An Effective Supervised & Unsupervised algorithm for OSN Inference Attacks

Mohini Parihar
Dept. of Computer Science & Engineering
Oriental College of Technology
Bhopal, India
mohiniparihar@gmail.com

Amit Dubey
Dept. of Computer Science & Engineering
Oriental College of Technology
Bhopal, India
amitdubey@oriental.ac.in

***Abstract*— The proposed methodology for predicting the links can be computed for time series as well as for the unweighted network. Here in the proposed methodology link prediction can calculated for weighted and unweighted network and for different supervised and unsupervised learning algorithms. From the result analysis it can be concluded that the supervised algorithms performs better as compared to unsupervised algorithms for both weighted and unweighted network. The Proposed Methodology implemented here to prevent Private Information from Various Attacks in Online Social Media Networks can be well predicted as well as performs better as compared to the existing methodology.**

***Keyword* — Online Social Networks (OSNs), Machine Learning, Decision Tree**

I. INTRODUCTION

At the present time, online social networks are attractive, most admired and dominant resources for citizens to join their friends and family members or etc. and also distribute their responsive information. Some of the information in social network is meant to be private and they are said to grow rapidly in near future. These social networks provide many means for sharing information among varied users for different work, the social network data like medical dataset or etc, was known by the social networking websites to the consultant for give some commercial poster of manufactured goods or etc. There is a growing interest in recognizing numerous explanations that exchange letters to a particular entity. First, organizations are interested in correlating user activities and aggregating information across multiple social networks to develop a more complete profile of individual users than the profile provided by any single social network. Second, social networks are interested in finding the entire explanations equivalent to a particular entity inside a particular social network websites. Clients are understood to unlock only one explanation in a social network i.e. Terms of Service; on the other hand some clients generate various accounts. Many unusual categories of social networks here in subsist such as friendship systems, telephone call set-ups, and academia co-authorship set-ups. In recent times the recognition of such online social networks websites is ever-increasing extensively. Such as, social network data could be utilized for advertising manufactured goods to the correct clients. Simultaneously, privacy apprehensions can avoid such attempts in put into practice [1]. They investigate how the online social network data could be utilized to forecast some entity confidential characteristic that a client is not enthusiastic to relate (e.g., political or religious association) and search the consequence of probable data refinement options on avoiding such private information outflow. This variance between aspirations exploit of data and entity privacy currents and opening for social network data mining – specifically, the detection of information and links from social network data. The privacy apprehensions of persons in a social network can be secret into one of two groups: privacy after data release, and private information leakage. But problem [2] of private information outflow for entity as an express consequence of their achievements as being measurement of an online social networks. They concentrated on different concerns associated to private information leakage in social networks. Here they show that using both acquaintance links and features simultaneously presents better predict-ability than details alone. A fundamental step in provided that confirmation about the small-world distinctiveness of OSNs has in recent times been accomplished with the publication of two descriptions by Facebook researchers on the Facebook complete social graph [3]. In addition, we explored the effect of removing details and links in avoiding susceptible information outflow. In the method, they discovered situations in which collective inference does not progress on using an effortless local categorization technique to recognize nodes. When they join the outcomes from the cooperative inference implications with the individual results, we begin to see that removing details and friendship links together is the most excellent way to decrease classifier correctness. This is almost certainly insufficient in sustaining the exploit of social networks. Other papers have tried to infer private information inside social networks. Recently, online social networking sites have exploded in popularity. Numerous sites are dedicated to finding and maintaining contacts and to locating and sharing different types of content.Online social media service [9], they usually are asked to create a profile and to give information about them. In many situations, the data needs to be published and shared with others. The data usually contain valuable information that can enable better social targeting of advertisements.

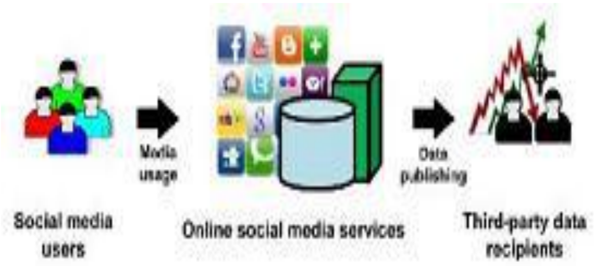

Figure 1. Online social media service [9].

In [3], author think techniques to understand private information using friendship links by creating Bayesian network from the links in the interior a social network websites. While they formulate slow improvement an authentic social network websites, Live Journal, they make use of theoretical characteristics to examine their knowledge algorithm. And also evaluated to [3], they provide performances that can facilitate with choosing the largest part efficient features or links that need to be removed for protecting confidentiality. At last, they try to elaborate the consequence of combined assumption performances in achievable for presumption attacks. Although the final explanation is intrinsic to the configuration of the OSN and to the maximum value they require on the number of make slow progressed customers, the earlier is fundamentally due to the confidentiality situations of the objectives' acquaintances and the OSN go-aheads. Our consequences give you an idea about that homophile in social networks [5] does not only allocate us to assume hidden attributes of OSN users locally, but also allows us to proficiently find the way in the direction of the objective. Note that we do not presume any earlier acquaintance about the network configuration and the user's allocation in the network. Additionally, malicious consumers repeatedly pretend to be honest consumers. For both cases, we necessitate equivalent methods to search the explanations of a single entity. The excessive recognition and quick expansion of these online social networks characterizes a distinctive occasion to learn, recognize, and influence their properties. Not only can an in-strength accepting of online social network arrangement and expansion assist in proposing and estimating existing schemes, it can show the way to enhanced plans of expectations online social network based arrangements and to a profounder appreciative of the collision of online social networks websites on the Internet. online social networks have become extremely admired and have become a explanation feature in today‘s knowledge. A consumer engaged in a social network converses with many public who may be either identified or unidentified. The resources that are distributed by the consumer in a social networking media are said to be extremely unconfident and are challenge on open attacks. Despite the fact that these networks make an effort to give protection and confidentiality over these resources still there are no ideal explanations for this difficulty. The most important difficulty happens when reserves are distributed among many citizens. There are no protected strategies offered to explain problems associated to several shared resources. Earlier research still has focused on particular social networks and did not examine social networks in comprehensive. Such as, researchers have considered the performance of consumers on Facebook or LinkedIn independently [6]. This only make available but an incomplete observation of a consumer. Mainly communications in Facebook will most probable only distinguish communications with friends and contacts in LinkedIn will most probable only distinguish interactions with colleagues. Expressive the identical accounts on several social networks make available the occasion to put together an enhanced representation of a consumer. Having a deeper appreciative of a consumer can subsequently show the way to enhanced personalized services or enhanced evaluations of consumers' proficiency. Various online social networks also present many valuable properties that can be influenced to improve information schemes, such as developments to organizing information proliferation novel paths for information exploration, repossession and innovative means of way of thinking about expectation. A few research efforts proposed distributed social networks to keep away from giving up confidential information to corporation [7], [8]. While these explanations will save from harm the personal data of users from big organization, it is still probable for a third-party to counterpart the explanations customers have on different private social networks.

II.LITERATURE SURVEY

In this paper [2], author has tried to discover how to open inference attacks using unrestricted social networking data to forecast private information. Here they give you an idea about that using both friendship links and features mutually gives enhanced forecast-capability than aspects by yourself. They also survey on the cause of eliminating details and links in avoiding responsive information leakage. In this development, they tried to exposed circumstances in which combined inference does not progress on using an effortless local categorization technique to recognize nodes. When they combine the consequences from the combined assumption suggestions with the entity outcomes they begin to observe that eliminating features and friendship links mutually is the most excellent way to extract classifier correctness. This is possibly infeasible in sustaining the exploit of social networks. Then, they also investigate the use of these methods and effort to exploit techniques of combined assumption to determine susceptible characteristics of the data set.

However, author [2] has also given you an idea about that by eliminating only features; they significantly diminish the correctness of local classifiers, which provide us the highest precision that we were proficient to accomplish through any arrangement of classifiers. They also demonstrate that they can reduce the efficiency of both local and relational categorization algorithms by using the sanitization techniques.

In this paper author [10], propose new techniques of social graph anonymization, center of attention only the design that by anonymizing both the nodes in the collection and the link arrangement, that one in that way anonymizes the graph as a complete. On the other hand, their process all center of attention on anonymity in the arrangement itself. Such as, all the way through the exploit of k-anonymity or t-closeness, depending on the quasi-identifiers which are chosen, much of the uniqueness in the data may be lost. The suggested technique of anonymity preservation, they sustain the complete distinctiveness in each node which allocates more information in the data post-release.

Here author Gross et al. [11], study definite procedure illustrations at Carnegie Mellon. Here author has also make a note of potential attacks, i.e. node re-identification or haunting that simply easy to get to data on Facebook could help with that procedure. They additional communication that at the same time as privacy controls may subsist on the consumer's end of the social networking websites a lot of entities do not acquire improvement of this instrument. To searching agrees very well with the sum of data that they were proficient to make slow progress using a much uncomplicated crawler on a Facebook network. They enlarge on their effort by experimentally investigative the correctness of some kinds of the demographic re-identification that they suggest before and after data refinement.

In this paper author has select the region of privacy inside a social network includes a huge breadth based on how privacy is characterized. Here they find the Anonymized Social Networks [12] think about an attack aligned with an anonymized network. In their representation, the network consists of only nodes and boundaries here feature values are not contained. The objective of the attacker is just to recognize citizens. Additional, their difficulty is extremely unusual than the one measured for the reason that they pay no attention to features and do not think about the consequence of the continuation of features on privacy.

In this papers author have tried to infer private information inside social network by Inference Attacks by Third-Party Extensions to Social Network Systems [13] recognize the hazard of social networks website API inference attacks, make available a arrangement of these attacks and propose a possibility evaluation method to facilitate users recognize the hazard of promising to a third-party application in an extensible SNS. Using that expansion of the metric to explanation for irregular attractiveness of confirmation inquiries and the propose of a secure API for extensible SNSs. They also generate a point of reference originate the achievability predicates and empirically evaluate the inference accurateness of the inference algorithms in the standard. This would permit us to empirically estimate the efficiency of the hazard evaluation method. One drawback of the hazard evaluation method is that it imagines all confirmation inquiries in the standard are uniformly accepted. An enhancement is to re-formulate the metric so that it acquires into explanation the irregular attractiveness of the confirmation problems. Here author find an interesting research problem would be to verify which description of the hazard metric is essentially more efficient in guiding consumers' confidentiality anticipations.

Due to certain condition, author has to find an efficient tag validation method is employed for authenticating the labeled consumer against the photo. In this paper here author it informs that if any consumers label themselves or additional in a photo then the photo holder will get obtain a tag notification. In such condition data owner will come to recognize about the exactness of the tagged consumers. Here author has used Facial recognition method [14] to identify people precisely in contents for example: photos, repeated tag validation is viable. Considering another condition, here author has informs about the possible authorization contact regarding an organizer's privacy first choice. By means of this purpose the photo owner will observe the consumers who are established to right to use the photo by the mutual permission which is not unambiguously approved by the data owner. At last the data owner can find out malicious actions in mutual control.

Here author Carminati [15] has initiated a new protection strategy for mutual right to use control in online social networks that fundamentally improve topology-based right to use control regarding a set of mutual clients. In this paper, author has proposed a new formal model is used for concentrate on the multiparty right to use control concern in Online Social Network Systems, along with guiding principle arrangement method and elastic disagreement decision method for mutual administration of mutual data in OSNs. Here proposed work can also accomplish an assortment of investigation undertakings on right to use control methods exploited in Online Social Network Systems, which is not concentrate on by earlier work.

III. PROPOSED METHODOLOGY

The link prediction problem in social network is a major issue as it totally depends that how efficiently and exactly the "future relationships" between two nodes in a social network can be predicted.

In 2011, Hially. R and Ricardo B.C , implemented "supervised learning algorithms" – Niave Bayes , J48 , IBk , and SVM on “co-authorship networks datasets” to compare that Link Prediction on unweighted networks have less performance than weighted networks, except in some case studies where weighted graph network taken in different has less precision rate then unweighted network in Naive bayes algorithm.

a. First we will select a dataset of an online social network and make out unweighted and weighted datasets from it.
b. We will make different versions of the weighted network datasets by taking in concern that our dataset should be able to increase accuracy.
c. We will implement our datasets using WEKA software.
d. Then we will apply different classification algorithms such as Niave Bayes , J48 , IBk , and SVM (all are present in WEKA environment) on both the weighted and unweighted social networks and will formulate our research on several parameters such as Accuracy , Precision and recall rate , and AREA under ROC curve and F-Measure obtained by each algorithm on different datasets.
e. We will compare the results generated and will come to an expected conclusion that weighted networks having temporal information in link prediction problem has more performance (with high accuracy) than unweighted networks having temporal information and another result will be that supervised strategy always results in accurate outcomes in link prediction problem.

As shown in the given figure is the outline of the proposed methodology used. The proposed methodology can be applied on the movie lens dataset which contains a set of attributes such as user is, tag id, time stamp, rating. Here in the proposed methodology the performance of supervised learning algorithm is shown.

3.1 Movie Lens Dataset

Here we consider two types of datasets weighted and unweighted. The weighted dataset contains a set of attributes which denotes a set of links between two nodes and their weights such as,

User ID	User ID1	weight	Movie ID	rating	timestamp

Here in this type of dataset 100 users is taken along with their links with weights. The non-weighted dataset contains a set of attributes such as,

User ID	User ID1	Movie ID	rating	timestamp

3.2 Applying Similarity Metrics

For the prediction of connecting links on the weighted and unweighted network, similarity metrics is one of the techniques to measure similarity between different nodes so that the similarity of different nodes can be detected. Here in the proposed methodology various similarity metrics have been applied to measure the similarity of weighted and unweighted dataset.

Number of Common Neighbors

It is defined as the total number of nodes that are connected directly in relationship with node x and y for unweighted network,

$$CN(x,y) = \varphi(x) \cap \varphi(y)$$

Where, $\varphi(x)$x) is the set of neighbors of nodes x.

$\varphi(y)$ is the set of neighbor of node y.

To calculate link prediction between nodes for unweighted network common neighbors can be calculated as,

$$CN(x,y) = \sum_{z\in\varphi(x)\cap\varphi(y)} w(x,z) + w(y,z)$$

Jaccard Coefficient

It is defined as the highest proportion of common neighbors to the total number of neighbors in the network. The jaccard coefficient can also defined for weighted as well for unweighted network. For unweighted network,

$$JC(x,y) = \frac{\varphi(x) \cap \varphi(y)}{\varphi(x) \cup \varphi(y)}$$

For weighted network,

$$JC(x,y) = \sum_{z\in\varphi(x)\cap\varphi(y)} \frac{w(x,z) + w(y,z)}{\sum_{a\in\varphi(x)} w(a,x) + \sum_{b\in\varphi(y)} w(b,y)}$$

Forecasting Models

Forecasting models are used for the future prediction based on the previous and present data analysis.

Moving Average

It can be defined as the mean of all the recent values that are prediction from the dataset very currently. It can be given as,

$$\widehat{X_t} = \frac{X_{t-1} + X_{t-2} + \cdots \ldots \ldots + X_{t-n}}{n}$$

Where, Xt is the time series forecasting

t=1,2….T be the time series.

Average

It can be defined as the mean of all the past values that are predicted.

$$\widehat{X_t} = \frac{X_{t-1} + X_{t-2} + \cdots \ldots \ldots + X_{t-T}}{T}$$

3.3 Supervised Learning Algorithm

Here various supervised learning algorithms are applied on the weighted and unweighted algorithm to test the performance of the link prediction. Here various supervised learning algorithms such as ID3, SVM, Random Forest, Simple Cart algorithms are implemented.

3.4 Unsupervised Learning Algorithms

Here various unsupervised learning algorithms are applied on the weighted and unweighted algorithm to test the performance of the link prediction. Here various unsupervised learning algorithms such as k-means, Hierarchical, Expected Maximization, Cob-web algorithms are implemented. As shown in the figure below is the proposed outline of our work.

Flow Chart

The Proposed Methodology implemented here for the security from Inference Attacks. First we will select a dataset of an online social network and make out unweighted and weighted datasets from it. We will make different versions of the weighted network datasets by taking in concern that our dataset should be able to increase accuracy. We will implement our datasets using WEKA software. Then we will apply different classification algorithms such as Niave Bayes , J48 , IBk , and SVM (all are present in WEKA environment) on both the weighted and unweighted social networks and will formulate our research on several parameters such as Accuracy , Precision and recall rate , and AREA under ROC curve and F-Measure obtained by each algorithm on different datasets.

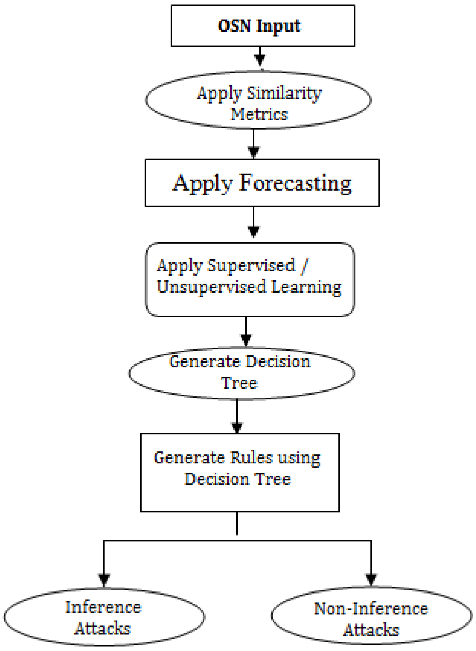

Fig. 2. Flow Chart of the Proposed Methodology

IV. RESULT ANALYSIS

The table shown below is the analysis and comparison of existing Naïve Bayes Algorithm and the proposed methodology. The result analysis shows the performance of the proposed methodology on various datasets. The proposed methodology provides higher precision in comparison. The table shown below is the analysis and comparison of existing Naïve Bayes Algorithm and the proposed methodology. The result analysis shows the performance of the proposed methodology on various datasets. The proposed methodology provides higher recall in comparison.

Table (1) Analysis of Precision

	Precision	
Datasets	**Naïve Bayes**	**SVM**
D-1	92.13	94.6
D-2	91.65	95.4
D-3	92.94	94.3
D-4	90.16	94.7
D-5	91.39	95

Table (2) Analysis of Recall

	Recall	
Datasets	**Naïve Bayes**	**SVM**
D-1	89.56	93.65
D-2	90.16	93.67
D-3	90.65	92.53
D-4	89.37	93.49
D-5	91.83	92.85

The figure shown below is the analysis and comparison of existing Naïve Bayes Algorithm and the proposed methodology. The result analysis shows the performance of the proposed methodology on various datasets. The proposed methodology provides higher precision in comparison.

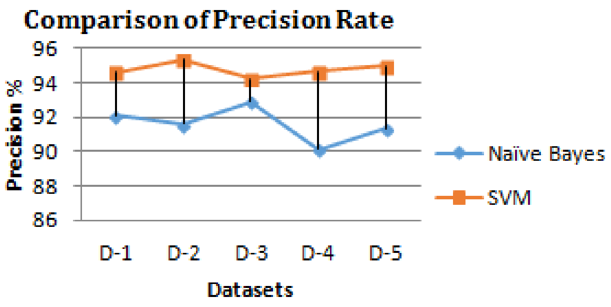

Fig 3. Comparison of Precision

The figure shown below is the analysis and comparison of existing Naïve Bayes Algorithm and the proposed methodology. The result analysis shows the performance of the proposed methodology on various datasets. The proposed methodology provides higher recall in comparison.

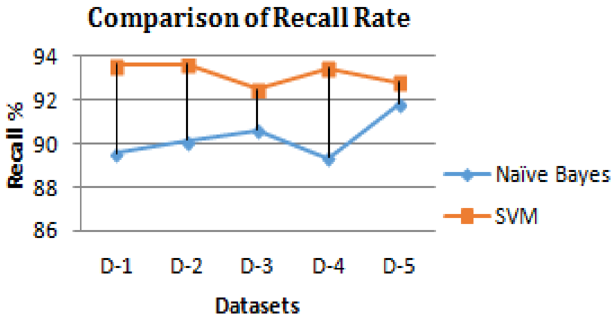

Fig. 4. Comparison of Recall

V. CONCLUSION

One fundamental issue in today's Online Social Networks (OSNs) is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. Up to now, OSNs provide little support to this requirement. To fill the gap, we propose a system allowing OSN users to have a direct control on the messages posted on their walls. This is achieved through an algorithm that includes clustering, classification and filtering and flexible rule-based system that allows user to customize the filtering criteria to be applied to their walls, and a Machine Learning-based soft classifier automatically labelling messages in support of content-based filtering.

REFERENCE

[1]. L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore Art Thou Anonymized Social Networks, Hidden Patterns, and Structural Steganography," Proc. 16th Int'l Conf. World Wide Web (WWW '07), pp. 181-190, 2007.

[2]. Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham, "Preventing Private Information Inference Attacks on Social Networks" IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 8, August 2013.

[3]. Backstrom, L., Boldi, P., Rosa, M., Ugander, J., Vigna, S.: Four degrees of separation. In: Proceedings of the 3rd Annual ACM Web Science Conference, 2011.

[4]. J. He, W. Chu, and V. Liu, "Inferring Privacy Information from Social Networks," Proc. Intelligence and Security Informatics, 2006.

[5]. Aiello, L.M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., Menczer, F.:Friendship prediction and homophily in social media. ACM Transactions on theWeb, 2012.

[6]. Valerio Arnaboldi, Marco Conti, Andrea Passarella, and Robin Dunbar. Dynamics of personal social relationships in online social networks: A study on twitter. In Proceedings of the First ACM Conference on Online Social Networks, COSN '13, pages 15–26, New York, NY, USA, 2013.

[7]. Luca Maria Aiello and Giancarlo Ruffo. Lotusnet: Tunable privacy for distributed online social network services. Computer Communications, 35(1):75–88, 2012.

[8]. Emiliano De Cristofaro, Claudio Soriente, Gene Tsudik, and AndrewWilliams. Hummingbird: Privacy at the time of twitter. In Proceedings of the 2012 IEEE Symposium on Security and Privacy, SP '12, pages 285–299, Washington, DC, USA, 2012. IEEE Computer Society.

[9]. Chethana Nair, Neethu Krishna, Siby Abraham, "Generalization Algorithm For Prevent Inference Attacks In Social Network Data" IJRCCT, 2014.

[10]. E. Zheleva and L. Getoor, "Preserving the Privacy of Sensitive Relationships in Graph Data," Proc. First ACM SIGKDD Int'l Conf. Privacy, Security, and Trust in KDD, pp. 153-171, 2008.

[11]. R. Gross, A. Acquisti, and J.H. Heinz, "Information Revelation and Privacy in Online Social Networks," Proc. ACM Workshop Privacy in the Electronic Soc. (WPES '05), pp. 71-80, http:// dx.doi.org/10.1145/1102199.1102214, 2005.

[12]. Backstrom, c. dwork, and j. Kleinberg, "Wherefore Art Thought: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," Proc. 16th Int'l Conf. World Wide Web (WWW '07), pp. 181-190 (2010).

[13]. Seyed Hossein Ahmadinejad, mohd anwar and philip w. l. fong, "Inference Attacks by Third-Party Extensions to Social Network Systems" (2010).

[14]. J. Choi, W. De Neve, K. Plataniotis, and Y. Ro, "Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks," IEEE Trans. Multimedia, vol. 13, no. 1, pp. 14-28, Feb. 2011.

[15]. B. Carminati and E. Ferrari, "Collaborative Access Control in Online Social Networks," Proc. Seventh Int'l Conf. Collaborative Computing: Networking, Applications and Work sharing (Collaborate-Com), pp. 231-240, 2011.

Author's Profile

(1) Name of Author - Ms. Mohini Parihar, M-Tech Sholar in Dept. of Computer Science & Engineering Oriental College of Technology Bhopal, India

(2) Name of Author - Mr. Amit Dubey, Asst Professor in Dept. of Computer Science & Engineering Oriental College of Technology Bhopal, India