Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP

Ms. Beri Keerthi

Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract*— *At present social network sites are part of the life for most of the people. Every day several people are creating their profiles on the social network platforms and they are interacting with others independent of the user's location and time. The social network sites not only providing advantages to the users and also provide security issues to the users as well their information. To analyze, who are encouraging threats in social network we need to classify the social networks profiles of the users. From the classification, we can get the genuine profiles and fake profiles on the social networks. Traditionally, we have different classification methods for detecting the fake profiles on the social networks. But we need to improve the accuracy rate of the fake profile detection in the social networks. In this paper we are proposing Machine learning and Natural language Processing (NLP) techniques to improve the accuracy rate of the fake profiles detection. We can use the Support Vector Machine (SVM) and Naïve Bayes algorithm.

I. INTRODUCTION

Social networking has end up a well-known recreation within the web at present, attracting hundreds of thousands of users, spending billions of minutes on such services. Online Social network (OSN) services variety from social interactions-based platforms similar to Facebook or MySpace, to understanding dissemination-centric platforms reminiscent of twitter or Google Buzz, to social interaction characteristic brought to present systems such as Flicker. The opposite hand, enhancing security concerns and protecting the OSN privateness still signify a most important bottleneck and viewed mission

When making use of social network's (SN's), one-of-a-kind men and women share one-of-a-kind quantities of their private understanding. Having our individual know-how entirely or in part uncovered to the general public, makes us excellent targets for unique types of assaults, the worst of which could be identification theft. Identity theft happens when any individual uses character's expertise for a private attain or purpose. During the earlier years, online identification theft has been a primary problem considering it affected millions of people's worldwide. Victims of identification theft may suffer unique types of penalties; for illustration, they would lose time/cash, get dispatched to reformatory, get their public image ruined, or have their relationships with associates and loved ones damaged. At present, the vast majority of SN's does no longer verifies ordinary users‟ debts and has very susceptible privateness and safety policies. In fact, most SN's applications default their settings to minimal privateness; and consequently, SN's became a best platform for fraud and abuse. Social Networking offerings have facilitated identity theft and Impersonation attacks for serious as good as naive attackers. To make things worse, users are required to furnish correct understanding to set up an account in Social Networking web sites. Easy monitoring of what customers share on-line would lead to catastrophic losses, let alone, if such bills had been hacked.

Profile information in online networks will also be static or dynamic. The details which can be supplied with the aid of the person on the time of profile creation is known as static knowledge, the place as the small print that are recounted with the aid of the system within the network is called dynamic knowledge. Static knowledge includes demographic elements of a person and his/her interests and dynamic knowledge includes person runtime habits and locality in the network. The vast majority of current research depends on static and dynamic data. However, this isn't relevant to lots of the social networks, where handiest some of static profiles are seen and dynamic profiles usually are not obvious to the person network. More than a few procedures have been proposed by one-of-a-kind researchers to realize the fake identities and malicious content material in online social networks. Each process had its own deserves and demerits.

The problems involving social networking like privacy, on-line bullying, misuse, and trolling and many others. Are many of the instances utilized by false profiles on social networking sites. False profiles are the profiles which are not specific i.e. They're the profiles of men and women with false credentials. The false Facebook profiles more commonly are indulged in malicious and undesirable activities, causing problems to the social community customers. Individuals create fake profiles for social engineering, online impersonation to defame a man or woman, promoting and campaigning for a character or a crowd of individuals. Facebook has its own security system to guard person credentials from spamming, phishing, and so on.

And the equal is often called Facebook Immune system (FIS). The FIS has now not been ready to observe fake profiles created on Facebook via customers to a bigger extent.

II. LITERATURE SURVEY

Understanding User Profiles on Social Media for Fake News Detection

Kai Shu, Suhang Wang, Huan Liu - 2018

Consuming news from social media is becoming increasingly popular nowadays. Social media brings benefits to users due to the inherent nature of fast dissemination, cheap cost, and easy access. However, the quality of news is considered lower than traditional news outlets, resulting in large amounts of fake news. Detecting fake news becomes very important and is attracting increasing attention due to the detrimental effects on individuals and the society. The performance of detecting fake news only from content is generally not satisfactory, and it is suggested to incorporate user social engagements as auxiliary information to improve fake news detection. Thus it necessitates an in-depth understanding of the correlation between user profiles on social media and fake news.

In this paper, we construct real-world datasets measuring users trust level on fake news and select representative groups of both "experienced" users who are able to recognize fake news items as false and "naïve" users who are more likely to believe fake news. We perform a comparative analysis over explicit and implicit profile features between these user groups, which reveals their potential to differentiate fake news. The findings of this paper lay the foundation for future automatic fake news detection research.

Identifying Fake Profiles In Linkedin

Shalinda Adikari, Kaushik Dutta - 2019

As organizations increasingly rely on professionally oriented networks such as LinkedIn (the largest such social network) for building business connections, there is increasing value in having one's profile noticed within the network. As this value increases, so does the temptation to misuse the network for unethical purposes. Fake profiles have an adverse effect on the trustworthiness of the network as a whole, and can represent significant costs in time and effort in building a connection based on fake information. Unfortunately, fake profiles are difficult to identify.

Approaches have been proposed for some social networks; however, these generally rely on data that are not publicly available for LinkedIn profiles. In this research, we identify the minimal set of profile data necessary for identifying fake profiles in LinkedIn, and propose an appropriate data mining approach for fake profile identification. We demonstrate that, even with limited profile data, our approach can identify fake profiles with 87% accuracy and 94% True Negative Rate, which is comparable to the results obtained based on larger data sets and more expansive profile information. Further, when compared to approaches using similar amounts and types of data, our method provides an improvement of approximately 14% accuracy.

A Feature Based Approach to Detect Fake Profiles in Twitter

Jyoti Kaubiyal, Ankit Kumar Jain - 2019

Social networking platforms, particularly sites like Twitter and Facebook have grown tremendously in the past decade and has solicited the interest of millions of users. They have become a preferred means of communication, due to which it has also attracted the interest of various malicious entities such as spammers. The growing number of users on social media has also created the problem of fake accounts. These false and fake identities are intensively involved in malicious activities such as spreading abuse, misinformation, spamming and artificially inflating the number of users in an application to promote and sway public opinion. Detecting these fake identities, thus becomes important to protect genuine users from malicious intents. To address this issue, we aim to use a feature-based approach to identify these fake profiles on social media platforms. We have used twenty-four features to identify fake accounts efficiently. To verify the classification results three classification algorithms are used. Experimental results show that our model was able to reach 97.9% accuracy using the Random Forest algorithm. Hence, the proposed approach is efficient in detecting fake profiles.

Method for Detecting Spammers and Fake Profiles in Social Networks

Yuval Elovici, Michael FIRE, Gilad Katz - 2019

A method for protecting user privacy in an online social network, according to which negative examples of fake profiles and positive examples of legitimate profiles are chosen from the database of existing users of the social network. Then, a predetermined set of features is extracted for each chosen fake and legitimate profile, by dividing the friends or followers of the chosen examples to communities and analyzing the relationships of each node inside and between the communities. Classifiers that can detect other existing fake profiles according to their features are constructed and trained by using supervised learning.

Social Networks Fake Profiles Detection Using Machine Learning Algorithms

Yasyn Elyusufi, Zakaria Elyusufi - 2020

Fake profiles play an important role in advanced persisted threats and are also involved in other malicious activities. The present paper focuses on identifying fake profiles in social media. The approaches to identifying fake profiles in social media can be classified into the approaches aimed on analysing profiles data and individual accounts. Social networks fake profile creation is considered to cause more harm than any other form of cybercrime. This crime has to be detected even before the user is notified about the fake profile creation. Many algorithms and methods have been proposed for the detection of fake profiles in the literature. This paper sheds light on the role of fake identities in advanced persistent threats and covers the mentioned approaches of detecting fake social media profiles. In order to make a relevant prediction of fake or genuine profiles, we will assess the impact of three supervised machine learning algorithms: Random Forest (RF), Decision Tree (DT-J48), and Naïve Bayes (NB).

Problem Statement

There are lots of issues that make this procedure tough to implement and one of the biggest problems associated with fraud detection is the lack of both the literature providing experimental results and of real-world data for academic researchers to perform experiments on. The reason behind this is the sensitive financial data associated with the fraud that has to be kept confidential for the purpose of customer's privacy. Now, here we enumerate different properties a fraud detection system should have in order to generate proper results:

The system should be able to handle skewed distributions, since only a very small percentage of all credit card transactions are fraudulent.

There should be a proper means to handle the noise. Noise is the errors that is present in the data, for example, incorrect dates. This noise in actual data limits the accuracy of generalization that can be achieved, irrespective of how extensive the training set is.

Another problem related to this field is overlapping data. Many transactions may resemble fraudulent transactions when actually they are genuine transactions. The opposite also happens, when a fraudulent transaction appears to be genuine.

The systems should be able to adapt themselves to new kinds of fraud. Since after a while, successful fraud techniques decrease in efficiency due to the fact that they become well known because an efficient fraudster always find a new and inventive ways of performing his job.

There is a need for good metrics to evaluate the classifier system. For example, the overall accuracy is not suited for evaluation on a skewed distribution, since even with a very high accuracy; almost all fraudulent transactions can be misclassified.

Disadvantages:

- The most of existing methods has ignored the poor-quality data like noise or Feature handled complex.
- The problems involving social networking like privacy, on-line bullying, misuse, not accurate analysis and trolling and many others.
- There are many of the instances utilized by false profiles on social networking sites.
- False profiles are the profiles which are not specific i.e, They're the profiles of men and women with false credentials.

III. METHODOLOGY

Proposed Work: A proper and thorough literature survey concludes that there are various methods that can be used to detect Fake profile detection. Some of these approaches are Machine Learning and NLP.

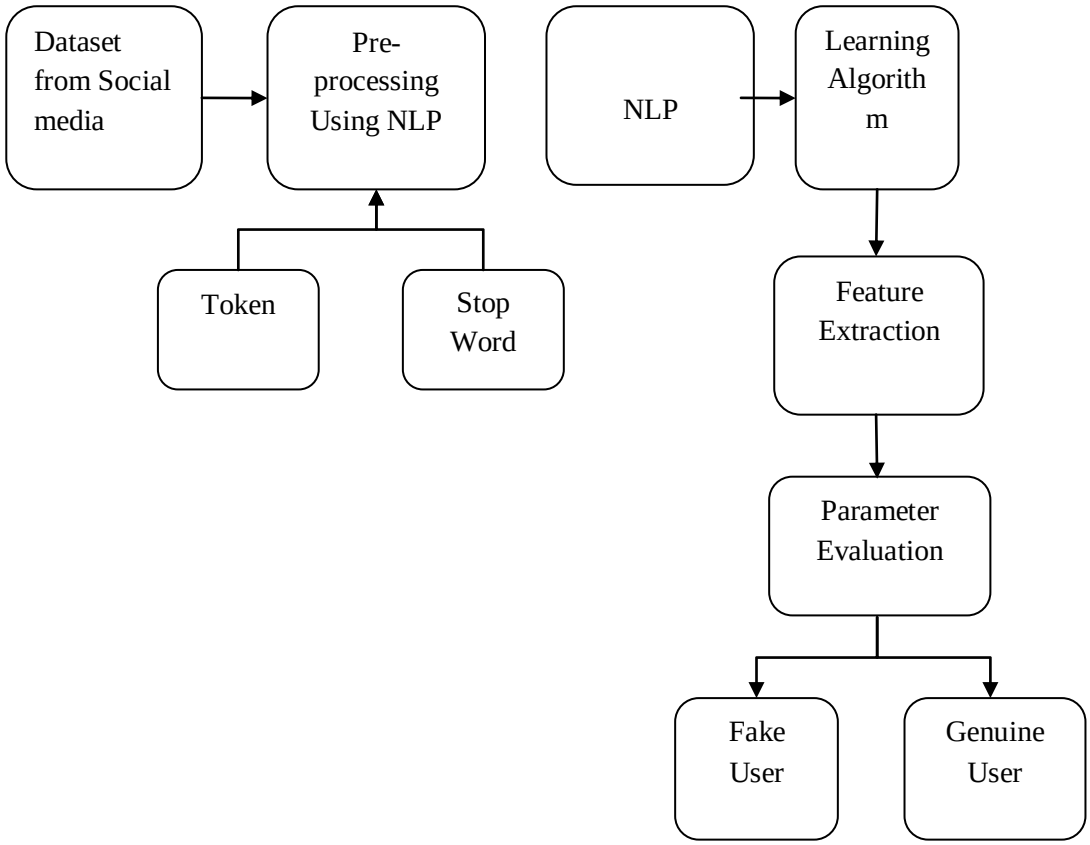
To analyze, who are encouraging threats in social network we need to classify the social networks profiles of the users. From the classification, we can get the genuine profiles and fake profiles on the social networks. Traditionally, we have different classification methods for detecting the fake profiles on the social networks. But we need to improve the accuracy rate of the fake profile detection in the social networks.

On this paper we presented a machine learning & natural language processing system to observe the unreliable users in on-line social networks. Moreover, we are adding the SVM classifier algorithm to increase the detection accuracy rate of the fake profiles.

In our research paper, as stated earlier, we will be emphasizing on the SVM algorithm and how it is used in fake news detection systems.

Advantages

- High accuracy is obtained and time consumption for detecting the Fake profiles.
- More datasets are included.
- We can find the all types of profiles on different social media application also.



The presented process used Facebook profile to notice false profiles. The working method of the proposed procedure includes three principal phases;

1. NLP Pre-processing
2. Principal Component Analysis (PCA)
3. Learning Algorithm

3.1 Data collection

To collect the Dataset from Facebook profiles. Collecting data for training the ML model is the basic step in the machine learning pipeline. The predictions made by ML systems can only be as good as the data on which they have been trained. Following are some of the problems that can arise in data collection:

- **Inaccurate data:** The collected data could be unrelated to the problem statement.
- **Missing data:** Sub-data could be missing. That could take the form of empty values in columns or missing images for some class of prediction.
- **Data imbalance:** Some classes or categories in the data may have a disproportionately high or low number of corresponding samples. As a result, they risk being under-represented in the model.
- **Data bias:** Depending on how the data, subjects and labels themselves are chosen, the model could propagate inherent biases on gender, politics, age or region, for example. Data bias is difficult to detect and remove.

3.2 Pre-Processing

Once the data is extracted from the twitter source as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like stop words, emoticons in order to make sure that non textual content is identified and removed before the analysis.

Text pre-processing is an essential a part of any NLP method and the significance of the NLP pre-processing are

- To minimize indexing (or knowledge) records dimension of the textual content records
    1. Stop words bills 20-30% of total phrase counts in a special textual content record
    2. Stemming may just diminish indexing size as much as forty- 50%
- To make stronger the efficiency and effectiveness of the IR method
    1. Stop words aren't valuable for shopping or textual content mining
    2. Stemming used for matching the similar words in a text record

3.2.1 Tokenization

Tokenization is the process of breaking a circulate of textual content into phrases, phrases, symbols, or different significant factors called tokens. The aim of the tokenization is the exploration of the phrases in a sentence. The list of tokens turns into input for further processing akin to parsing or textual content mining. Tokenization is valuable both in linguistics (where it's a form of textual content segmentation), and in laptop science, the place it forms a part of lexical analysis. Textual knowledge is simplest a block of characters at the starting.

All strategies in know-how retrieval require the words of the data set. For that reason, the requirement for a parser is a tokenization of records. This might be sound trivial because the text is already saved in computing device-readable codecs. However, some problems are nonetheless left, like the removing of punctuation marks. Different characters like brackets, hyphens, and so on require processing as well.

3.2.2 Stop word Removal

Stop phrases are very more often than not used fashioned phrases like ‘and’, ‘are’, ‘this’ etc. They don't seem to be useful in classification of records. So, they must be removed. However, the development of such stop phrases record is problematic and inconsistent between textual sources. This process also reduces the text knowledge and improves the approach performance. Each textual content report offers with these phrases which are not vital for text mining applications.

3.2.3 Stemming and Lemmatization

The aim of both stemming as well as lemmatization is to scale down inflectional types & mostly derivationally associated varieties of a phrase to a fashioned base kind.

Stemming usually refers to a crude heuristic process that chops off the ends of words in the hope of accomplishing this goal accurately more often than not, and quite often involves the removal of derivational affixes.

Lemmatization often refers to doing matters competently with the usage of a vocabulary and morphological analysis of phrases, in most cases aiming to eliminate inflectional endings only and to come back the base or dictionary type of a word, which is often called the lemma.

3.3 PCA

Principal Component Analysis purpose is to extract the fundamental understanding from the table, to symbolize it as a suite of new orthogonal variables known as major accessories, and to show the sample of similarity of the observations and of the variables as elements in maps.

3.4 Train the Model using Algorithm

In this proposed system we are using two machine learning algorithms named as Support Vector Machine (SVM) and naïve Bayes algorithms.

3.4.1 Support Vector Machine (SVM)

An SVM classifies information by means of finding the exceptional hyperplane that separates all information facets of 1 type from those of the other classification. The best hyperplane for an SVM method that the one with the biggest line between the two classes. An SVM classifies data through discovering the exceptional hyperplane that separates all knowledge facets of one category from those of the other class. The help vectors are the info aspects which are closest to the keeping apart hyperplane.

3.4.2 Naive Bayes

Naive Bayes algorithm is the algorithm that learns the chance of an object with designated features belonging to a unique crew/category. In brief, it's a probabilistic classifier.

The Naive Bayes algorithm is called "naive" on account that it makes the belief that the occurrence of a distinct feature is independent of the prevalence of other aspects. For illustration, if we're looking to determine false profiles based on its time, date of publication or posts, language and geo-position. Even if these points depend upon each and every different or on the presence of the other facets, all of these properties in my view contribute to the probability that the false profile.

3.5 Evaluation

- The NLP pre-processing techniques are used to analyze the dataset and machine learning algorithm such as SVM and Naïve Bayes are used to classify the profiles.
- To classify the fake profile or genuine profiles in Facebook

IV. IMPLEMENTATION

TABLE 1
MODELS COMPARISON USING CROSS-VALIDATION

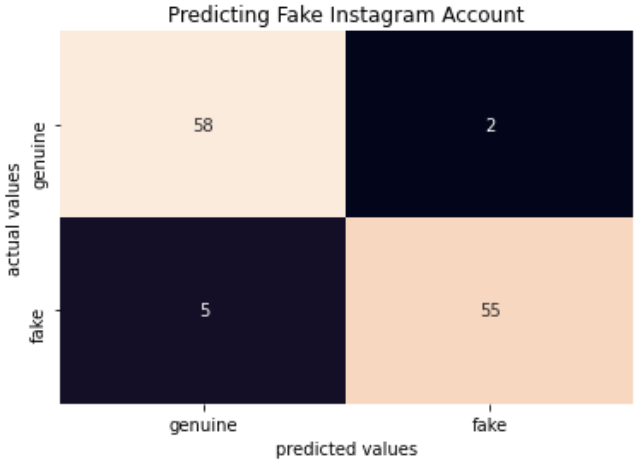
Model	Train_Score	Validation_Score
Gradient Boosting Classifier	1.000	0.980
Random Forest Classifier	1.000	0.980
Logistic Regression	0.977	0.974
SVC	0.936	0.936
Gaussian NB	0.779	0.793

We will select the two best models to continue.

These models are Random Forest Classifier and Gradient Boosting Classifier.

Test score: 0.922

Final Evaluation	Precision	Recall	F1-Score	Support
genuine	0.92	0.97	0.94	60
fake	0.96	0.92	0.94	60
Accuracy			0.94	120
Macro Avg	0.94	0.94	0.94	120
Weighted Avg	0.94	0.94	0.94	120



We can see our model predicted around 91.5% fake accounts and 90.2% genuine accounts correctly. The model only predicted 7 accounts wrong

V. CONCLUSION

In this paper, we proposed machine learning algorithms along with natural language processing techniques. By using these techniques, we can easily detect the fake profiles from the social network sites. In this paper we took the Facebook dataset to identify the fake profiles. The NLP pre-processing techniques are used to analyze the dataset and machine learning algorithm such as SVM and Naïve Bayes are used to classify the profiles. These learning algorithms are improved the detection accuracy rate in this paper.

REFERENCES

[1] Romanov, Aleksei, Alexander Semenov, Oleksiy Mazhelis, and Jari Veijalainen. "Detection of fake profiles in social media-Literature review." In *International Conference on Web Information Systems and Technologies*, vol. 2, pp. 363-369. SCITEPRESS, 2018.

[2] Adikari, Shalinda, and Kaushik Dutta. "Identifying fake profiles in linkedin." *arXiv preprint arXiv:2006.01381* (2020).

[3] Kaubiyal, Jyoti, and Ankit Kumar Jain. "A feature-based approach to detect fake profiles in Twitter." In *Proceedings of the 3rd International Conference on Big Data and Internet of Things*, pp. 135-139. 2019.

[4] Elovici, Yuval, F. I. R. E. Michael, and Gilad Katz. "Method for detecting spammers and fake profiles in social networks." U.S. Patent 9,659,185, issued May 23, 2019

[5] Elyusufi, Y. and Elyusufi, Z., 2019, October. Social networks fake profiles detection using machine learning algorithms. In *The Proceedings of the Third International Conference on Smart City Applications* (pp. 30-40). Springer, Cham.

[6] Ozbay, F.A. and Alatas, B., 2020. Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: Statistical Mechanics and its Applications*, *540*, p.123174.

[7] Gurajala, S., White, J.S., Hudson, B. and Matthews, J.N., 2015, July. Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach. In *Proceedings of the 2015 International Conference on social media & Society* (pp. 1-7).

[8] Ramalingam, D. and Chinnaiah, V., 2018. Fake profile detection techniques in large-scale online social networks: A comprehensive review. *Computers & Electrical Engineering*, *65*, pp.165-177.

[9] Ojo, Adebola K. "Improved model for detecting fake profiles in online social network: A case study of twitter." *Journal of Advances in Mathematics and Computer Science* (2019): 1-17.

[10] Meel, Priyanka, and Dinesh Kumar Vishwakarma. "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities." *Expert Systems with Applications* (2019): 112986.