

# **FACEBOOK POST CLASSIFICATION AND ADULT CONTENT CHECK AND CATEGARIZATION USING HADOOP**

**<sup>1</sup>Aishwarya Pande, <sup>2</sup>Pranoti Patil, <sup>3</sup> Pooja Pawar, <sup>4</sup> Anil Rasal**

*<sup>1,2,3,4</sup> Student, Computer Department, Rajarshi Shahu College of Engineering ,Pune, (India)*

## **ABSTRACT**

*The proposed work is to avoid the adult contents and accessing the Facebook posts from the developers account and Tokenizing, The posts and removal of stop words by apply stemming. This system is used to classify posts in to adult content or not, based on adult words which are stored in a file. Categorization of the posts on the basis of “sports”, “politics”, “business”, etc. The data which is fetch from the Facebook and the data is divided and the every block of memory and each data block will have a token and process according to the token data is processed stop words filtering: consists in eliminating stop-words.*

***Index Terms: Adult, Business, Categorization, Classification, Politics, Sports, Streami.***

## **I INTRODUCTION**

To develop a system that classifies Facebook posts on Hadoop (installed on cloud) with the help of NLP. Extract texts from posts, images, videos of social media as Facebook to know how people feel about different posts and do sentiment analysis, using Hadoop Classifying the post gives probable effect of the original post, so its easier to understand social affect of any post on Facebook or any such social media. So we focused on Facebook statuses, which we can view as opinions of users or their reaction on concern we want to analyze. We develop tool status puller that automatically collects random Facebook statuses. Then we make classifier that performs classifications on that corpus collected from Facebook. The dramatic and exponential growth of content available on web and its classification has now become an efficient methodology to make the contents of large repository in an organized manner. Social networking websites are the new era of expressing views. Today every fifth person put their opinions, views, comments on these micro-blogging and social sites like, FACEBOOK and many more. Now people are using internet as a communication tool among their social network including friends, family, friends of friends to these micro-blogging and social network sites. In this we gradually put and share their opinions among their friends on these sites which finally becomes huge and relevant repository for any of particular entity or organization.

## II LITERATURE SURVEY

[1] **M. Lovelin Ponn and Felciah R.Anbuselvi “A Study on Sentiment Analysis of Social Media Reviews “,IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIIECS**

**Abstract-** With the rapid growth of web technology, there is a huge volume of data present in the web for internet users. It also become the place for online learning and exchange ideas .The information gathering has become more important to find out what other people think about any product, service or organization. This gives an enormous growth and availability of rich-opinion data in various social media such as on-line review site, twitter, face book and personal blogs and so on. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which yields to computational treatment on various opinion and sentiment in text. This survey focus on various techniques and methods for classifying the opinion from social media datasets and its future aspects.

**Keywords:** Sentiment Analysis, social media, classification, Opinion mining.

**In this paper they have proposed :**

A sentiment analysis of social media review

**From this paper we have referred :**

Finding sentiment in a text

[2] **Rajdeep Singh, Roshan Bagla and Harkiran Kaur” Text Analytics of Web Posts Comments Using Sentiment Analysis”, 2015 IEEE International Conference on Text Analytics IEEE, 2015.**

**Abstract-**Social networking websites are the most common platform to share and voice one’s opinion. The posts and the corresponding comments written on one's wall, such as facebook assist people in decision making under various situations based on the opinions of other people. Many a times these comments influence one's thought process while making decisions due to higher number of likes and comments viewed on such posts. It has often been observed that users press the like button on any post within fractions of seconds without even completely reading the post. The outcome is though the user likes a particular post but enters a negative comment for it. This paper proposes and implements the sentiment analysis approach to determine the actual popularity of a post on social networking websites. This approach will provide the actual statistics to facilitate the adjudication, that is, if the idea or thought conveyed by the user supports the post or not. The comments are analyzed using a Lexicon based sentiment analysis approach. In this approach, the actual numbers of negative and positive comments are discovered, thus providing the actual statistics and preventing a post from attaining false admiration.

**Keywords:** Sentiment Analysis; Social Networking; Text Analysis; Facebook; Lexicon based approach

**In this paper they have proposed :**

A analysis of web post and comment

**From this paper we have referred :**

Post categorization and analysis

**[3]. Po-Cheng Huang , Jain-Shing Wu and Chung-Nan Lee “Negative Emotion Event Detection for Chinese Posts on Facebook”, 2015 International Conference on Cloud Computing and Big Data.**

**Abstract :** As most people would like to post their articles in social network to express their feeling, it would benefit to collect and analyze these information to figure some signs before some misfortunes happened. Hence, in this paper, we propose a novel emotion analysis system not only to detect the Chinese posts with negative emotions on facebook in time sequence but also extract the places related to those posts. 9820 posts from facebook are used as a training set and 2334 posts from facebook as testing samples to verify the system accuracy. Experimental results show that the precision of negative emotion classification of the proposed system is 74.8%, and the recall rate is 78.7%, both of the precision and recall the proposed system are better than traditional methods (SVM and Naïve Bayesian) 8%~17%. In addition, the proposed system is not only able to extract the posts with negative emotions, but also to find the correlation between emotion and places.

**Keywords:** negative emotion detection; data mining; emotion classification I.

**In this paper they have proposed :**

A percentage of detecting the emotion as negative and positive post

**From this paper we have referred :**

Emotion detection for post on facebook, negative post detecting technique

**[4]. Deepika Sherawat “ Mining Emotions (Anger & Fear) from Indian Army Fans Page on Facebook”, 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA) IMS Engineering College, Ghaziabad, India**

**Abstract:** Usage of social media is becoming an inescapable part of all of us. When people with similar interest interact with each other they eventually lead to a discussion. While a person expresses his point of view then there is an impact of his words those who are listening by having some perceptions in their mind. How a person perceives or accepts any view of some other person also has an impact on their thinking as well. They either have a positive impact or a negative impact on them, and this impact leads to some positive or negative comments by that person. And at certain stages the negativity in the group increases so much that it leads to some ill-impacts on the society.

**Keywords :** SNS, EmoAngrFear, Anger, Fear, Opinion Mining

**In this paper they have proposed :**

The moods of emotion from post on facebook

**From this paper we have referred :**

The moods detection technique

[5]. Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen and Xiaofei He” Interpreting the Public Sentiment Variations on Twitter”, IEEE transactions on knowledge and data engineering.

**Abstract:** Millions of users share their opinions on Twitter, making it a valuable platform for tracking and analyzing public sentiment. Such tracking and analysis can provide critical information for decision making in various domains. Therefore it has attracted attention in both academia and industry. Previous research mainly focused on modeling and tracking public sentiment. In this work, we move one step further to interpret sentiment variations. We observed that emerging topics (named foreground topics) within the sentiment variation periods are highly related to the genuine reasons behind the variations. Based on this observation, we propose a Latent Dirichlet Allocation (LDA) based model, Foreground and Background LDA (FB-LDA), to distill foreground topics and filter out longstanding background topics. These foreground topics can give potential interpretations of the sentiment variations. To further enhance the readability of the mined reasons, we select the most representative tweets for foreground topics and develop another generative model called Reason Candidate and Background LDA (RCB-LDA) to rank them with respect to their “popularity” within the variation period. Experimental results show that our methods can effectively find foreground topics and rank reason candidates. The proposed models can also be applied to other tasks such as finding topic differences between two sets of documents.

**Keywords :** Twitter, public sentiment, emerging topic mining, sentiment analysis, latent Dirichlet allocation, Gibbs sampling

**In this paper they have proposed :**

The interpretation of sentiment variation

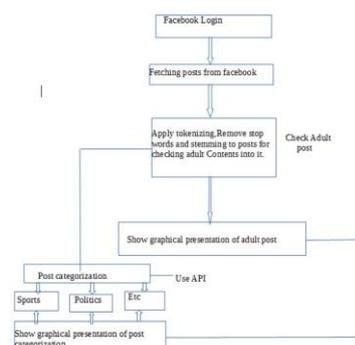
**From this paper we have referred :**

Data mining technique

## III PROPOSED TECHNIQUE

### 3.1 Working

It is real time system. first we have to login to the facebook . then fetching of data is done. Then it apply the tokenizing technique.and remove stopwords , streamings . The it will do post for checking adult comments into it. After classification the graph is generated. At the same time categorization of post is done in different categories like sports, politics etc. Then it shows the graphical representation of categorized posts.



## Fig.1.System Architecture

### 3.2 Algorithm

Step 1: Login with normal user who have the facebook account along with the account added into facebook developer.

Step 2: Add Posts on the facebook account for easy retrieval of posts.

Step 2.1: Get U (Posts that is mostly related to Adult content) as Input to IE.

Step 2.2 : for i=0 to MAX

//MAX = Maximum no of Posts

Step 2.3 : visit i (facebook portal) when logged in by user .

Step 2.4 : go to step 2.2 till MAX

Step 2.5 : Output as CP is able to view all policies and schemes.

Step 3: call RC function

Step 4: call to IE Function

Step 4.1 : Get CP (Facebook posts) as Input .

Step 4.2 :Call Function NLP

Step 4.3: Process NLP as Removing Stop Word.

Step 4.4 : Get Relevant Information of positive, negative and neutral comments along with the adult content check.

Step 5: Display Result in the graphical form how many posts contains adult content.

Step 6: Stop.

### 3.3 Mathematical Model

$S = \{S, s, X, Y, T, f_{main}, DD, NDD, f_{friend}, memory\ shared, CPU_{count}\}$

**S (system)**:- Is our proposed system which includes following tuple.

**s (initial state at time T )** :-GUI of Facebook post classification,adult content check and categorization using Hadoop. The GUI provides space to enter a query/input for user.

**X (input to system)** :- Input Query. The user has to first enter the query. The query may be ambiguous or not. The query also represents what user wants to search.

**Y (output of system)**:- List of URLs with Snippets. User has to enter a query into Facebook post classification,adult content check and categorization using Hadoop then Facebook post classification,adult content check and categorization using Hadoop generates a result which contains relevant and irrelevant URL's and their snippets.

**T (No. of steps to be performed)**:- 6. These are the total number of steps required to process a query and generates results.

**f<sub>main</sub> (main algorithm)** :- It contains Process P. Process P contains Input ,Output and subordinates functions. It shows how the query will be processed into different modules and how the results are generated.

**DD (deterministic data)**:- It contains Database data. Here we have considered

Facebook posts classification which contains number of ambiguous queries. Such queries are user for showing results. Hence, Facebook posts classification is our DD.

**NDD (non-deterministic data):-** No. of input queries. In our system, user can enter numbers of queries so that we cannot judge how many queries user enters into single session. Hence, Number of Input queries are our NDD.

**f<sub>friend</sub> :-** WC And IE. In our system, WC and IE are the friend functions of the main functions. Since we will be using both the functions, both are included in f<sub>friend</sub>

function. WC is Web Crawler which is bot and IE is Information Extraction which is used for extracting information on browser.

**Memory shared:** - Database. Database will store information like list of receivers, registration details and numbers of receivers. Since it is the only memory shared in our system, we have included it in the memory shared.

**CPU<sub>count</sub>:** - 2. In our system, we require 1 CPU for server and minimum 1 CPU for client. Hence, CPU<sub>count</sub> is 2.

#### **Subordinate functions:**

Identify the processes as P.

$$S = \{I, O, P, \dots\}$$

$$P = \{RC, IE\}$$

Where,

WC is the online Web Crawler technique where posts from facebook are being fetched.

IE is Information Extraction.

P is processes.

$$WC = \{U, MAX, CP\}$$

## **IV CONCLUSION**

In this project accessing the Facebook post from the developers account. tokenizing the posts and removal of stop words, apply stemming. Classify posts into adult content or not, based on adult words. Categorization the post on the basis of “sports”, “politics”, “business”, etc. To find the categories of the post API has been used. Final output will be on the basis of adult post graph and category of post graph. At the end Facebook posts has been stored in to Hadoop for further analytics. In this project accessing the Facebook post from the developers account. tokenizing the posts and removal of stop words, apply stemming. Classify posts into adult content or not, based on adult words. Categorization the post on the basis of “sports”, “politics”, “business”, etc. To find the categories of the post API has been used. Final output will be on the basis of adult post graph and category of post graph. At the end Facebook posts has been stored in to Hadoop for further analytics

## REFERENCES

- [1] M. Lovelin Ponn and Felciah R.Anbuselvi “A Study on Sentiment Analysis of Social Media Reviews “,IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIECS
- [2] Rajdeep Singh, Roshan Bagla and Harkiran Kaur” Text Analytics of Web Posts Comments Using Sentiment Analysis”, 2015 IEEE International Conference on Text Analytics IEEE, 2015.
- [3]. Po-Cheng Huang , Jain-Shing Wu and Chung-Nan Lee “Negative Emotion Event Detection for Chinese Posts on Facebook”, 2015 International Conference on Cloud Computing and Big Data.
- [4]. Deepika Sherawat “ Mining Emotions (Anger & Fear) from Indian Army Fans Page on Facebook”, 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA) IMS Engineering College, Ghaziabad, India
- [5]. Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen and Xiaofei He” Interpreting the Public Sentiment Variations on Twitter”, IEEE transactions on knowledge and data engineering.