MONITORING SUSPICIOUS AND ILLEGAL DISCUSSION ON ONLINE FORUM

Tharani.V, Dhivya preetha.S, Snegha.J *(Students[1,2,3])* .
R. Charanya *Assistant Professor (Senior)* .
Vellore Institute of Technology, Vellore, Tamilnadu, India

ABSTRACT

Nowadays, people are passionate about using internet in their day to day life. This rapidly increases the usage of online forums. Online forums are nothing but a medium to share their thoughts, feelings and emotions towards specific pictures, videos and paintings etc.., It leads to happening of many legal and illegal activities. Those illegal activities are online black money transactions, dissemination of copyright movies and usage of illegal words. The law enforcement needs a system to handle this problem effectively. In this project we will be using the data mining techniques to analyses the suspicious and criminal activities that are happening in the forums. Our proposed system will be using the text data mining techniques like step word selection algorithm, stemming algorithm, emotional algorithm, text classification and ID3 Decision tree to analyses those suspicious activities by downloading the postings from selected discussion forums randomly and apply the above data mining techniques .all these are done using the text data mining concept and the system will decide the legal and illegal activities based on the admin instructions and will give the warning messages to the users who involved in illegal activities.

1. INTRODUCTION

In this chapter we are going to discuss about the features of our project. The basic features of the project have been discussed here. It includes about suspicious activities in the internet, how to handle the problems that was occurred in the existing systems, new effective data mining methods and text mining methods to control the illegal activities, over view of the project, problem definition, existing system, proposed system.

2. LITERATURE SURVEY

This paper presents a way for detecting skeptical discussions on the online forums, by which we will be able to encounter the users skeptical activities and engrossment. As we know that forum is a huge in conduct area, a way to convey the opinions of the users, about anything in their mind, with some sort of suggestions for advertise and relays. [1] every online forums suggests the user preferences by patterning their point of view. Observing your illegal activities is the only choice to know the users allegiance. The advancement in the modern technology also increases the usage of online forums. Forums are being used by millions of users, so this causes and aids the malicious people to send and share their illegal things through the online forums. In order to avoid all these illegal activities our system of law needed a system that tackles this problem, so this system downloads the postings which are mostly in text formats from the susceptible areas and detect the skeptical activities and

Give the users a warning message. In most of the virtual spaces contents are stored in text formats so the work lies mainly on text posts.

The algebraic and hypermedia technology influences human behaviors and communal interactions. This paper is all about developing a system that detects the illegal activities and interests of all users[1]. The real performances of online forums retrieval data analysis remain debatable due to lack of Information retrieval analysis from online forums. The purpose of this system is to monitor suspicious discussions automatically on online forum. Text analysis is used to detect suspicious posts in online forums [1]. This system also focuses on plan execution time, automated classification to identify more significant suspicious discussions.

In today's modern world social media's has become a place to talk about people's views and thoughts about the current issues which is happening in the world. The annoying posts or irritating posts causes violence and disturbs the peace in the society , These irritable posts refers to the religious, politics, high standard people's these things are not liked by the people and it causes violence among the people. So such posters or posts must be stopped or deleted before entering into the society which causes trouble or tension in the nation. So our article talks about the methods and possible techniques which is used to find the disreputable posts and rectify them as soon as possible before creating problems in the society. For that we used a Algorithm called "ID3 DECISION TREE"[2]. In which 200 sets of data's like

comments is taken and it is given to three people to check the abusive comments. After that they have to give the ratings to the comment accordingly by the scale factor from one to five, if they gave one it represents the comment is being less abusive. If they gave five it represents the comment is being more abusive. At last the average set of comments is calculated by the people and it is known as intensity factor. It is very successful process, which is mostly used in social media's like Instagram , Facebook , and Twitter which is most helpful and useful to remove the unnecessary comments in day today life activities.

In this article ,which is used to detect the mistrustful users who uses the bad text or words in the social media. This technique is mainly focused on Text Investigation. In order to identify the Malicious users. we need to check the user's comments , tweets, etc. Malicious users activity may also leads to terrorist activities[2]. So that we must check those details often. And there are many possible techniques and ways can be performed to detect those illegal activities.

the website involving some illegal acts with the use of quick internet. It can be thought as a series of actions. The illegal information is rapidly increased on the webpage and majority of illegal information are in the form of text.

In this paper is used to provide an evaluation of mining to extract useful information by using Data mining techniques. The series of step is to extract knowledge from large set of

data in the database for that they applying some artificial intelligence method to finding illegal information or unlawful act in the website which is make use of recognizing criminal characters[3].

Data mining is a suitable method for recognizing criminal characters which is in the form of text documents etc., Data mining tool is a very useful tool for point out the criminal characters and it is also support and very useful for police force. The first will be the recognizing the crime character after that we extract the useful information and knowledge by means of data mining techniques. The website involving an illegal act can be thought a series of actions. Therefore we can analyses the information from the view of actions to solve the problem in the website. The majority of information collected in the form of text like e-mails, chat system, text document, blogs etc.,

Basically, the text data is in the form of unstructured like missing letters so the researcher uses some tool to extract the knowledge and useful information in the text.

Therefore, the experience is important in analysis the data and search the information. Sometimes the information will be missed if it is having criminal hide.

3. PROPOSED SYSTEM

Monitoring Suspicious Discussions on Online Forum by Data- mining. The technique used is data-mining in which few data is extracted from huge amount of data. The system uses text mining to extract suspicious words from entire chat. The system not only provides the forum to chat but also helps to reduce the use of illegal words during chat and provides the database for criminal investigation, if any crime occurred by the person using that forum.

If the system detects the suspicious word in the chat the word is replaced by *** and if this occur thrice the user account is block for 24hours and if Account blocks for 3 times the user will be out of forum permanently.

The Social Networking sites are affecting human life. Hence this system successfully detects the suspicious words from chats and prevents the suspicious activities. This system is applicable to every department where there is need. Not only in social networking is sites this system applicable in forest department, disaster management system to prevent illegal activities. Text mining technology used to detect suspicious words from discussion forum. Figure 1 explained about the architecture of the proposed system. User must login in. After user logged in they must register after that input will be saved in the database and the process continues. Also Table 1 explained about the comparison of the proposed work with the existing system.

ADVANTAGES:

• This system will reduce illegal activities held on internet.

• This system will provide security for many websites.

• This system will act as an evidence for investigation

FEATURES:

- System will analyze online plain text sources using text data mining.

- This system can be used for security purpose.

- This system will scan post written in a natural language and will cluster this post into different groups and will identify illegal topics.
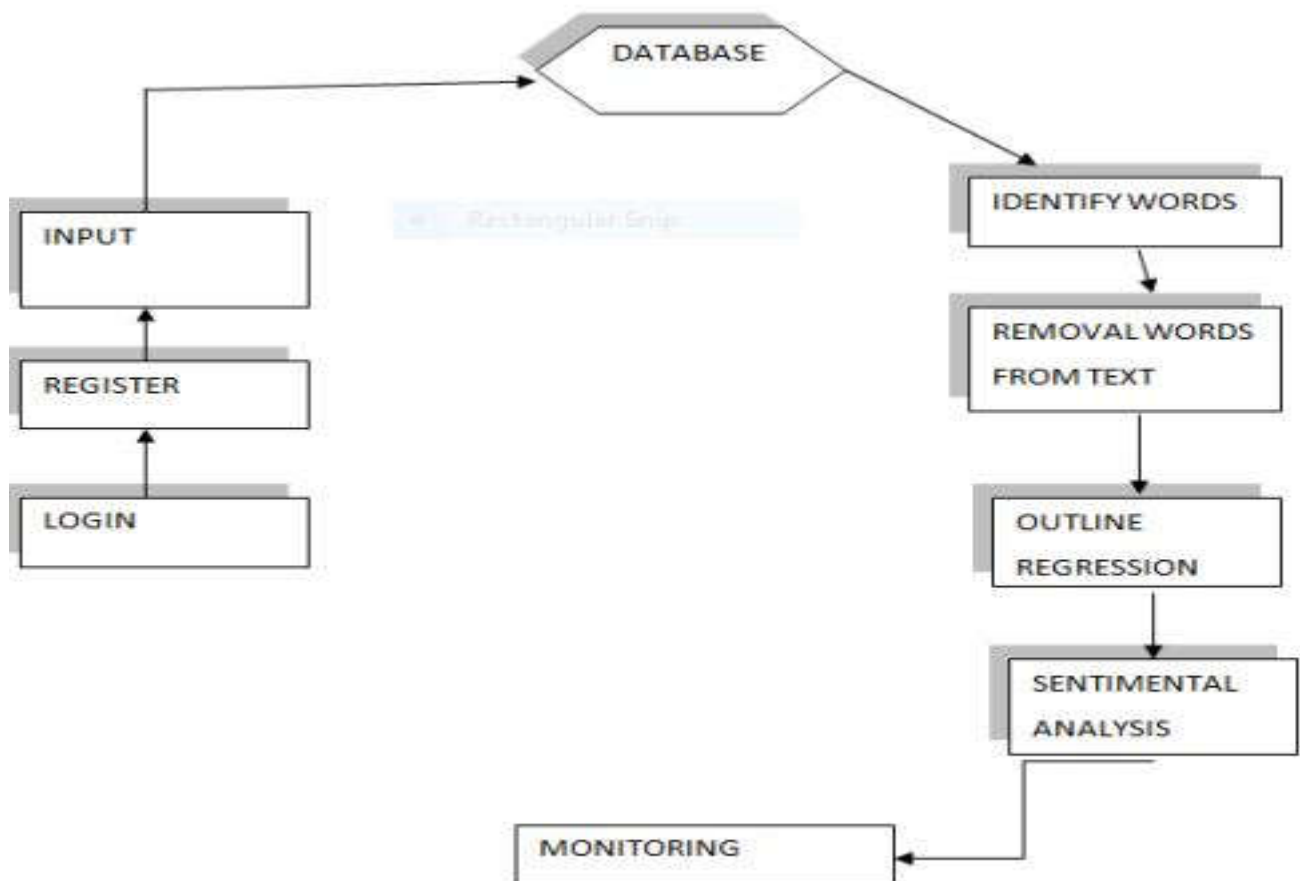

Figure 1: Architecture Diagram

TABLE OF COMPARISION:

EXISTING SYSTEM	PROPOSED SYSTEM
The internet having more technology to communicate between peoples one another. It having some channels and webpage for online discussion. Unluckily, some poisonous users take advantage of this technology. They use if for some unlawful purpose or illegal act or crime activities.	Monitoring Suspicious Discussions on Online Forum by Data- mining. The technique used is in which few data is data-mining extracted from huge amount of data. The system uses text mining to extract suspicious words from entire chat.
In internet especially social media, some users involving illegal act and they exchange in the form of skeptical post like video, image etc. The majority of illegal act in the form of text so we focus on the text data.	If the system detects the suspicious word in the chat the word is replaced by*** and if this occur thrice the user account is block for 24hours and if Account blocks for 3 times the user will be out of forum permanently

4. CONCLUSION

This article represents the methods for detecting suspicious discussions on the social media's. In this method we can identify the disreputable posts, activities, and illegal users.

The necessity of this system to monitor the illegal discussion and activity which takes place on social media's. Text Analytics, Text Mining, Step by Step Algorithm is used to identify the annoying posts in the social media's and the procedure is to collect the large data set and it is simplified, and it is implemented in this process. Also necessity of the article to give reviews about data mining technique how it is useful to identify such illegal activities.

REFERENCE

[1]. M.Suruthi Murugesan, R.Pavitha Devi, S.Deepthi, V.Sri Lavanya, Dr.Annie Princy, "Automated Monitoring Suspicious Discussions on Online Forums Using Data Mining Statistical Corpus based Approach," In Proceedings of 2016 IEEE Imperial journal of Interdisciplinary Research(IJIR), Volume 2, Issue 5,2016.

[2]. Algorithm to Monitor Suspicious Activity on Social Networking Sites usingData Mining Techniques by Suhas Pandhe and Suhil ,2015

[3]. Detecting Suspicion Information on the Web Using Crime Data MiningTechniques by J Hosseinkhani, 2014.

[4]. Salim Alami, Omar el Beqqali, "Detecting Suspicious Profiles using Text Analysis within Social Media," In Proceedings of 2015 IEEE Journal of Theoretical and Applied Information Technology, Volume 73, Issue 3,2015.

[5]. M. F. Porter. An algorithm for suffix stripping .Program, 14(3):130–137, 1980.

[6]. B. Connor, R. Balasubramanyan , B. R.Routledge , and N. A. Smith."From tweets to polls: Linking text sentiment to public opinion time series". In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media 2010.

[7]. M. F. Porter. An algorithm for suffix stripping .Program, 14(3):130–137, 1980.

[8]. Samuel Marquiz "classificateur de Kolmogorov sur le web " 7 Juin 2004.