Spam Tweet Detection Based On Machine Learning Approach

Miss. Salke Bhagyashri A.[1] Miss. Phad Kanchan R.[2] Miss. Bhosale Supriya B.[3]

[1]*UG Student PREC LONI, SPPU, Maharashtra, India*
[2]*UG Student PREC LONI, SPPU, Maharashtra, India*
[3]*UG Student PREC LONI, SPPU, Maharashtra, India*

ABSTRACT

Online social networking is very vast growing growth today's world but attacks on it is more common, Amongst them one of the attack is twitter attack in this Spammers spread various malicious tweets which may have form like as links or hash tags on the website and online services , which are too harmful to real users. In order to prevent this attacks training tweets are added and further this issues is addressed by extracting 12 lightweight features such as account age, no of followers, no of following, no of tweets, no of re-tweets etc. For streaming tweet spam detection a feature discretization is important to spam detection performance. In system there is a big ground truth which includes total 600 public tweets based on the URL based security tool. Spam detection mainly builds the classification model which includes the binary classification and further it can be solved by the machine learning based algorithm. The behaviour of models. System reported the impact of the data related factors, such as spam to non-spam ratio, training data size, and data sampling, to the detection performance. The feature of implemented system is simple and time varying spam tweet detection. The System is shows as the spam detection is big challenge and it bridge the gap between the performance evaluation and mainly focus on the data, feature and model to identify the genuine user and report the spam user by giving the answer in binary value.

Keyword: - *Feature Extraction, Machine Learning, Feature Discretization, Training Tweets, Sampling,*

INTRODUCTION:

In this scenario, only information available in a tweet that was captured by Twitter's Streaming API can be used for classification. In order to better understand ML algorithm's power in classifying streaming spam tweets, System provided a fundamental evaluation in this work to achieve this goal [1]. In order to build this paper the summary is as follows:

- System created a big ground-truth for the research on spam tweet detection.
- System reported the impact of the data related factors, such as spam to non-spam ratio, training data size, and data sampling, to the detection performance.
- System extracted 12 lightweight features for streaming tweet spam detection
- System investigated machine learning algorithms to build up the tweet spam detection model.

1.1 Scope

- Use of this system is in online social networking for spam detection.

- Feature of spam tweets seems to be time varying.
- It is unable to detect the categorization of tweets on the basis of their types.

1.2 Objective

- To categories the Spam and Non-spam tweets.
- To work on a performance evaluation such as Precision, Recall, F-measure.
- To categorize the tag based tweets and link based tweets.

2. LITERATURE SURVEY

The severe spam problem on Twitter has already drawn researcher's attention. Some researchers have studied the characteristics of spam after that several significant works to detect Twitter spam have been proposed. As a result, System discusses prior related works by organizing them into two categories:

- **2.1 Characterizing Twitter Spam**

In order to better understand Twitter spam some analysis has been carried out. In 2010, Grier et al. analysed found that 2 million URLs were spam, which gives about 8% of all searching unique URLs [4]. Grier et al. also examined the performance of blacklists, and the results indicated that blacklists delay failed to stop the spread of spam on Twitter. In 2011, Thomas et al. analysed spam characteristics on a huge dataset of 1.8 billion tweets, of which 80 million were spam. They characterize the behaviour of spammers and found five large campaigns. The 89% spam accounts were rarely setting up social connections with users. Instead, 52% accounts made use of unsolicited mention and 17% accounts were hijacking trending topics. To establish the spammer's relationship, Scientist Yang et al. first carried out an analysis on the cybercriminal ecosystem, which was composed of criminal account community and criminal supporter's community on Twitter.

- **2.2 Detecting Twitter Spam**

In response to detect Twitter spam, there have been a few works system introduced. Most of these works are utilizing machine learning algorithm to separate spam and non-spam [1]. Some preliminary works, including made use of account and content features, such as account age, number of followers or followings, URL ratio, and the length of tweet to distinguish spammers and non-spammers. These features can be extracted efficiently but also fabricated easily. Consequently, some works proposed robust features which rely on the social graph to avoid feature fabrication.

- **2.3 Survey on paper**
  - In 2011, Song et al. [3]. System use distance and connectivity as the features which are hard to manipulate by spammers and effective to classify spammers.

  Pros- This paper mainly use relation feature instead of account feature. In this paper sender receiver relationship is used to detect the spam message.

  Cons- Here, the relation feature approach is very difficult to calculate.

  - In 2010, although there are few works such as [4], the irregular behaviour of user profile is detected and based on that the profile is developed to identify the spammer.

  Pros- This paper automatically detects spammer which based on the machine learning algorithm.

Cons- These papers mainly require the historical information to build the social graph.

- In 2010, K. Lee et al. [6]. System analyses how spammers who target social networking sites operate. To collect the data about spamming activity, system created a large set of "honey-profiles" on three large social networking sites.

Pros- Key component are:

I. The deployment of social Honey pots for harvesting deceptive spam profiles from social networking

II. Statistical analysis of these spam's profiles.

Cons- The features used is mainly Time consuming and resource consuming for the system.

- In 2012**,** Yang et al [2]. Carried out an analysis on the OSN to verify the fake accounts. It works on the extracted knowledge from the network so it detects, verify and remove the fake accounts.

Pros- This paper carried out an analysis on the OSN However; OSNs suffer from abuse in the form of the creation of fake accounts, which do not correspond to real humans.

Cons- This work is carried out manually so it is time consuming and expensive based on CAPTCHA.

3. SYSTEM OVERVIEW

System overview includes following steps which are as follows:

- Account Authentication
- Addition of Training Tweets.
- Feature Extraction
- Machine Learning Algorithm
- Spam Analysis Detection based on classification model
- Evaluations of Performance Matrices

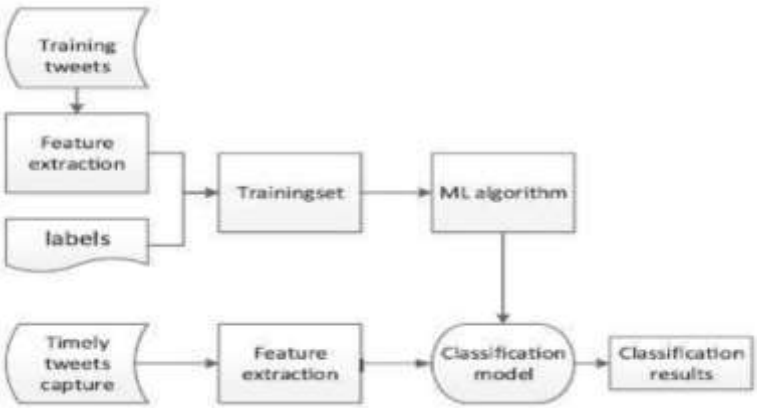

Fig -1: System Architecture

The solving approach or modules are required for above solution is as follows:

- Solving Approach -

By using Twitter's Streaming API to collect tweets with URLs. A tweet is retrieved as JSON format the returned tweet by the Streaming API contains many attributes of the tweets, such as the text, "the number of re tweets," "contained has tags, URLs,"Etc. Dataset with ground-truth is needed to perform a number of challenging machine learning-based streaming spam tweets detection tasks. Here system will describe large dataset with over 600 million tweets, including more than 6.5 million spam tweets.

- Implementation steps:

- Feature Extraction: Extraction of 10-12 features and categories as Tag based features and URL based features. User-based features were extracted from the JSON object "user," User-based features, like no_of followers, no_of followings, no_userfavourites, no_lists, and no_tweets, can be directly parsed from the JSON structure. Tweet-based features include no_retweets, no_hashtags, no_usermentions, no_urls, no_chars, and no_digits. While no_chars and no_digits need a little computing, i.e., counting them from the tweet text, others can also be straightforwardly extracted.

- Feature Statistics: System evaluate the spam detection performance on dataset by using machine learning algorithms.

- ML- Based SPAM Tweets Detection: This consist of,
  - Naïve Bays: This is mainly used for filtering the spam tweets and also used in text classification. It is mainly based on probability calculation to detect the spam message.
  - SVM: This mainly helps in data classification. The classification step is build after the training process of tweets. Timely captured tweets also label in this classifier.

Modules -

- User

  Individual access is given to the application. System has the different feature of posting any comment or status in the wall or tweet anything.

  1) Authentication of User

  By using this feature, a user is authentication is done. Various attribute information match on this. After successful attribute matching, a user is logged in the system

  2) Posting a Tweet

  In that user send tweets and it has right to the system.

  3) Spam Analysis Detection

  Detection based on which the system will generate a list containing Valid and Spam post from current user.

- Admin Module

  The admin can view List of Valid and Invalid i.e. spam post.

  1) Add training tweet

This Module will facilitate Admin to specify sample tweet content containing valid or spam tweet. This email will be further analysed to store some training parameters as valid or spam tweet.

4. CONCLUSIONS

In this System, Classifier based approach is given to solve the detection of spam messages. A classification model is mainly based on machine learning algorithm which gives the output in the form of binary value. Here the feature extraction is important phase of project to add more benefits to the system. A performance evaluation is carried out on a large dataset which includes around 600 tweets to identify the spammer also system helps to categories the spam and non-spam message.

5. ACKNOWLEDGEMENT

I would like to take this opportunity to express my thanks to my guide Proof Ghorpade P.P.for his esteemed guidance and encouragement. His guidance always helps me to succeed in this work. I am also very grateful for his guidance and comments while designing part of my research paper and learnt many things under his leadership.

6. REFERENCES

[1] Chao Chen, Jun Zhang, Yi Xie, and Yang Xiang, "A Performance evaluation of machine learning-based streaming spam tweets detection," in *IEEE transaction on computational social system,* 2015, Vol-2 No-3.

[2] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proc. Symp. Netw. Syst. Des. Implement. (NSDI),* 2012, pp. 197–210.

[3] J. Song, S. Lee, and J. Kim, "Spam filtering in Twitter using sender receiver relationship," in *Proc. 14th Int. Conf. Recent Adv. Intrusion Detection,* 2011, pp. 301–317.

[4] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammer on Twitter*," in the 7th Annu. Collab. Electron.Messaging Anti-Abuse Spam Conf., Redmond, WA, USA,* 2015.

[5] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honey pots + machineLearning," in. *Proc 33rd Int. ACM SIGIR Conf. Res.Develop. Inf. Retrieval*, 2010, pp 435-442.

[6] Nathan Aston, Jacob Liddle and Wei Hu*, "Twitter Sentiment in Data Streams with Perceptron,"in *Journal of Computer and Communications,* 2014, Vol-2 No-11.