International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

# The Facebook Analysis using Natural Language Processing (NLP)

**Ms. Megha S Anchatgeri[1], Mrs. Thejaswini S[2]**

*[1]Department of Computer Science, Siddaganga Institute of Technology, Tumkur, India*
*[2] Assistant Professor of Computer Science, Siddaganga Institute of Technology, Tumkur, India*

---

**Abstract -** *Digital development in modern society and the use of Information Technology which are very useful and creative become a big part in our daily life. On the other hand, it may lead to negative affect when it is used improperly in person. User data from social media in specifically Facebook whose shared posts are positive comments and negative comments. But, the negative side comments lead to negative affect personally or can be further extended to the community and nation level when the communication or transaction is used improperly. Therefore, it is necessary to identify users who have commented on the negative side that may be a lawbreaker on Computer related crime. On this which beneficial about investigation for legal proceeding and it facilitate for the police or people who take a part in the operation on law. The problem with this issue is, without forensic evidence, many offenders in complex criminal cases are still free and may repeat offending against the laws and harm the others. On the other side, A problem of children/ person being exposed to pornographic web sites on the internet has led to their safety issues. To prevent the children/person from these inappropriate materials, an effective web filtering system is essential. Therefore, the proposed system explores to classify the shared posts in to positive, negative comments, identifies particular user who have commented on the negative side on the online social networking specifically on the facebook and also handles the web filtering technique to filter pornography websites. To address the security problems and safety problems, the proposed system concentrates on the safety issues and security issues.*

***Key Words***: **Naive Bayes algorithm, Web Filtering, Image Filtering, Semantic Analysis, Support Vector Machine (SVM).**

## 1. INTRODUCTION

In today's situation, the advancement of computerized life in a present society and the utilization of Information technology will turn into an enormous effect in our progressing life. Here, the part of online networking will have affected to an individual to convey and in addition to share their views and emotions on open platforms. Some of the peoples or communities will jabber things on online systems it might prompts an adverse effect on social media as well as the community and national level where there is a possibility of tremendous exchanges. The client's information from web-based social networking like Face book, twitter and so on., are considered and presents a paper with respect to the constructive and pessimistic post that may prompt antagonistic effect for the people or group or national level. The reason for this paper is to recognize the clients who have remarked on the negative side that might be a culprit on Computer related crime. The one way is to control the negative impacts on social media is the Forensic Investigation with Naïve-Bayes classification algorithms calculations, Support Vector Machine(SVM). Without measurable confirmations, numerous culprits do illicit things and furthermore they will damage to others. Thus, in many developed countries such as Japan, European countries and the US have applied existing scientific knowledge and technology to identify evidences in criminal trials to gain true scientific results which is very useful in criminal investigations. Particularly in Japan, shows it having over 90% of the murderers are arrested by applying scientific equipment and invented and innovated technology to be used in forensic science. Recently, many researches present the social media data that are analysed in order to apply for correlation analysis and also the further result of existing data. Particularly from the study of researchers in forensic field, data analysis in forensic[1] shows that the social media data can be divided into major groups to be used in the study such as the researches that analyse data in text form, and the researches that analyse data in image form.

In Thailand todays, the use of social media is un-closed that the information of social media users in text form is popular and it also reflects the use of social network such as Facebook is highly accepted and its growth rate is multiple. The Naive Bayes calculations elucidate the information into two noteworthy parts. The initial part is related to share their feelings and furthermore their thoughts on positive side and the second part are on negative side. The motivation is to distinguish the clients who have commented on the antagonistic side which may cause unwanted impact personally or can be further extended to the group and country level. The problems of children being unmasked to porn-content websites such as on the public platform internet has led to their vulnerability issue. Pornographic websites without mining on public platform interpret can cause children to act out sexually against other children it can also mislead children's sexual mental state and adjustment, in order to prevent children from these ill-suited materials an efficient web filtering solution is essential.

To address the security problems and safety problems, the proposed system concentrates on the safety issues and security issues. Firstly, Safety issue is handled by applying the content based web filtering technique in particularly to filter the audio information's. Secondly, Security issue is handled by identifying a person (law breaker, offender) who have commented on the negative side by classifying shared posts into positive comments and negative comments. Thus, the proposed system provides a security and safety solution

by identifying particular users who have commented on the negative side, also helps the criminal analysts to render the offender to criminal justice and to prevent the children being exposed to pornographic websites respectively. To address the security problems and safety problems, the proposed system concentrates on the safety issues and security issues. Firstly, Safety issue is handled by applying the content based web filtering technique in particularly to filter the audio information's. Secondly, Security issue is handled by identifying a person (law breaker, offender) who have commented on the negative side by classifying shared posts into positive comments and negative comments. Thus, the proposed system provides a security and safety solution by identifying particular user(s) who have commented on the negative side, also helps the criminal analysts to render the offender to criminal justice and to prevent the children being exposed to pornographic websites respectively.

According to our survey this is the first proposed paper which includes security and safety issues in NLP. Therefore, the proposed system provides a security and safety solution by identifying particular users who have commented on the negative side, also helps the criminal analysts to render the offender to criminal justice and to prevent the children being exposed to pornographic websites.

The organization of the paper is as follows: In section II discussion of background and related work is discussed. Proposed methodology is discussed in section III and finally, conclusions is discussed in section IV.

## 2. LITERATURE SURVEY

**Uraz Yavanoglu, Busra Caglar, Ozlem Milletsever, Medine Colak, Semra Cakir and Seref Sag_roglu , vol. 58, 2010.** Proposed the Intelligent Approach for Identifying Political Views over Social Networks. It is a research-based analysis of political views by analyzing Social Network data through Artificial Neural Networks: ANN model and Data mining. The data used in this research is taken from Twitter which is a public data. Therefore, this work helps to analysis thoughts and ideas from Twitter users both supporting or opposing the government [2].

**V. B. Cline. Pornography's Effects on Adults and Children. In New York: Morality in Media, vol. 11, 1990.**The growth in the number of pornographic web sites coupled with the increase in the number of children using the internet leads to a heated debated about safety issues. From several studies, the psychologists show that pornography can make the children be victims of sexual violence and cause them to act out sexually against other children. Moreover, pornographic exposure can mislead child's sexual attitude and orientation [3].

**H. Saif, Y. He, and H. Alani, Semantic Sentiment Analysis of Twitter. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.** Introduces a method based on the disambiguation of named entities to improve the performance of systems of sentiment analysis of Twitter. Their approach consists in adding semantic features to the named entity identified (for example, the word "product" is associated with "iPhone" and "iPhone with "Apple") and then associate negative/positive sentiment extracted from tweets. Semantic features extracted from tweets are incorporated into a naive Bayes classifier. Three other types of features are compared: unigram features, part-of-speech features and sentiment-topic features. Three corpuses composed of tweets are tested: Stanford Twitter Sentiment Corpus (STS), Health Care Reform (HCR) and Obama-McCain debate (OMD). They obtain best precision (84.5%) when semantic features are incorporated into the classifier, with the STS corpus [4].

**A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," Business horizons, vol.53, no. 1, pp. 59–68, 2010.**Cyberbullying can be defined as aggressive, intentional actions performed by an individual or a group of people via digital communication methods such as sending messages and posting comments against a victim. Different from traditional bullying that usually occurs at school during face to-face communication, cyberbullying on social media can take place anywhere at any time. For bullies, they are free to hurt their peers' feelings because they do not need to face someone and can hide behind the Internet. For victims, they are easily exposed to harassment since all of us, especially youth, are constantly connected to Internet or social media [5].

**S. J. Kavanagh, Protecting Children in Cyberspace. pp 58-59 Springfield. VA: Behavioral Psychotherapy Center. 1997.** Pornographic exposure can persuade child's attitude into a bad way and mislead child's development. According to Dr. Victor Cline (a psychologist) reported that, there are a certain critical period of childhood where a child's brain is being programmed for sexual orientation. During this period, the mind will develop a feeling for what the person will be attracted to. Being exposed to healthy sexual attitudes during this critical period can result in the child development in a healthy sexual way. Conversely, exposure to deviant pornography during this period can permanently imprint the abnormality of the child sexual orientation [6].

**M. Thelwall, D. Wilkinso, S. Uppal, "Data Mining Emotion in Social Network Communication: Gender Differences in MySpace", Journal of the American Society for Information Science and Technology, 2010.**Thelwall et. al. constructed an intelligent model for sentiment analysis from MySpace member's comments with the help of data mining. The data mining approach consists of opinion mining or sentiment analysis phases. The methodology is to have opinion mining approach to gender differences in the expression of emotion applied for MySpace sites member's comments. The sample dataset was collected from MySpace social network sites member's profiles containing 819 public comments or from 387 comments of normal U.S MySpace members. The experiment result showed that two thirds of the comments included positive emotion excluding 9% of 20

parts. Females have more positive attitude than males for positive comments but there is no distinction for negative comments by female or male [7].

**Aniello Castiglione, Giuseppe Cattaneo and Alfredo De Santis** Proposed A Forensic Analysis of Images on Online Social Networks which is a research-based analysis of visual information that may violate the copyright law or engage in illegal activity on social networks. In this research, the analysis mainly focuses on processing the uploaded images and what changes are made to some of the characteristics. The pixel resolution and related metadata are studied together [8].

**Zhaochun Ren, David van Dijk, David Graus, Nina van der Knaap, Hans Henseler, Maarten de Rijke. Semantic Linking and Contextualization for Social Forensic Text Analysis. IEEE 2013.** proposed the Semantic Linking and Contextualization for Social Forensic Text Analysis which is a research-based analysis of the connective data between two social networks and after that analysis data connection setting within the context [9].

**Shankar Setty, Rajendra Jadit, Sabya Shaikh, Chandan Mattikalli and Vma Mudenagudi. Classification of Facebook News Feeds and Sentiment Analysis. IEEE 2014.**A learning based classifier is built using various classification algorithms such as Bi-nary Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Bayes Net and J48 and This experiments on the 2016 live news feeds showed that the proposed approach could achieve significantly improved performance for structuring the data on Facebook using SVM classifier learning model [10].

## 3. PROPOSED METHODOLOGY

Forensic science is the applications of scientific principles to matters of legal problems and criminal investigation. The Law (the rule) is a legal principle enacted by authoritarians to force people to follow unconditionally, and any person who fails to comply with the law will be provided adverse actions. The Laws may be enacted to define the rules of the relationships between people or people and the State or in governing the country.

The main objective of the proposed system is to classify the shared posts in to positive comments and negative comments, identifying particular user who have commented on the negative side in specifically on the facebook, Prevents children/person from inappropriate materials(audio) by filtering. Figure 1 below shows the proposed methodology.

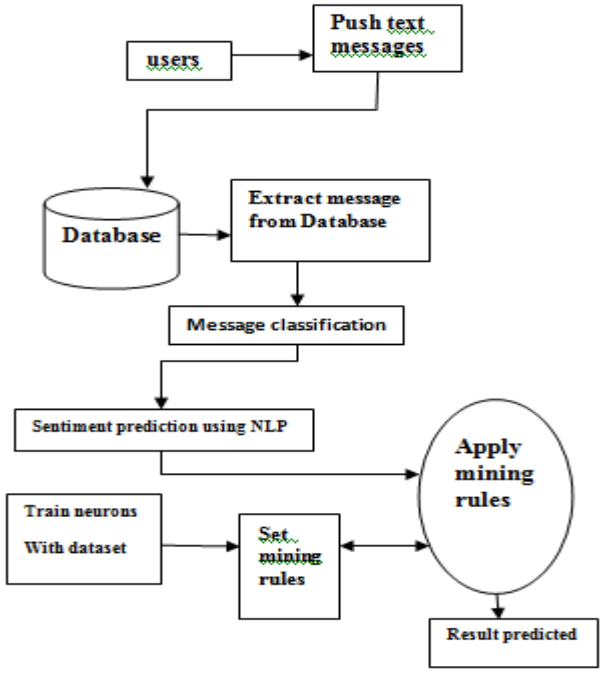


**Figure 1:** shows the proposed mythology of facebook analysis

The proposed system consists of the following components/Phases:

**Phase 1**(Data collection): In this phase, users can send a friend request to all the people and only those who accept his/her friend request, the user can start chatting with him/her. In the same manner other people can also send a friend request and it is up to the user to accept the friend request or delete the friend request. If the friend request sent by the user is deleted then the user is not able to chat with that person. Once the friend request is accepted among users then the users can share images, post messages and can start chatting with their friends.

**Phase 2** (Filtering): In this phase, data gets stored in the database, the transaction between the messages in the database is viewed by creating an interface and query is written. Natural Language processing(NLP) is used to do sentiment analysis and is stored in the database. Weka tool is used to predict the data and also stored in the database gathers all the text messages from the database, and then the classification of the messages is done, for the classified messages the sentiment prediction is done [11]. Based on training the neurons using training dataset, finally the result of the messages is predicted.

**Phase 3**(Analysis): Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. Training dataset is used to train

the machine, then for the test dataset that is the messages entered by the user the data mining rules [12] are applied and then finally the result is predicted using J48 algorithm.

The working of the forensic analysis using social media process is as follows**:**

Step 1: Social media data is taken & divided into two categories namely data type and its attributes. Data type contains User detail and opinion detail and its attributes are User Id, User Name and Opinion Id, Opinion Text, Time Create respectively.

Step 2: Prepare forensic keywords based on positive & negative views. [13]

Step 3: Define weight for each forensic keyword. The Table1 shows the different forensic keywords and their weights respectively.

**Table 1:** Example of forensic keywords

| Forensic Keyword | Weight | Forensic Keyword | Weight |
|---|---|---|---|
| ความรัก ( love) | 1 | คิดถึง (miss) | 8 |
| น้ำใจ (kindness) | 2 | ชอบ ( like) | 9 |
| ขอบคุณ (thanks) | 3 | หัวใจ (heart) | 10 |
| ปรานี (mercy) | 4 | ยิ้ม (smile) | 11 |
| ยินดี ( joy) | 5 | สวย (pretty) | 12 |

Step 4**: Search and Data Extraction**. At this stage, the divided Forensic keyword data is processed to search, compare, and find out the Frequency of Forensic keyword that are shown by the user reviews. To process on this, the String Matching technique is used in mapping Forensic Keyword together with Opinion Text data in order to prepare for data analysis in next step.

Step 5: **Data analysis:** In this process, the information examination process behavior is done by grouping with Naïve Bayes procedure that a standard of likelihood measure is connected to clarify the exploration technique. The posterior probability method is used to predict an attribute has been occurred in the data set or not. If the probability of an attribute is 0 indicates that the pattern for their attribute has not occurred in the training data set. Based on the positive & negative opinions results have been classified in to two classes:1) class positive: social media users whose views are positive. 2) class negative: social media users whose views are negative.

Step 6: **Getting result of analysis**. In this procedure, it is an analysis of the results from using Naïve Bayes technique in analysing social media users based on positive & negative views.

## 4. CONCLUSION

Natural language processing is a branch of artificial intelligence & computer science and it uses text mining to make the interaction between human and computer, though its purpose is to have interaction among natural language of human beings and computers. The proposed system explores to classify the shared posts in to positive, negative comments, identifies particular user who have commented on the negative side on the online social networking specifically on the facebook and also handles the web filtering technique to filter pornography websites. To address the security problems and safety problems, the proposed system concentrates on the safety issues and security issues. Firstly, Safety issue is handled by applying the content based web filtering technique in particularly to filter the audio information's. Secondly, Security issue is handled by identifying a person (law breaker, offender) who have commented on the negative side by classifying shared posts into positive comments and negative comments. The effective Naïve-Bayes classifier, Support vector machine, Auto MLP is used in order to classify positive side and negative side group of users Thus, the proposed system provides a security and safety solution by identifying particular user(s) who have commented on the negative side, also helps the criminal analysts to render the offender to criminal justice and to prevent the children being exposed to pornographic websites respectively. Hopefully, the researchers are confident that this research should be further developed for the benefits of forensic process as well as building peaceful coexistence in today's society.

## REFERENCES

[1] Uraz Yavanoglu, Busra Caglar, Ozlem Milletsever, MedineColk, Semra Cakir, Seref Sag_roglu. Intelligent Approach for Identifying Political Views over Social Networks. IEEE 2014.

[2] Chris Howden, Lu Liu, Zhijun Ding, Yongzhao Zhan, KP Lam. Moments in Time: A Forensic View of Twitter. IEEE 2013.

[3] Shankar Setty, Rajendra Jadit, Sabya Shaikh, Chandan Mattikalli and Vma Mudenagudi. Classification of Facebook News Feeds and Sentiment Analysis. IEEE 2013.

[4] Zhaochun Ren, David van Dijk, David Graus, Nina van der Knaap, Hans Henseler, Maarten de Rijke. Semantic Linking and Contextualization for Social Forensic Text Analysis. IEEE 2013.

[5] Noora Al Mutawa, Ibtesam Al Awadhi, Ibrahim Baggili and Andrew Marrington. Forensic artifacts of Facebook's instant messaging service. IEEE 2012.

[6] Reza Soltani, Abdolreza Abhari. Identity Matching In Social Media Platforms. IEEE 2011.

[7] Aniello Castiglione, Giuseppe Cattaneo, Alfredo De Santis. A Forensic Analysis of Images on Online Social Networks. IEEE 2011.

[8] Norulzahrah Mohd Zainudin, Madjid Merabti. Online Social Networks As Supporting Evidence A Digital Forensic Investigation Model and Its Application Design. IEEE 2011.

[9] M. J. Jones & J. M. Rehg. Statistical Color Models with Application to Skin Detection. International Journal of Computer Vision. 2004.

[10] R. R. Korfhage. Information Storage and Retrieval. John Wiley & Sons. Canada. 2004.

[11] Office of the Royal Society. [Online]. Dictionary. Source: IEEE 2000

[12] Eakasit Pacharawongsakda. [Online]. Data Mining Trend. Source: IEEE 2000

[13] D. R. Hughes. The Internet safety site of Enough.org. http://www.protectkids.com/, Last accessed: February 1999