Comparative Study of Spam Detection in Twitter by Different Approaches of Sentimental Analysis and Machine Learning Algorithm

Nikhil. K. S1, Anand. R2, Dr. Anil. G. N3, Jagadish. P4
M. Tech Student1, Assistant Professor2,3, Professor4
B.M.S Institute of Technology and Management, Yelahanka, Bengaluru, Karnataka, India

Abstract:
In this era, micro blogging websites such as Twitter and Facebook are filled with opinions and data. With rapidly growing online social network, spammers find these platforms easily accessible to trap users in mischievous and awful activities by posting spam messages. So, Spam detection is very necessary. The proposed framework will detect spam tweets by two methods: In first method, Tweets are sent to six machine learning classifiers to predict the spam tweets based on different parameters. In Second method, Tweets are sent to VEDAR sentimental analysis algorithm followed by machine learning classifiers. Results are compared. By comparing the results, it is easy to choose the best method to be followed to detect the spam tweets and also the best classifier can be identified by evaluating the results in terms of Accuracy and F1 Score. For validating, each tweet is sent to textbox of UI and can see the output which shows a message that the respective input tweet is spam of ham (not spam).

I. INTRODUCTION:
Twitter, a social networking and micro blogging site, is undeniably popular online social media platforms available today, where people send and read in short messages called “Tweets”, with 110 billion daily users will be active and 800 billion tweets are sent daily. For anyone who follows you on Twitter short information is shared is called Tweeting, with the intention that your posts are interesting to people. Usually everyone can read tweets, but messages can be kept private by sender. Private tweets can be read only by their followed friends. With rapidly growing online social network, spammers find these platforms easily accessible to trap users in mischievous and awful activities by posting spam messages. Many people across the world use twitter to post their daily routine, breaking news to outside world.

Parallely, spammers try to hack their account via sending spam messages which may contain URLs, photos on which clicking redirects to malicious sites and by hacking spammers post vulgar messages on individuals’ story or on their twitter wall. Due to this, Individuals reputation will go down and there comes the situation where individuals have to face many controversies to answer all questions by others. So, Spam detection is very necessary.

II. LITERATURE SURVEY:
Google Safe Browsing and Twitter’s BotMaker tools detecting and blocking spam tweets but fails to give protection for the user in real time in earlier stage. So, to stop spammers, they proposed a framework which considers tweet text feature along with user tweets to classify the tweets which helps to detect the spammer even if he creates a new account. They have used three ML algorithms - SVM, NN and RF. They have achieved an accuracy of 91.65%.

Tweets can be classified as positive, negative or attention is most valuable for certain companies, organizations. The concentrate’s principle center is to group negative, positive and nonpartisan methodologies of three clarified twitter datasets. Impact of oversampling, unigram highlights and different highlights on what’s more, class-put together precision proportions are worked with respect to the datasets. Benchmark is come to in dataset-2 tests. 88% generally speaking exactness was seen in dataset-1 tests which beats the earlier art. Unigram highlights indicated noteworthy impact on generally exactness, class-based precision balance.

With a specific end goal to handle Twitter spam drift issue, a plan called Lfun plot is utilized which can discover changed spam tweets from unlabeled tweets and wire them into classifiers preparing process. The new preparing dataset is utilized to prepare another dataset containing unlabeled tweets which will bring about finding of spam tweets. Our proposed plan will modify preparing information, for example, dropping too old examples after certain time which will dispose of unwanted data sparing space.

Twitter is one of the greatest micro blogging organizing stages; every day it post billions of tweets by many user client. Such an adaptability and wide spread of utilization, Twitter effectively gets interrupted along noxious exercises. Noxious exercises incorporate malware interruption, spam dispersion, social assaults, and so forth. Spammers utilize social building assault technique to send tweets which are spam, URLs which
The effect invigorates analysts to build up a model that break down, recognizes, recuperates from disparaging activities in twitter. Twitter organize is immersed with a huge number of phony spam profiles which may risk the typical client's safety and protection. To improve genuine client’s wellbeing and ID of spam profiles become important pieces of their exploration.

**M Rathi, D Varshney, A Malik,, Rachita S, S Mendiratta**

“Sentiment Analysis of Tweets using Machine Learning Approach [8]” International Conference on Contemporary Computing (IC3), 2018

The principle accentuation of this exploration is on the grouping of feelings of tweets' information accumulated from Twitter. Before, analysts and researchers were making using of machine learning techniques for sentiment classification, but the results were not up to expectation. So as to improve grouping results, they came up with using ensemble machine learning techniques to increase the reliability and efficiency of proposed approach. For the equivalent, we are combining Support Vector Machine with Decision Tree and trial results demonstrate that our proposed methodology is giving better order results regarding f-measure furthermore, accuracy rather than individual classifiers.

**III. SYSTEM DESIGN**

Fig 1 explains about the architecture of the system, the tweets from online are fetched to train the machine learning models. Obtained tweets data are sent for data pre processing where the data are lemmatized stemmed and words with less than three characters are removed which is followed by feature engineering process. In this, important features, words in the document are stored as a bag of words with their scores. The words with their scores are used to train the machine learning models such as SVM, RF, KNN, MNB, MLP and DT. These models with be evaluated by using bags of words and their scores, this process is continuous as new bag of words keeps on changing with the parameters such as Accuracy, Confusion Matrix and F1-Score. New input tweet which is to be checked whether it is spam or not is pre-processed and sent to the machine learning models when are ready. A model using trained data and bag of words of new tweets predicts the output as spam or ham (not spam).

**IV. IMPLEMENTATION**

Fig 2 Flow diagram of the proposed system
Fig 2 explains, Online tweets data of Twitter are extracted. The data will be in the form of raw data, so the data is sent for pre processing. In first method, raw data are directly sent to machine learning classifiers to calculate accuracy and F1 score. In second method, using VEDAR algorithm and LDA, raw data are pre-processed by removing punctuations, words less than 3 characters, stemmed and lemmatized which results in Bag of Words where only important words from raw tweets data is collected. These Bag of Words are then sent to Tf-idf vectorization, such that for each word the scores are been estimated. Higher the score of the word, higher is its importance in the document; accuracy and F1 score is calculated. Two methods results are tabulated and can is noticed that second method i.e. VADER followed by machine learning classifiers yields better accuracy and F1 score to detect the spam tweets. For validation, the new tweets can be sent to machine learning classifiers to check whether the tweet is spam or not.

V. SEQUENCE DIAGRAM

VI. RESULTS:

Below table shows the comparative results of VADER sentimental analysis followed by machine learning classifiers with accuracy and F1 score as its parameters. Last column shows the positive percentage which is increase in the accuracy of detecting spam.

<table>
<thead>
<tr>
<th>ML Methods</th>
<th>ML classifiers NOT followed by VADER</th>
<th>ML classifiers followed by VADER</th>
<th>% Variation Found between two methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>95.4341</td>
<td>93.8148</td>
<td>98.3857</td>
</tr>
<tr>
<td>NB</td>
<td>94.4802</td>
<td>92.8476</td>
<td>97.4018</td>
</tr>
<tr>
<td>DT</td>
<td>87.6915</td>
<td>73.3729</td>
<td>90.4036</td>
</tr>
<tr>
<td>MLP</td>
<td>86.7874</td>
<td>72.6165</td>
<td>89.4900</td>
</tr>
<tr>
<td>RF</td>
<td>92.2152</td>
<td>87.2421</td>
<td>94.7085</td>
</tr>
<tr>
<td>KNN</td>
<td>89.4861</td>
<td>71.1198</td>
<td>91.3901</td>
</tr>
</tbody>
</table>

VII. CONCLUSION

The detection of the spam tweets is very important so that the user will be aware and won’t be in the situation to face controversies. The proposed methodology uses two methods where; in first method tweets are sent to sentimental analysis algorithm followed by six machine learning classifiers, in second method tweets are directly sent to six machine learning classifiers for detecting whether the tweets are spam or not. Six Machine learning models using are built using six classifiers and trained tweets data which are fetched through online twitter. Each model has its own accuracy, confusion matrix and F1 Score as its parameters. When a new set of tweets or single tweet is send to machine learning models, it results showing whether the tweets are spam or ham (not spam). By comparing the results, it is observed that using sentimental analysis algorithm (i.e. VADER) followed by machine learning classifiers leads to better result in detecting spam rather than not using sentimental analysis algorithm. When sentimental analysis algorithm is followed by Machine learning classifiers we found that SVM has increased its accuracy by 2.95% and F1 score by 2.90%; NB has increased its accuracy by 2.91% and F1 score by 2.90%; DT has increased its accuracy by 2.71% F1 score by 2.26%; MLP has increased its accuracy by 2.71% and F1 score by 2.28%; RF has increased its accuracy by 2.49% and F1 score by 2.02%; and KNN has increased its accuracy by 1.90% and F1 score by 6.63%

FUTURE WORK

The future work is implementing the proposed method in real time so that spam tweets are detected earlier before it reaches to twitter account user so that user will be free from all malicious links, URLs.

VIII. REFERENCES:


[5]. M Rathi, D Varshney, A Malik, Rachita S, S Mendiratta “Sentiment Analysis of Tweets using Machine Learning Approach [8]” International Conference on Contemporary Computing (IC3), 2018

[6]. https://en.wikipedia.org/wiki/Twitter

[7]. https://blockgeeks.com/guides/what-is-spams-and-types/