FAKE PROFILE DETECTION IN USING MACHINE LEARNING

CHEKURI. SRI DIVYA NAGA DURGA[1], S.VENNELA [2],
[1] **Assistant professor, MCA DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
EMAIL ID: divyachekuri4@gmail.com
[2]**PG Student of MCA, Dantuluri** Narayana Raju College**, Bhimavaram, Andharapradesh**
EMAIL ID: vennela9900@gmail.com

Abstract: The majority of people use social networking sites as part of daily life. People create their profiles on social networking sites every day and communicate with others there regardless of their location or time. The majority of their time is spent on social networking sites by people of all ages. Social networks all over the world are used to create and share enormous amounts of data. These motivations have led to the emergence of fake users who prey on social network members. We need classify the social networks profiles of the users. We may obtain the actual user profiles from the categorization. To increase the accuracy rate of the false profile identification, we employ machine learning and natural language processing (NLP) approaches. Additionally, Support Vector Machine (SVM) also be supported

Index Terms: - Term Frequency Inverse Document Frequency, Tiff Matrix**,** Machine Learning Techniques.

I Introduction

Spam poses a serious danger to the web's utility. In order to transmit their content to users, spammers disguise it as beneficial or pertinent content. The legitimate users ingest this spam data because they believe it to be pertinent to their informational requirements. A communication channel isn't valuable until the spammers arrive, according to Clay Shirky. Spam is difficult to eradicate. Spam emails continue to circulate online despite the fact that email providers like Gmail, Microsoft, and others have been effectively detecting them for a number of years. According to these services, email spam accounts for 90 to 95 percent of all email exchanges. Companies are unable to halt spammers even after successful spam detection, which guarantees the financial gain spammers receive when they trick a user into clicking on a spam link

With the development of online social networks, the scope of the threat posed by spam has grown. Twitter is one of the most well-known online social networks that has been severely impacted by spam. Twitter spam is more dangerous because it targets Twitter's trending topics and is therefore a little simpler to access, thanks in part to the hash-tag operator. Twitter's diverse readership is another factor that makes it an easier and more lucrative target for spammers. Twitter users come from various walks of life, including teachers, students, politicians, celebrities, customers, and even the ordinary public. They belong to a variety of age categories, although the majority of us use Twitter most frequently between the ages of 55 and 64. About 60% of people access Twitter on their mobile devices. With 288 million active users each month, Twitter is a rapidly expanding social networking site. There are over 400 million tweets posted every day, with each user's account posting an average of 208 tweets. A user encounters numerous issues with search results that share recurrent and relevant information as a result of this constant distribution of information. This might occasionally be highly troubling because a user has browse through all the information to acquire a broad overview of the subject. Due to the prevalent use of URLs, acronyms, casual language, and modern linguistic concepts, spam detection on the twitter network is challenging. Here, outdated techniques for spam information detection fall short. Studies on a variety of methods for spotting spam on Twitter and blogs using various features are currently available. Knowing the significance of spam on Twitter, we draw inspiration or motivation from

In this research, we offer a spam detection approach for detecting spam tweets. This approach is based on emotive aspects of a tweet. This user needs and decision to build and develop enhanced approaches to detect spams on twitter. The goal is to capitalize on the strategy spammers employ to persuade users to click on a specific link. They undoubtedly utilize some inspirational language (such as "the finest website," "amazing service," etc.) to influence people's perceptions. The results demonstrate that this exploitation of emotive aspects is successful. Of the user and use statistical in a specific tweet (examples of different spam tweets are presented in the table I. Another strategy is given for spam identification in twitter network. They analyses the spread of spam in the network.

2. Literature survey

Prior to 2004, email spam classification research was suffering from considerable diversity and controversy in the methods

employed for spam filtering and the ways by which these methods were evaluated. It was unclear, which method was best and showed promise for improvement. Three different communities were focusing on these issues (Lyman, 2009):

- The community of developers and practitioners with the motive of developing tools for instantaneous deployment;
- The community of spam filter vendors with the motive of selling spam filters;
- The community of researchers with the motive of inventing new facts and validating existing theories and algorithms.

Several spam filtering methods were experimented and investigated by users, practitioners, vendors and researchers and classified into three groups:

- Manual inspection,
- System oriented approaches,
- Content-based filtering.

Apart from all, Content-based filters can be further classified as

- Ad-hoc Rule-based filters,
- Practical learning filters,
- Machine learning research.

3 Implementation Study

Shen *et al.* [29] investigated issues of detecting spammers on Twitter. The proposed method combines characteristics withdrawal from text content and information of social networks. The authors used matrix factorization to determine the underline feature matrix or the tweets and then came up with a social regularization with interaction coefficient to teach the factorization of the underline matrix. Subsequently, the authors combined knowledge with social regularization and factorization matrix processes, and performed experiments on the real-world Twitter dataset, i.e., UDI Twitter dataset.

Washha*et al.* [31] described the Hidden Markov Model for filtering the spam related to recent time. The method supports the accessible and obtainable information in the tweet object to recognize spam tweets and the tweets that are handled previously related to the same topic.

Jeong*et al.* [17] analyzed the follow spam on Twitter as an alternative of dispersion of provoking public messages, spammers follow authorized users, and followed by authorized users. Categorization techniques were proposed that are used for the detection of follow spammers. The focus of the social relation is cascaded and formulated into two mechanism, i.e., social status filtering and trade significance

profile filtering, where each of which uses two-hop sub networks that are centered at each other. Assemble techniques and cascading filtering are also proposed for combining the properties of both trade significance profile and social status. To check whether a user is fake or not, a two-hop social network for each user is focused to gather social information from social networks.

Meda*et al.* [21] presented a technique that utilizes a sampling of non-uniform features inside a machine learning system by the adaptation of random forest algorithm to recognize spammer insiders. The proposed framework focuses on the random forest and non-uniform feature sampling techniques. The random forest is a learning algorithm for the categorization and regression that works by assembling several decision trees at preparation time and selecting the one with the majority votes by individual trees. The scheme integrates bootstrap aggregating technique with the un-planned selection of features.

Disadvantages

- There is no filtering system based on a preprocessing schedule and on Naïve Bayes algorithm to discard the tweets containing inaccurate information,.
- Less security due No URL Based Spam Detection.

3.1 Proposed Methodology

In the proposed system, the system elaborates a classification of spammer detection techniques. The system shows the proposed taxonomy for identification of spammers on Twitter. The proposed taxonomy is categorized into four main classes, namely, (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. Each category of identification methods relies on a specific model, technique, and detection algorithm.

The first category (fake content) includes various techniques, such as regression prediction model, malware alerting system, and Lfun scheme approach. In the second category (URL based spam detection), the spammer is identified in URL through different machine learning algorithms. The third category (spam in trending topics) is identified through Naïve Bayes classifier and language model divergence. The last category (fake user identification) is based on detecting fake users through hybrid techniques.

Advantages

The average numbers of verified accounts that were either spam or non-spam and (ii) the number of followers of the user accounts.

The fake content propagation was identified through the metrics that include: (i) social reputation, (ii) global engagement, (iii) topic engagement, (iv) likability, and (v) credibility. After that, the authors utilized regression prediction model to ensure the overall impact of people who spread the

fake content at that time and also to predict the fake content growth in future.

4. Methodology

The proposed approach is divided into three stages

Spam Detection

Spam Detection Based on the above identified features, we proceed to use traditional classifiers to help detect spammers. In this work, several classic classification algorithms such as Random Forest, Naïve Bayesian, Support Vector Machines, and K Nearest neighbours are compared. The Random Forest classifier is known to be effective in giving estimates of what variables are important in the classification. This classifier also has methods for balancing error in class population unbalanced data sets. The Naïve Bayesian classifier is based on the well-known Bayes theorem. The big assumption of the naïve Bayesian classifier is that the features are conditionally independent, although research shows that it is surprisingly effective in practice without the unrealistic independence assumption. To classify a data record, the posterior probability is computed for each class is a normalized factor which is equal for all classes, only the numerator needs to be maximized in order to do the classification for the Naïve Bayesian classifier. The Support Vector Machine method we used is the SMO scheme implemented in the Python programming. This SMO scheme, designed by J.C.Platt[16],uses a sequential minimal optimization algorithm to train a support vector classifier using polynomial or RBF kernels. The SMO classifier has been shown to outperform Naive Bayesian classifier in email categorization when the number of features increases. The K-Nearest Neighbour method implemented in the Python programming is the IBK classifier.

$$P(Y \mid X) = \frac{P(Y)\prod_{i=1}^{d} P(X_i \mid Y)}{P(X)} \quad (2)$$

Data Collection

We downloaded tweets from an Online source using Tweets which download the data in a csv file and cover tit to txt. It gives us the list of all followers, followings ,tweets of the particular selected account. After basic pre-processing, we are left with around 70k tweets which are classified into following (a) Legit Users (b) Legit User Tweets (c) Spammer Tweets (d) Spammer Users. Manual annotation of these tweets was done with spam or not-spam labels using two annotators A and B. Kappa score for this annotation was found satisfactory (0.82) to proceed with the experiments. We decide to use standard metrics for measuring the usefulness of our approach and hence precision, recall, and F-measure are used.

Features Performance Comparison

Here we will discuss our proposed features spam detections performance by using five selected classifiers (SVM, Random Forest, Naive Bayes, Bays Network and J48). We have compared the performance of different features by making different We have discussing just one combination" all proposed features with baseline features combination"

5 Results and Evolution Metrices

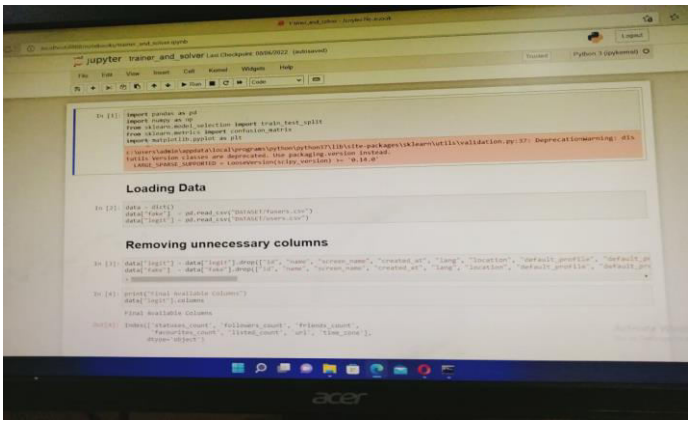

Fig1:

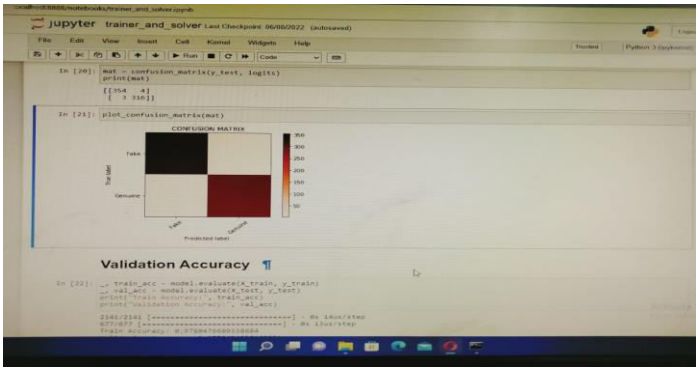

Fig2:

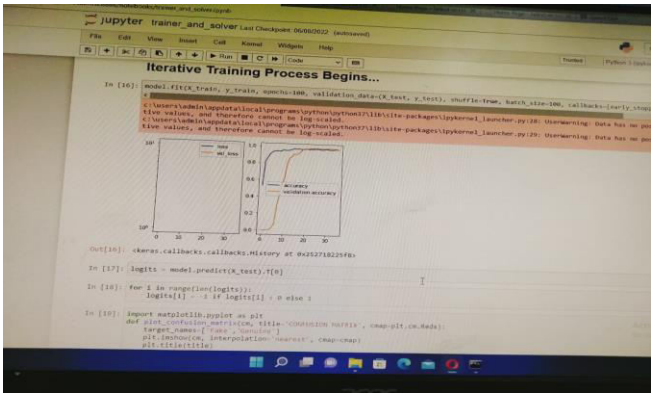

Fig3:

6 Conclusion

In this project, we propose some user-based and content-based elements that can be utilised to distinguish between legitimate users and spammers on the well-known social networking website Twitter. The Twitter spam policies and our observations of spammers' actions have an impact on these suggested enhancements. Then, we employ these attributes to detect spammers. Using the Twitter dataset we have compiled, we assess the efficacy of these variables in spammer detection using conventional classifiers like Random Forest, Nave Bayesian, Support Vector Machine, and K-NN neighbour schemes. Our findings demonstrate that the Random Forest classifier performs at its peak. Our suggested features can attain precision and F-measure using this classifier. Based on our dataset, our features yield somewhat superior classification outcomes. Next, we'll assess our identification method utilising a larger Twitter dataset

7 References

[1] How to; 5 Top methods & applications to reduce Twitter Spamhttp://blog.thoughtpick.com/2009/07/how-to-5-topmethods-applications-to- reduce-twitter-spam.html

[2] Rish. "An empirical study of the naïve bayes classifier". Proeedings of IJCAI workshop on Empirical Methods in Artificial Intelligence,2005.

[3] L. Bilge et al, "All your contacts are belong to us: automated identifty theft attackson social networks", Proceedings of ACM World Wide Web Conference,2009.

[4] T.N.Jagaticetal,"SocialPhishing",CommunicationsofACM,Vol50(10):94-100,2007.

[5] S. Yardi et al, "Detecting Spam in a Twitter Network", First Monday, Vol 15(1),2010.

G. Stringhini, C. Kruegel, G. Vigna, "Detecting Spammers on Social Networks", Proceedings of ACM ACSAS'10, Dec,2010