Abstract:

The main idea of the project is how to build practical applications like Automatic Tweet Spam Detection using machine learning techniques , we have read and understood how spammers try to post tweets and based on that we extracted the features which will help to detect which the tweet is spam or not . We built three different models, including: Decision tree , Naïve Byes classifier and Neural Network and we used the Python language and its libraries like Scikit learn related to machine learning algorithms.

Proposted work :

We build the project and program it based on the following steps :

1- pre-processing :

This step the first thing we must to do ,it is very important for this project ,it is essential for make everything is ready for extract features .for example , we fill the misleading values in dataset in our project we fill the misleading values with zero value or the average value in the column because the misleading values causes the problems in the training model. Also we extract some charecters in the tweet and the location in order to build the high performance classifier .

2-extract features :

As we know ,The dataset contains useful and non-useful information.so we extract the features the which affect to the to the classifier strongly.and ignore the other non useful information we can extract several features from one one column in raw data ,in our project we founf the extract features as the following:

1-number of hasthges:According to our searches, we found that the spammers use hashtages a lot to attracts attention through followers.

2-number of mention :

3-special content in Tweet: we fount that the spammers use special url in their tweets as this (<https://t.co/> ….) this used a lot in tweets of spammers.

4-follwing and follwers : we found the relationship between following and follwre and Type of tweet. We found f the following of spammers greater than its follwers then the type of sweet is spam s, we we exrract feature compare between following and follwers .

5-content url in tweet : we found that the spammers use a url a lot in order to spread their malicious links and harmful content to people.

6-special character in the location : we found the tweet with misleading values in location will be not spam ,and the same time we found the spammers use special character a lot in location as this (“:?!+…)

Models Training:

In our project we implement three different classifiers as the following :

Desion Tree:

This used as Supervised learning and with application which have the output labels

Naïve Byes classifier : this algorithm based on the byes theorem

Neural Network : This used as Supervised learning and with application which have the output labels

In all above algorithms , we can split the dataset to 0% of data to training the mode to be able to predict the tweets in new data and 20% of data to test the model and check the performance of the model like accuracy ,presion and recall .

Conclusion:

We have learned in this project a lot of new and useful topics in the machine learning, so that we were able to build a model in different ways in order to detect the tweet whether it is a spam or not. We conclude that the python language and its libraries form a strong and important part in building the Ai Applications We built three Classifiers and we saw the difference between them through the results like Accuracy , Presion and Recall and the results were excellent.

Summarization:

The Twitter application is one of the most important social media in this modern , and it is similar to the Facebook application, which enables people to communicate on the Internet and share the opinions between them. As there are more than a billion users on Twitter, as a result of this large number of tweets, some users have appeared who publish spam (unwanted) tweets in various fields. Based on previous statistics, every person who used social media encountered spam of some kind.

There are many forms that social media users are exposed as in the following:

Viruses :spammers use the twitter as a platform to spread malicious data in the system of users.

Phishing attacks : Sensitive user information is obtained by impersonating a genuine user.

Spammers - send spam messages or post spam tweets to the users of social networks.

fake attack - attacker obtains multiple fake identities and pretends to be genuine in the system in order to harm the reputation of honest users in the network.

Clone and identity theft attacks: Hackers create a social media account with the same personal data as a genuine user to defraud their friends.

TYPES OF SPAMMERS:

Spammers are defined as malicious users who create fake accounts and information to pose a threat to the security and privacy of users and social media, so, They are categorized as:

Phishers: They pretend to be genuine users to get personal data of other genuine users

Fake Users: They are users who impersonate genuine people to deceive their friends

Promoters : They are the ones who send malicious links to advertisements or promotional links to sites that steal user data

Summers goals in general:

a) Disseminate pornography b) Spread viruses c) Phishing attacks d) Compromise system reputation

As we know Twitter uses a chirping bird as his emblem thereforeTwitter name. Users can access it for frequent exchange Information called "tweets" which are messages of up to 140Long characters that anyone can send or read.

Users share these tweets which may contain news,Reviews, photos, videos, links and messages. Below Standard terms used on Twitter and relevant to our work:

Tweets : A message on Twitter containing maximum length of 140 characters.

Followers & Followings [3]: Followers are the users who are following a particular user and followings are the users whom user follows.

Retweet: A tweet that has been reshared with all followers of a user.

Hashtages: twitter uses this # symbol so that users can target specific words or topics that are known and easily accessible to other users.

Mentions: Twitter will use the @ symbol in order to allow users to tag their friends.

List : Twitter provides a mechanism to list users you follow into groups

Direct Message [3]: private masseges between users.

URL : Spammers post a lot of these bad url in their tweets

Spam Words- Spammer’s tweets mainly consist of spam words.

HTTP links- if tweets contain maximum number of www or http://, then they are posted by spammers.

Duplicate tweets- spammers tend to post duplicate tweets with different @usernames in tweets.

There are some Existing Methods for detections of spam profiles in twitter like :

1-Benevenuto et. al. [7] detected spammers on the basis of tweet content and user based features. Tweet content attributes usedare - number of hashtags per number of words in each tweet, number of URLs per word, number of words of each tweet, number of characters of each tweet, number of URLs in each tweet, number of hashtags in each tweet, number of numeric characters that appear in the text, number of users mentioned in each tweet, number of times the tweet has been retweeted. Fraction of tweets containing URLs, fraction of tweets that contains spam words, and average number of words that are hashtags on the tweets are the characteristics that differentiate spammers from non spammers. Dataset of 54 million users on Twitter has been crawled with 1065 users manually labelled as spammers and non-spammers. A supervised machine learning scheme i.e. SVM classifier has been used to distinguish between spammers and non spammers. Detection accuracy of the system is 87.6% with only 3.6% non-spammers misclassified.

2-Twitter facilitates its users to report spam users to them by sending a message to “@spam”. So Gee et. al. [12] utilized this feature and detected spam profiles using classification technique. Normal user profiles have been collected using Twitter API and spam profiles have been collected from “@spam” in Twitter. Collected data was represented in JSON then it was presented in matrix form using CSV format. Matrix has users as rows and features as columns. Then CSV files were trained using Naive Bayes algorithm with 27% error rate then SVM algorithm has been used with error rate of 10%. Spam profiles detection accuracy is 89.3%. Limitation of this approach is that not very technical features have been used for detection and precision is also less i.e. 89.3% so it has been suggested that aggressive deployment of any system should be done only if precision is more than 99%

During this research, it was found that different ways of detecting spam systems and each time some improvements are made to get a higher detection rate.

1. Since Twitter has millions of active users and this number is constantly increasing. And almost all the authors have used very small testing dataset to see the performance of their approach. So there is a need to increase the testing dataset to see the performance of any

approach.

2. Need to develop a multivariate model.

3. Need to develop a method that can detect all kinds of spammers.

4. Need to test the approaches on different combinations of spammers and non-spammers.

Through, the most widely used classifier like : SVM, Naïve Byes And we extracted auxiliary features to detect based on features for the user or for the content These systems were validated on a small data set have not been even tested with different combinations of spammers and non-spammers. Combination of features for detection of spammers has shown better performance in terms of accuracy, precision, recall etc. as compared to using only user based or content based features.

.