Monitoring Suspicious Discussions On Online Forums Using Data Mining

P. Venkateswara Rao
Professor ,Department of Computer Science Engineering, Narayana Engineering College Gudur

Saila Konda
B.Tech, Department of Computer Science Engineering, Narayana Engineering College Gudur

Reshma Allam
B.Tech, Department of Computer Science Engineering, Narayana Engineering College Gudur

Bindhu Lasya Nandimandalam
B.Tech, Department of Computer Science Engineering, Narayana Engineering College Gudur,

ABSTRACT

The use of online forums has been increased rapidly due to increase of internet technology. As internet technology has advantages it also has disadvantages. The discussion of news are done at online forums rather than actual mass media because of this many illegal activities takes place such as dissemination of copyrighted movies, online gambling, warning messages etc. Here we mainly focus on stress. The security force wants to monitor these online forums to download suspected postings as evidence for investigation. Here we introducing a system as a solution for this problem. To detect illegal activities and illegal postings we are using a data mining algorithm. This system helps to download postings from discussion forums, monitors and analyze online text sources such as comments, news in the internet, blogs and classify the text into groups whether the post is legal or illegal for security purpose. It will use text data mining technique.

Keywords: Online forums, Illegal activities, Illegal postings.

1. INTRODUCTION

People today are very fond of using internet and social media became a platform for chatting and expressing their views and thoughts. Based on the system some words can be legal or illegal. If an illegal word is occurred once the next time the word will be blocked and cannot be used in postings. To tackle with this problem our system helps to monitor illegal words and illegal postings. As data is constantly increasing in online forums[8] it is difficult to manage data so data mining is optimal choice to gather data and to classify words for identifying illegal posts. It helps to alert the law enforcements to monitor online forums. Stop words, Stemming algorithms are used in monitoring illegal posts.

2. LITERATURE SURVEY

This paper gives information about the provoking posts and how to deal with it.. This divides the words into some categories such as emotions, hacking, privacy and fraud etc.[1]

In this work the author discusses about the feelings and emotions on discussion forums. Here EmoTxt is used to find the feelings and categorize and outputs in comma separated value (CSV) format. This follows a structured hierarchal format.[2]

This paper talks about discussion forums and the way it spreads the information and malicious posts. This paper uses a data mining algorithm Naive-Bayes theorem. This theorem analyze the words into positive and negative.[3]

The paper deals about suspicious activities that happen on online forums that caused b exchanging of information such as text, images, video etc. As many of the postings is in text format it focuses on text data. This uses Stemming algorithm, Brute force algorithms, Suffix Stripping algorithms, Index compression etc.[4]

In this paper the author deals with the illegal words and that word is replaced with *** if it continues than the account will be blocked for 1 day. If the gets blocked more than 3 times than the account will be out of the forum completely.[5]

In this paper [6] the author mainly focuses on the stock market. This finds the postings from stock forums and discovers change in investors. The approach here used is monitoring opinions whether it is positive, negative or neutral.

The author talks [7] about rapid increase of internet technology and growth of malicious posts. Here to tackle with this problem he uses stemming, stop word selection, Levenshtein algorithm.

This paper talks [9] about gathering postings from online forums continuously and applies data mining techniques hot topics and clusters them into groups. Here different techniques are employed to collect data.

3. METHODOLOGY

Stop words is a technique helps to focus on the key words rather than commonly used words as it eliminates those words [2]. For example In English words such as ‘how’, ‘she’, ‘them’ are commonly used and stop word technique eliminates such words. This helps to find the illegal posts easily.

Stemming is a process of reducing a word to its base form. The Stemming algorithm removes the suffixes from English words and converts into its root word, for example: The words “Politician”, “Politicians”, “Policy” have suffixes and will be removed during the information retrieval and will leave word “Politics” as the root word.

In Levenshtein algorithm [1], a large set of words are compared with each other and tells how different the two strings are. For example,

Kitten → Sitten (substitution of "s" for "k")

Sitten → Sittin (substitution of "i" for "e")

Sittin → Sitting (insertion of "g" at the end).

Levenshtein distance is a measure of similarity between two words. Here in our example the difference between “kitten” and “sitting” is 3.

The Levenshtein distance between two strings is shown below:

$$\mathrm{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j) + 1 \\ \mathrm{lev}_{a,b}(i,j-1) + 1 \\ \mathrm{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

a = string 1

b = string 2

i = The character position in string 1

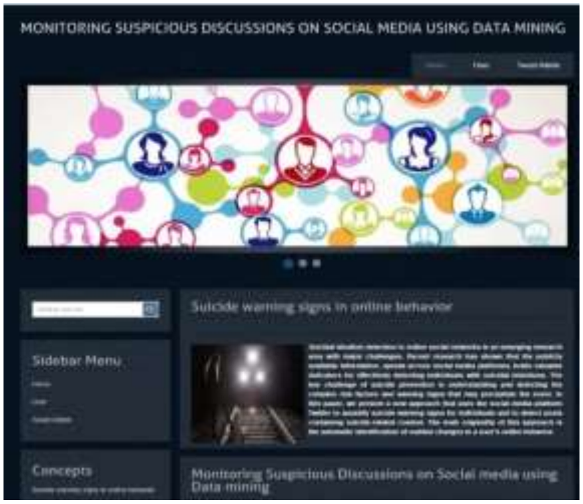
j = The character position in string 2.

Naïve-Bayes theorem is a classifier algorithm used here which is used to analyze and classify the words into legal and illegal. It is a data mining technique.

Research is being carried out using web mining [3]. The data is collected in large number of web pages using Web mining as it requires a user query interface for predicting crime from various web pages. The techniques used in web mining are classification, association analysis, outlier analysis and cluster analysis. Clustering and classification techniques identify and group the similar items in classes. The association rules mining and sequential pattern mining techniques are similar and they both identify frequently occurring sets and extract a pattern. It is more complex using all these techniques in web mining.

4. RESULTS & DISCUSSIONS

Home page of the discussed online forum and some of the related posts are shown in Fig 1..



MONITORING SUSPICIOUS DISCUSSIONS ON SOCIAL MEDIA USING DATA MINING

Sidebar Menu

Concepts

Suicide warning signs in online behavior

[illegible]

Monitoring Suspicious Discussions on Social media using Data mining

Fig. 1 Home Page

Admin login page is shown in Fig. 2 here the admin needs to give correct user id and password to login.

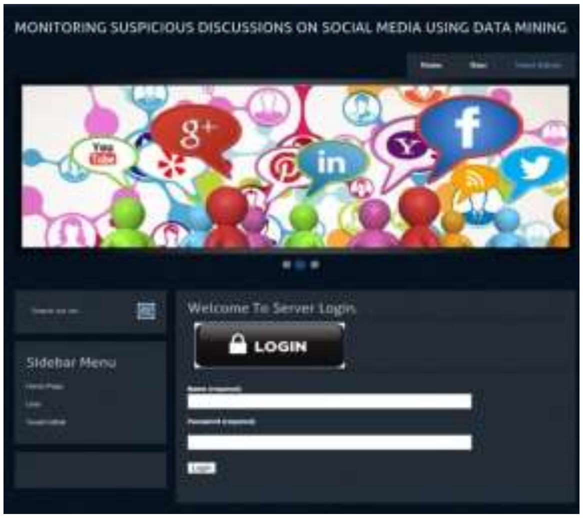


Fig.2 Admin login page

End user registration page, Here the user needs to give necessary information to get registered as shown in Fig. 3

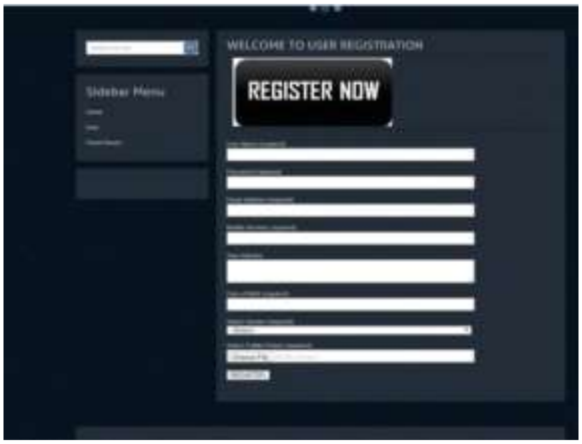
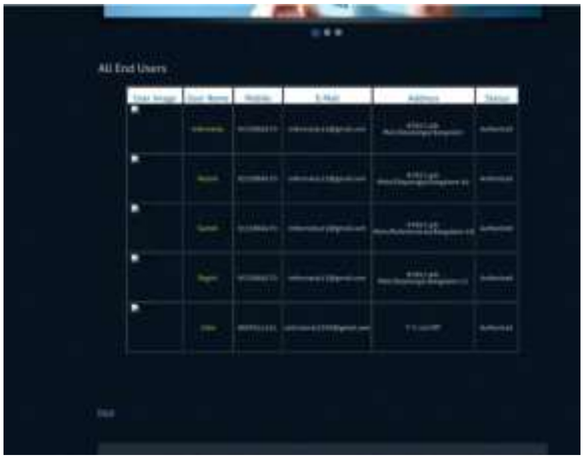


Fig 3

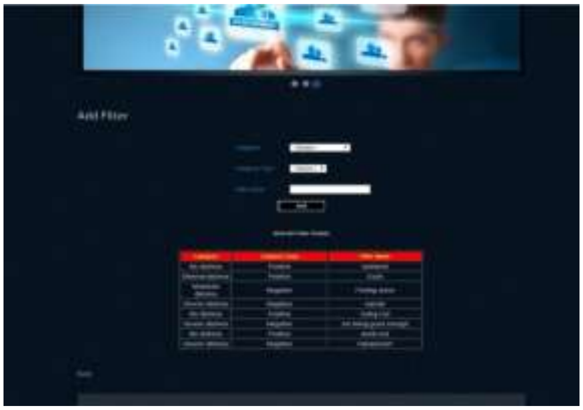
All end users and Friend request and responses are shown in Fig. 4, In this the admin can view all the end users, friend requests and their details.



All End Users

Fig 4. Friends request

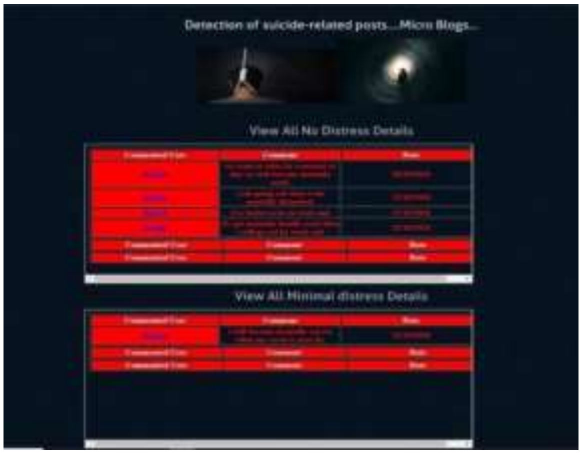
Add categories and category type using add filter page as shown in Fig 5. Here we mainly used two category types such as positive and negative.



Add Filter

Fig 5. Add

Some of the posts that related to suspicious activities are shown in Fig 6.



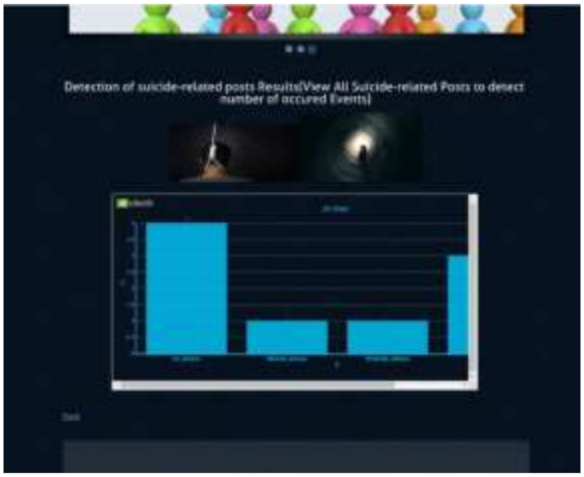
Detection of suicide-related posts....Micro Blogs...

View All No Distress Details

View All Minimal distress Details

Fig 6. Suspicious post

Suspicious posts that occurrences are represented in the form of charts as shown in Fig. 7



Detection of suicide-related posts Results(View All Suicide-related Posts to detect number of occured Events)

5. CONCLUSION

Monitoring Suspicious Discussions On Online Forums Using Data Mining paper talks about the suspicious activities on online forums and how to reduce it. Here it monitors the text data and classify them into different categories such as legal and illegal which is posted on the online forums. This helps to reduce the illegal activities that is happening on online forums and helps the law enforcements to easily tackle this problem.

6. REFERENCES

[1] Tanya Srivastava, R.Mangalagowri, Shailesh S.Dudala, 'MONITORING OF SUSPICIOUS DISCUSSIONS ON ONLINE FORUMS USING DATA MINING' International Journal of Pure and Applied Mathematics Volume 118 No. 22 2018, 257-262.

[2]Javed Hosseinkhani, Mohammad Koochakazei, Solmaaz Keikhaee and Yahaya Hamedi Aman 'DETECTING SUSPICION INFORMATION ON WEB CRIME USING CRIME DATA MINING TECHNIQUES' International journal of advanced computer science and information technology(IJACSIT) vol.-3,No.1,2014,page 32-41.

[3] Shet Nitish Nagesh, Yashaswini, Rahul Anil Prabhu, Rajatha J Shetty 'MONITORING SUSPICIOUS DISCUSSIONS ON ONLINE FORUMS USING DATA MINING' International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019.

[4] M.Suruthi Murugesan, R. Pavitha Devi, S. Deepthi, V.Sri Lavanya & Dr. Annie Princy Ph.D, 'AUTOMATED MONITORING SUSPICIOUS DISCUSSIONS ON ONLINE FORUMS USING DATA MINING STATISTICAL CORPUS BASED APPROACH' Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-5, 2016.

[5]Bavane A.B, Ambilwade Priyanka V, Bachhav Mourvika D, Dafal Sumit N, Fulari Priyanka Y, 'MONITORING SUSPICIOUS DISCUSSION ON ONLINE FORUM BY DATA MINING' International Journal Of Advanced Engineering &Science Research(IJAES), Volume 5, Issue 1, March 2017.

[6] Priyanka B.Hulde, Prof. Priyanka Dhudhe, ‘MONITORING MALICIOUS DISCUSSIONS ON ONLINE FORUMS USING DATA MINING’ International Conference on Emanations in Modern Engineering Science & Management (ICEMESM-2018).

[7] Harika Upganlawar, Nilesh Sambhe, ‘SURVEILLANCE OF SUSPICIOUS DISCUSSIONS ON ONLINE FORUMS USING TEXT DATA MINING’ International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835 Volume-4, Issue-4, Aprl.-2017.

[8] Review on Techniques and Applications Involved in Web Usage Mining B Bhavani,V.sucharita,K.V.V.Satyanarayana International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 15994-15998

[9]MISS.Surbhi Chavhan, PROF. Vaishali Agrey, ‘REVIEW STUDY ON AUTOMATIC ONLINE MONITORING AND DATA MINING INTERNET FORUMS’ International Journal of Research in Computer & Information Technology (IJRCIT), Vol. 4, Issue 2, March-2019.