 3 shows the overall process involved in the proposed system. The first stage in the proposed work is the process initialization, where the data collection, selection and transformation process are done.

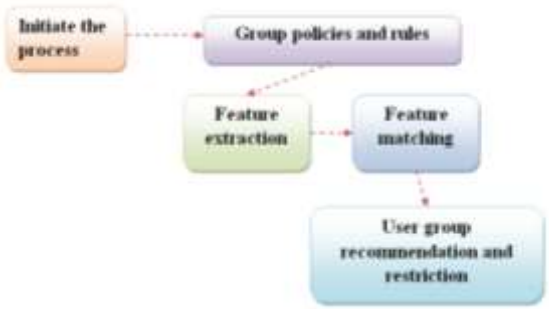


Fig 3 Overall process of proposed system

Algorithm *EDT* (labeled example S, set of variables X)

Input: A set S of labeled examples, a set X of variables.

Output: rule set

1. B = empty.
2. Partition the training data S into C subsets Sc by the class value c or year attribute.
3. for each training data set Sc
   - Compute the Mean M (X_i;X_j) and the Mode □(X_i;X_j) between each starting to end tuple of variables X_i and X_j.
   - Compute W (X_i) for each variable X_i.
   - For all variables X_i in X
     - Add all the variables X_j with W (X_j) > W (X_i) to the parent set □$_{Xi}$ of X_i.
     - Add arcs from all the variables Xj in X_i to X_i.
   - Add the resulting network B_c to B.
4. Return B.

(EDT)

The system develops a new decision tree algorithm using many algorithmetic features. The current study deals the problem of combinatorial optimization problem, which is not handled in the literature. So it created a new algorithm named as EDT, which is based on C5.0 algorithm. The system performs an iterative method to improve the accuracy and efficiency. This also aims at reducing the training complexity in the classification.

The EDT exploits frequent policies and rule set to build accurate EDT method. Consider a labeled structured data set D, a frequency and the rule threshold, and a test case T. EDT exploits the novel approach to estimate, for each class in D. EDT is a novel iterative boosting classifier. It evaluates and selects on-the-fly a set of eligible item sets when a new test case T has to be classified. To allow efficient pattern retrieval, EDT first performs data storage in a disk-based compact data representation. The followings are the major process included in the proposed study.

1. Dataset collection
2. Feature extraction
3. R-UG (Restricting and recommending User group)

In this section, the EDT is thoroughly described. This section presents the training phase of EDT. The test phase of EDT, based on the basic statistical evaluation process.

RESULTS AND DISCUSSION

These section discuses about the implementation and results of the proposed LUG technique using EDT for user group analysis.

Performance of proposed system

Measure the performance of the each iteration then measure the results the overall time of the P-LUG. Member recommendation accuracy is evaluated by comparing the members which assigned by the reputation measures.

User id	Profile based	Friends group based	Token and other
1	3.5	2.3	1.7
2	5.5	4.5	2.9
3	8.4	7.1	3.4
4	12	11	9.6
5	23	20	14.4

Table 1 performance measurement table

In this section, through statistic analysis, the system finds the average time taken for each process in P-LUG. This is initially calculates the individual times such as profile based, interpersonal based and intra personal based. This shows the impact of the amount of user information (user's number of rated items and number of friends) to the accuracy of the proposed model and compared models in e-member.

From the below chart 4 shows the time taken at each stage of P-LUG. The user list is specified from 1 to 5. For each process, the time calculated and finally the system provides the overall time taken by the P-LUG . This shows for 5 Products the system takes 57.4 mille second time.

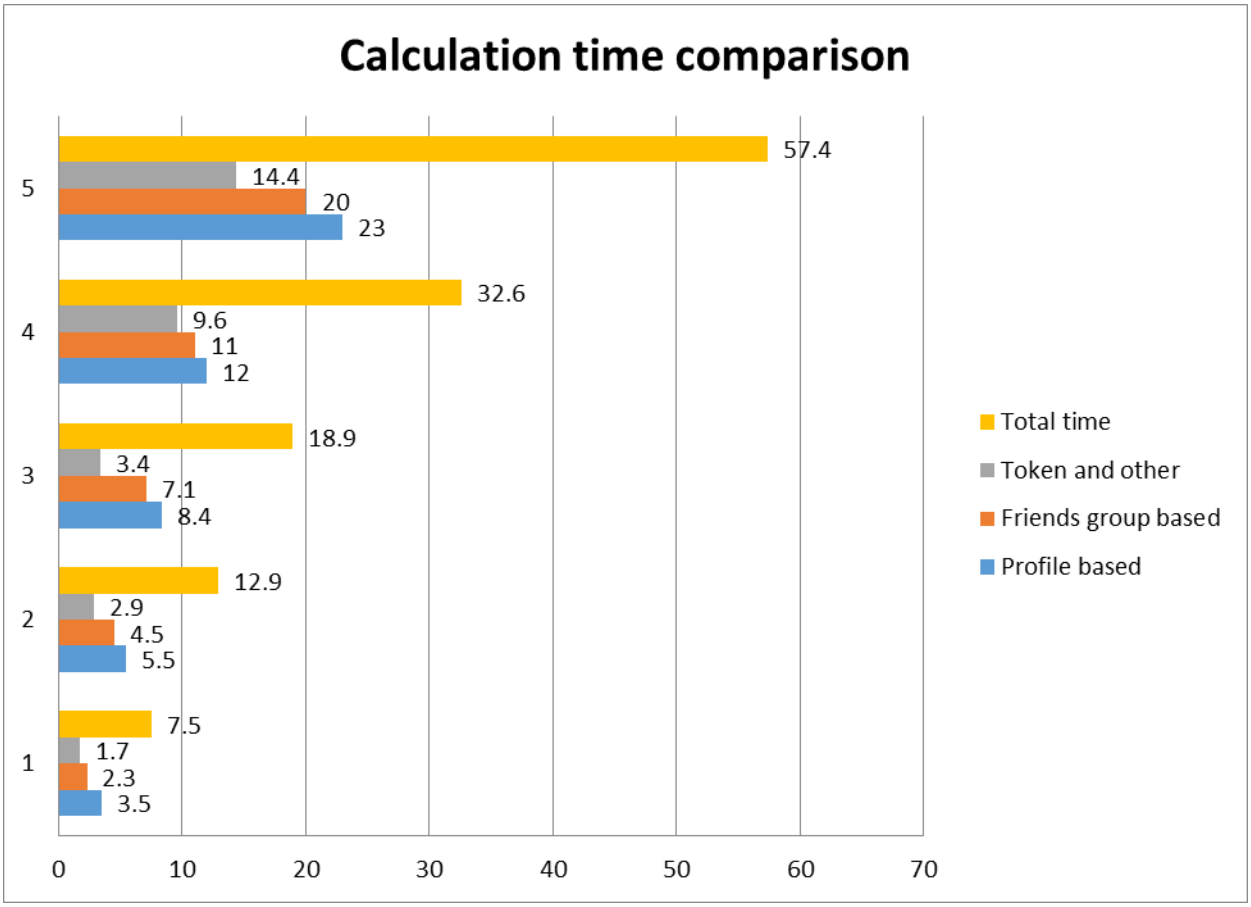


Fig 4 Time comparison calculation

Accuracy calculation:

The system finally performs the analysis to shows (fig 5) the accuracy of the proposed system. Accuracy refers to the proportion of data classified an accurate type in total data, namely the situation TP and TN, thus the accuracy is

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$ * 100%

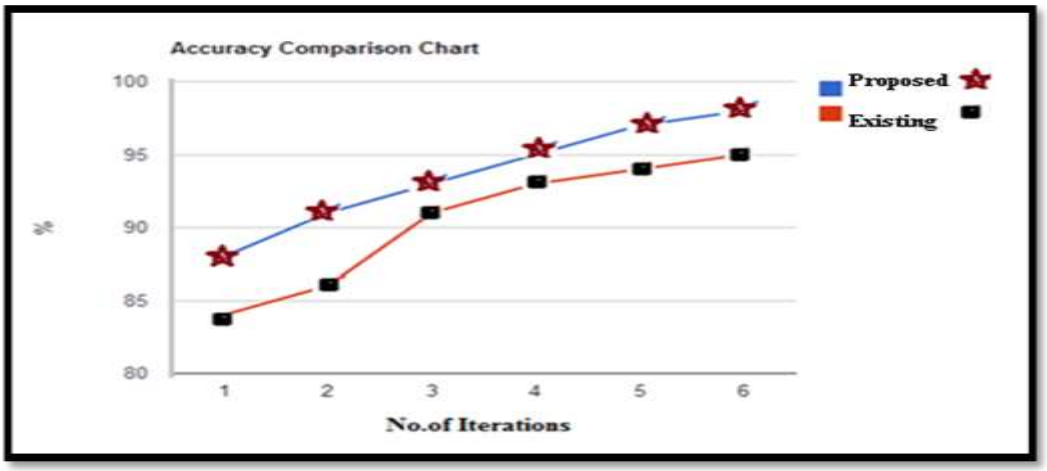


Fig 5 Performance and accuracy analysis

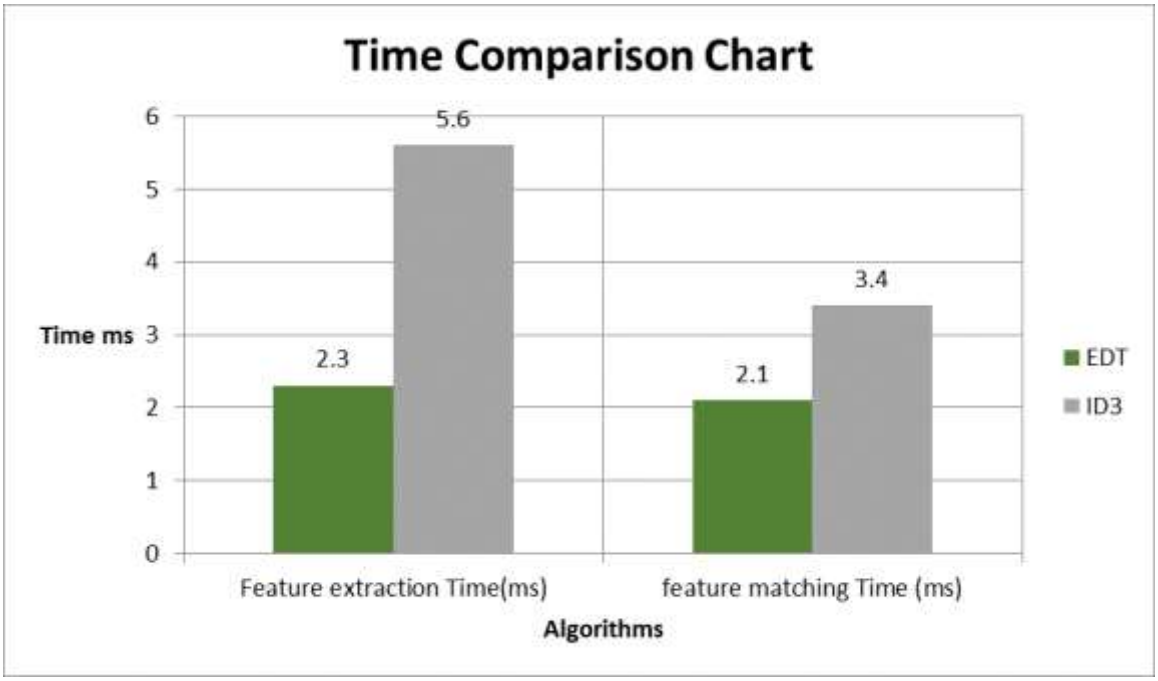


Fig 6 Time comparison between existing ID3 and proposed EDT

The above figure 6 shows the comparison between existing and proposed system based on the Training time. This section compared the training time of the other classification algorithms with the EDT. Below fig 7 shows the precision comparison between existing ID3 and proposed EDT

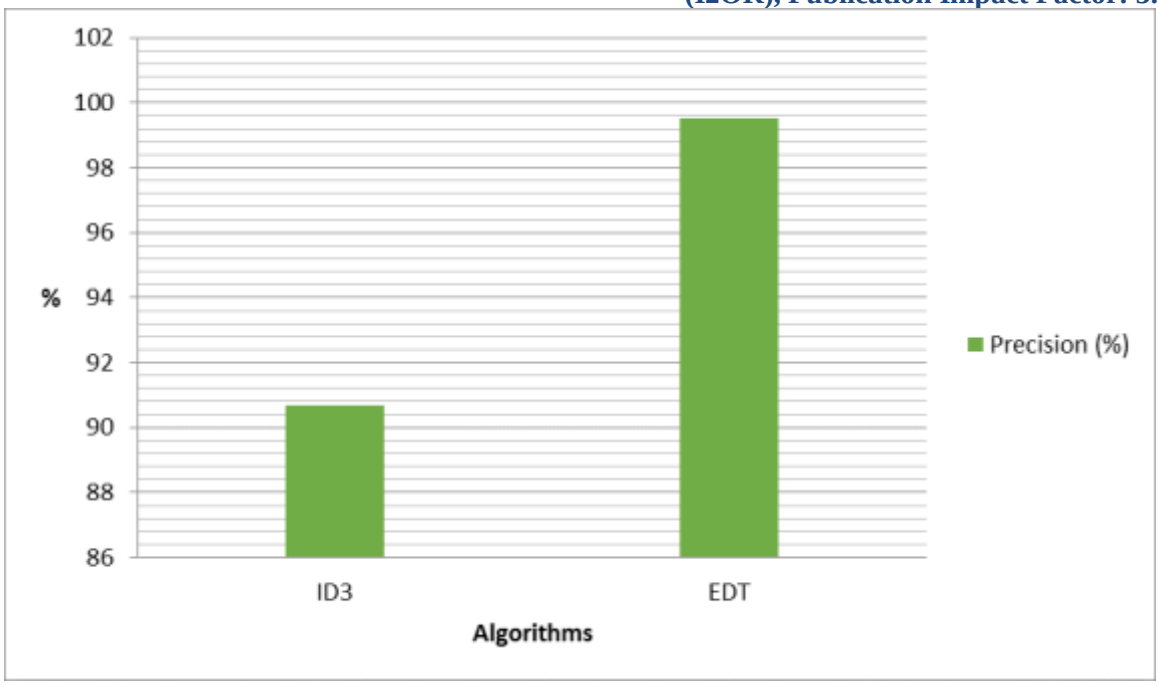


Fig 7 precision comparison between existing ID3 and proposed EDT

CONCLUSION

The main purpose of the study is to investigate the major factors causing the privacy issues in SNS, and best privacy policy matching practice to recommend and restrict users according to the group preference. This also finds the most suitable members without revealing their privacy policies in the current Limited User Group scenario. The proposed system used effective data mining techniques for appropriate policy and rule finding. The system proposed a new approach EDT an improved decision tree with reputation calculation, which facilitates accurate identification of members in SNS who are all suitable for the group. Based on the review of the related literature, synthetic data has been collected and performed the policy matching process using the decision trees.

REFERENCE

[1] Wallbridge, R. (2009). How safe is Your Facebook Profile? Privacy issues of online social networks. *The ANU Undergraduate Research Journal Volume One 2009*, 85.

[2] Azimi, A., & Ghomi, A. A. (2011). Social Networks Privacy Issues That Affect Young Societies. In *International Conference on Emerging Trends in Computer and Image Processing (ICETCIP'2011) Bangkok Dec* (pp. 35-39).

[3] J. M. Kleinberg. Challenges in mining social network data: processes, privacy, and paradoxes. In KDD '07:Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining,pages 4–5, New York, NY, USA, 2007. ACM.

[4] Liccardi, A. Ounnas, R. Pau, E. Massey, P. Kinnunen, S. Lewthwaite, M.-A. Midy, and C. Sarkar. The role of social networks in students' learning experiences. SIGCSE Bull., 39(4):224–237, 2007.

[5] Gharibi, W., & Shaabi, M. (2012). Cyber threats in social networking websites.*arXiv preprint arXiv:1202.2420*.

[6] Stringhini, G., Kruegel, C., & Vigna, G. (2010, December). Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference* (pp. 1-9). ACM.

[7] Ahn, J. (2011). The effect of social network sites on adolescents' social and academic development: Current theories and controversies. *Journal of the American Society for Information Science and Technology*, *62*(8), 1435-1445.

[8] Beato, F., Kohlweiss, M., & Wouters, K. (2009). Enforcing access control in social network sites. *Hot Topics in Privacy*, 1-11.

[9] Hui, P., & Buchegger, S. (2009, July). Groupthink and peer pressure: Social influence in online social network groups. In *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in* (pp. 53-59). IEEE.
[10] Chou, W. Y. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P., & Hesse, B. W. (2009). Social media use in the United States: implications for health communication. *Journal of medical Internet research*, *11*(4).
[11] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook. com.*Social networks*, *30*(4), 330-342.
[12] Rogers, E. M. (1987). Progress, problems and prospects for network research: Investigating relationships in the age of electronic communication technologies.*Social Networks*, *9*(4), 285-310.
[13] Ellison, N. B., Lampe, C., & Steinfield, C. (2009). FEATURE Social network sites and society: current trends and future possibilities. *interactions*, *16*(1), 6-9.
[14] Baruah, T. D. (2012). Effectiveness of Social media as a tool of communication and its potential for technology enabled connections: A micro level study.*International journal of scientific and research publication*, *2*, 5.
[15] Pallis, G., Zeinalipour-Yazti, D., & Dikaiakos, M. D. (2011). Online social networks: status and trends. In *New Directions in Web Data Management 1*(pp. 213-234). Springer Berlin Heidelberg.
[16] Kunegis, J., Lommatzsch, A., & Bauckhage, C. (2009, April). The slashdot zoo: mining a social network with negative edges. In *Proceedings of the 18th international conference on World wide web* (pp. 741-750). ACM. Domingos, P. (2005). Mining social networks for viral marketing. *IEEE Intelligent Systems*, *20*(1), 80-82.