International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

Detecting Offensive Language on Social Media

Mr. Sameer Sawant[1]
Zeal College of Engineering and Research, Pune
Student of Computer Engineering

Mr. Sanket More [2]
Zeal College of Engineering and Research, Pune
Student of Computer Engineering

Ms. Nilam Pawar[3]
Zeal College of Engineering and Research, Pune
Student of Computer Engineering

Ms. Shraddha Pawar [4]
Zeal College of Engineering and Research, Pune
Student of Computer Engineering

Prof. Rupali T. Waghmode[5]
Zeal College of Engineering and Research, Pune
Assistant Professor Computer Engineering Department

Abstract - *The high popularity of social media has highly enriched people's lives, allowing online users to share their feelings through posting comments. However, the rapid climb of Facebook has triggered a dramatic increase in spam volume and class. Spammers post their status or comment in Page to send spam content to their friends or other users within the network. To detect spam comments in Facebook, we employ Lexical analysis to build the dataset dictionary which compares itself with the comments posted on Facebook and according to that in case we find any vulgar word in that sentence we are going to delete such kind of comments automatically as soon as they are posted. And the user on whose post the comment is posted will receive a mail with that comment and get to know the name of person also. So, the user can take necessary actions. Hence all the spam comments will be removed and thus all the post will be clean and positive.*

***Key Words*:** lexical analysis, vulgar, discontinuous, inadequate, dataset, social media

1. INTRODUCTION

The social networks of the modern web allow users to communicate with each other by expressing their opinions in the form of comments related to some given posts. . It is easier to read online reviews of products or to read online advice about trips, then to ask friends or relatives on their opinions. However, such online reviews can be spam reviews that do not provide useful content. The purpose of spam comments is to promote content or services that have little rating or to promote products, which are less familiar on the market. An example of a social media site, which allows users to communicate and share opinions by means of comments, is Facebook.

The site provides a method to eliminate a spam comment if it is a vulgar. This mechanism offers protection to the Facebook user community so that offensive or off-topic comments are no longer displayed to them. Clearly, due to the continuing and increasing stream of comments, it is impossible for the community to read and rate all the comments.

The methods we present in this paper can help to automatically filter spam comments. In our research we aim to identify different types of spam comments based on various features such as: discontinuous text, inadequate and vulgar content and comment topic is not related to the specific context. Our approach is to first eliminate comments which contain discontinuous text, inadequate text and vulgar language. Our project aim is to find vulgar comments and do not display them on website.

In Existing System if one person post a status or photo, Friends passes comments even other people who are not your friends can also pass comments on that post. Sometimes unknown people can post vulgar comments and the people who are active on social media can see those comments immediately.

So we are developing a system that will automatically delete such comments as soon as they are posted, the person on whose post the comment is done will receive a mail with that comment.

2. LITERATURE SURVEY

Due to the popularity of social media, such as Twitter, Facebook many research works have been conducted on spam comments detection. The existing research mainly focused on two aspects: detecting spammers and analyzing content features of spam comments.

To detect spam comments in Chinese social media, some researchers used semantic analysis to build the self-extensible spam dictionary which updates and extends itself with new cyber words automatically and text-based features to construct classifier for detecting spam comments [1]. However, they are not able to detect spammer and delete comments.

Some researchers proposed to detect spam comments by analyzing the comments content. They thought that low similarity between microblogs and comments means that comment might be spam. However, from our observation, many normal comments are short and inconsistent with microblogs' theme. Liu et al. proposed spam dictionary and Proportion-Weight Filter model to detect two kinds of spam comments (advertisement and vulgar comments), and achieved an average accuracy value of 87.6%[2]. Their

results have much scope for improvement because they ignored other text-based features.

Some researchers detected spammers by analyzing the users' attributes and representative behaviors, like registration date, repeated reposting and aggressive following. Lin et al. assumed that spammers' attributes and behavior are different from those of normal users [3]. However, social spammers may continue to change their behaviors and try to behave like normal users.

In some research, they focus on the problem of detecting spam accounts on Facebook. Features for detection of spammers could be content-based and user social behavior-based methods. They tried to collect a large range of user comments from selling Facebook pages in Vietnam, combining with user social behavior to build a dataset of users. They also have applied Maximum Entropy (Maxent) method on the dataset to create a model which detects spam accounts [4]. However, it's not difficult for spammers to get lots of ossified fans. Thus, these methods based on social networking may not work well.

These previous studies are focused on analyzing and detecting spam contents from micro-blogs and posts in English or identifying the behavioral characteristics of the spammers. Our approach differs from the existing approaches in two aspects. First, we introduce Lexical analysis to detect spam comments. Lexical analysis provides us with additional useful features to detect spam comments. Second, we have developed a system through which spam comments won't be shown and also get deleted by system automatically and people can feel free and save to post.

3. MOTIVATION

The use of social media is increasing in today's age. Adolescents use social media more than adults. But with the increasing popularity, it also comes with the negative part. People use abusive or offensive language on other's pictures or videos. The impact of the negative influences of social media on teenage users increases with a rise in the use of offensive language in social conversations. This increase could lead to frustration, depression, and a large change in their behavior.

So it affects the good decorum of social media. Therefore, we have developed a system through which such types of comments won't be shown and also get deleted by the system automatically. People can feel free to post.

4. SYSTEM ARCHITECTURE

In above system architecture User registration process and User authentication process are shown. In User registration process user have to sign up and fill the information then first set the text password and then set the graphical password. To set graphical password user select pass point in multiple images. In User authentication process user login with user id and enter the text password if text password is correct then select pass points of images for graphical password. Verification of user id and password from database if image selection is wrong then authentication failed if image selection is correct then authentication successful.

After authentication, if the user commented on posts then the lexical analyzer will check that comment with the given dataset. If the comment is vulgar then that comment will be deleted and the response is sent to the post owner. So owners can unfriend the user.

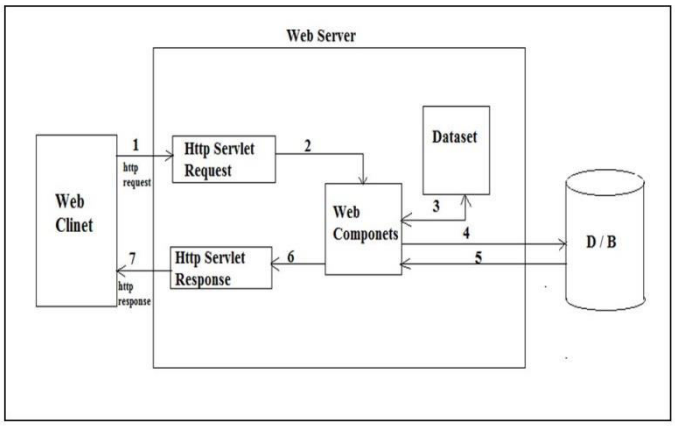

Fig -1: System Architecture

5. IMPLEMENTATION AND RESULTS

In this system, we have used a lexical analysis algorithm to detect vulgar comments. Lexical analyzer scans the entire source code of the program. It identifies each token one by one.

The system checks each comment with the dataset. The dataset contains vulgar words. If the comment matches with datasets words then that comment automatically gets deleted and mail notification is set to the post owner as shown in the below figures. Otherwise, it will be shown on the post.

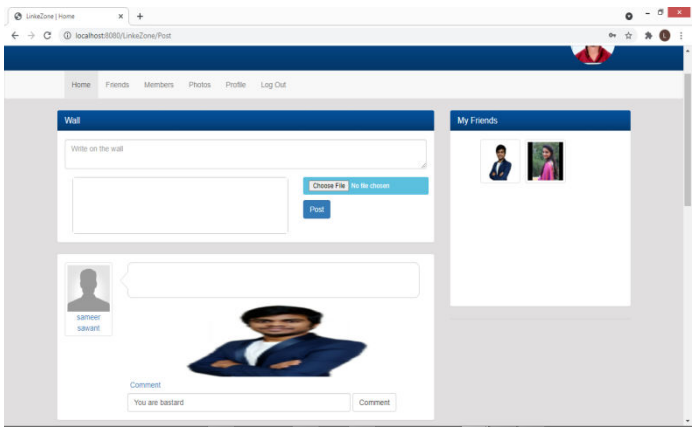

Fig -2: While posting vulgar Comment

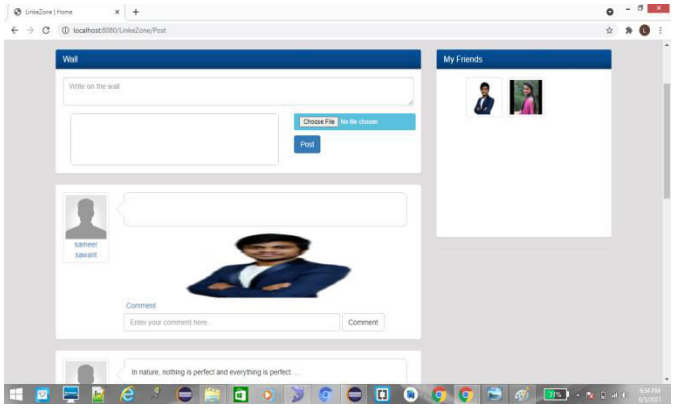

Fig -3: Hide vulgar comment from post

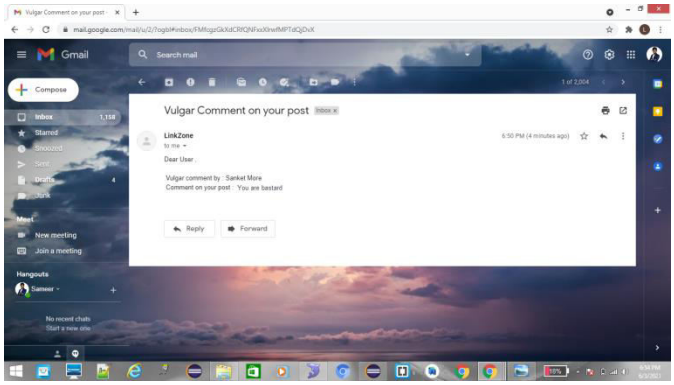

Fig -4: Mail notification sent to post owner

6. CONCLUSIONS

In this paper, we conducted an in-depth analysis on the comments posted on social sites. To detect vulgar type comments more accurately and effectively, we are going to construct a Dataset Dictionary and we are comparing that dataset with the comments posted on social media and according to that in case we find any vulgar word in that sentence we are going to delete such kind of comments automatically as soon as they are posted. Giving warning and blocking unwanted users is main aim of our system. Hence, we are removing all negative comments and thus all the post will be clean and positive.

7. FUTURE SCOPE

In future, we can get the OTP on in registration phase by SMS (Short Message Service).

Comment monitoring and removing is beneficial for E-commerce sites where some fake comments or bad comments are provided by customers on others products to decrease productivity or sailing of their products.

On Social media like Facebook, twitter while peoples are commenting on others post it is possible to provide dictionary words to monitor comments

It is also possible to intimate the user about his vulgar comment while commenting.

REFERENCES

[1] Qiang Zhang, Shangru Zhong, Kai Lei "*Spam comments detection with Self-extensible dictionary and text- based features*", Shenzhen Key Lab for Cloud Computing Technology & Applications School of Electronics and Computer Engineering(SECE) Peking University, SHENZHEN 518055 P.R.CHINA , 2017.

[2] Chenwei Liu, Jiawei Wang, Kai Lei "*Detecting spam comments in micro blogs using self-extensible spam dictionary*", Institute of Big Data Technologies Shenzhen Key Lab for Cloud Computing Technology & Applications School of Electronics and Computer Engineering(SECE) Peking University, SHENZHEN 518055 P.R.CHINA, 2016.

[3] Jong Myoung Kim, Zae My oung Kim, Kwangjo Kim "*An approach to spam comments detection through domain independent features*",2016 International Conference on Big Data and Smart Computing (BigComp).

[4] Thi-hong Voung, Vanhien tran, Minh-Due Nguyen" *Social-spam Profile Detection based on Content Classification and User Behavior*", Eighth International Conference on Knowledge and Systems Engineering (KSE), 2016.

[5] Alessandro Maisto, Serena Pelosi, Simonetta Vietri, Pierluigi Vitale "Mining Offensive Language on Social Media", University of Salerno Department of Political, Social and Communication Science Via Giovanni Paolo II, 132.

[6] Cicero Nogueira dos Santos, Igor Melnyk, Inkit Padhi "Fighting Offensive Language on Social Media with Unsupervised Text Style Transfer", IBM Research T.J. Watson Research Center.

[7] M. K. Girish Khurana, "Review: Efficient Spam Detection on Social Network," ISSN, vol. 3, no. 6, pp.2321-9653, 2015.

[8] L. Liu, Y. Lu, Y. Luo, R. Zhang, L. Itti, and J. Lu,"Detecting "Smart" Spammers On Social Network: A Topic Model Approach," CoRR, vol. abs/1604.08504, 2016.

[9] X. Zhang, and X. Zheng, "A novel method for spammer detection in social networks," in ICSDM 2015, pp. 115-118.

[10] S. Rayana, and L. Akoglu, "Collective Opinion Spam Detection: Bridging Review Networks and Metadata," in KDD 2015, pp. 985-994.