

# Analysis of Social Media Behavior in Banking Institutions using Text Mining

<sup>1</sup>D. Hemavathi, <sup>2</sup>Dr.H. Srimathi

**ABSTRACT---***Social networks are the foremost vital communication channels in recent years, that well-liked among the various social teams. These networks affected the ideas and policies of people, teams and communities. Every day, a lot of tweets on twitter being revealed. These tweets mirror opinions and beliefs of their publishers and have an effect on others in addition. Therefore, it's vital to investigate these tweets and establish and classify trends of various users. A set of options have been extracted to portray every cluster exploitation various text mining strategy and store these characteristics within the info. Text mining, sentiment analysis, and opinion mining techniques are wont to accomplish this extraction. During this paper, the important time tweets were fetched on banking establishments, Later the fetched tweets applied to many steps to investigate the social media behavior by using text mining.*

**KEYWORDS---***Social media behavior, implementation, banking institutions, twitter data, Text mining*

---

## I. INTRODUCTION

Industries and organizations don't have demand to store and perform operations and analytics on information of the clients in the last ten years. Even though the necessity came from 2005 to enhance all data into information is route interested to meet the requirements of the individual people. So large data came in form of image among the \$64000 time business analysis of data processing. From twentieth century ahead this Modern internet has changed the means that of expressing their views. Gift situation is completely they are communicating their issues through on-line sites, discussion forms and conjointly some on-line applications like Twitter, Face book so forth. Within the event that we tend to simply accept Twitter as our illustration concerning 1TB of knowledge is generating each week with the different types of tweets. So, by this, it's additional clearly anyway this internet is systematically change the strategies for living and class of individuals. Among these tweets are going to be sorted by the hash value labels that they are remarking and posting their tweets. During this approach, as of currently many enterprises and conjointly the summary company zone units abuse this for taking part in out some investigation determined they're going to foresee the success rate of their product or in addition they're going to show the different read from the data that they have collected for analysis. But, to calculate their views are extraordinarily hard throughout a conventional means that by taking this serious data that unit of measurement near to generate day by day.

---

<sup>1</sup>Assistant Professor, SRM Institute of Science and Technology, Tamil Nadu, INDIA  
(Email id : hemavatd@srmist.edu.in)

<sup>2</sup>Assistant Director- Admissions, SRM Institute of Science and Technology, Tamil Nadu, INDIA  
(Email Id : srimathh@srmist.edu.in)  
(Corresponding author: D.Hemavathi : Email id: hemavatd@srmist.edu.in)

Text mining has clad to be a most popular methodology to estimate and understanding immense datasets not done by the traditional techniques. These tools are faced extent of information problems, kind of like the understanding topics in social media or give the way to retrieve the information in unstructured data. Text mining is additionally a extraordinarily great tool within the early stages of analysis exploration, granting the matter data to counsel issues and ideas to the data scientist in the course of the analysis. what is more, these tools may also aid in development and structuring text-based data for future investigation in illustration or entirely completely different graphical techniques. And, besides the tangible analysis blessings, text mining is additionally a abusive and fruitful system of discovery. Text Mining is that the most frequent however robust task faced by the consultants of IP / analytics. The foremost necessary challenge is one should fully assess the underlying patterns in text, that too manually.

## II. Literature Survey

Researchers [1] concisely defined a method of knowledge collecting and therefore the chief areas of information discovery, knowledge mining, and information visual image from Twitter information. In this paper author creates a twitter app and it fetched in the real time twitter tweets on a particular topic and stored it into R and then it can apply several text mining procedures on the tweets to pre-process the tweet text and then it is analyzed the pre-process data by picturing them.

Reflection of some background information about text mining and how natural language inputs can be transformed into useful new knowledge. Also, different processes and steps required for sentiment analysis are investigated and some of them are implemented. Four different methods [2] were used to achieve the task and they vary from normal word scoring and classification of more advanced methods i.e. Recursive Tensor Neural Networks. Public sentiment analysis based on twitter data of staple food price variations in Indonesia has been analyzed and examined in various ways[3]. Text mining is used for ordering tweets into negative and positive sentiment. After that link between the kind of staple foods and sentiment classes were evaluated using Marascuillo procedure and Chi Square test [3]. Results show that Support Vector Machine (SVM) classifier yields higher accuracy than a Decision Trees and Naïve Bayes. Also, the price of red onion, eggs and milk had the most substantial association of the negative sentiment paralleled to other commodities.

Some of the researchers investigated that the usage of TM on policy official papers fell short in both quantitative and qualitative aspects [4]. The consumption of IE to extract allusions between policies, together with the use of automatic summarize and keyword extraction, to retrieve a concise illustration of the official papers and a set of eloquent labels (tags) respectively, were evaluated both statistically and by experts. The results showed that to a large extent, these techniques are capable of analyzing internal policies and extracting reliable information from them and it represents the harmonious use of three different techniques and the results that derive from their utilization.

Dissimilar text mining approaches to get numerous textual commands on social media websites. One of the surveys [5] defines how studies in social media have utilized text analytics and text mining techniques for the intention of isolating the crucial themes in the data. And focused on investigating the text mining studies interrelated

to the face book and twitter, the two leading social media in the world. As a result can serve as the baselines for future text mining research. Opinion scores are used to categorize the tweets into neutral, positive, and negative classes. Then, utilizing various machine learning classifiers the accuracy of anticipated classification with respect to actual classification is being estimated and compared using a controlled learning model [6]. Along with constructing a sentiment classification model, investigation of tweets is being carried out by picturing the word cloud of tweets using R.

Text mining tool to analyze the major social media sites, with YouTube, LinkedIn, Face book, Twitter and others, with a focus on a model of hotels. The measurements analyzed were passion and sentiments reach. A requirement was found between numerous variables obtained by financial performance and text mining. The results indicate that analysis of social media using these techniques can be a method to improve financial performance [7].

Accuracy is influenced by interactions between LDA generated topic models and the correct construction of both dictionaries [15]. These results also reveal that text mining can produce new visions into variables that have been broadly studied in hospitality industry, including that hotel food generate ordinary positive sentiments for the case studied, while hospitality makes both strong positive and ordinary feelings.

Some Reviews interprets the Client retention is the best important factor to be investigated in today's competitive business environment [8]. And also scam is a significant problem in the banking sector. Identifying and preventing fake is difficult, because cheats develop new patterns every time, and the patterns grow more and more erudite to elude easy detection. This paper evaluates the data mining skills and its uses in banking sector like risk management, marketing, customer retention, and detection and fraud prevention.

Some research studies accompanying to detect monetary fraud using data mining tools within one period and communicate the existing trends to academic specialists and industry, physicians, the prominent data mining tool in detecting monetary fraud with a 13 percent of usage [9].

### **III. OBJECTIVES**

- To gather tweets about the Banking Institutions
- To normalize and clean the retrieved tweet
- To analyze the normalized tweets.

### **IV. METHODOLOGY**

Text mining assists a corporation to derive most likely valuable business awareness from text matter of a document like email, postings and word documents on social media streams like Twitter, Face book and LinkedIn. Text mining plays a crucial role in creating decisions, through this method we will analyze the data and on the proposal of the result. Presently social media sites like twitter are maximum used to share user opinions on varied topics, twitter offers a platform to user to contribute their thoughts on the Banking sector and there are petabytes of information generated by twitter per day.



Figure 1: Analytical approach of text mining

Tweets are flexible for practitioners and research people through the public Twitter APIs. Twitter REST API and the Streaming API are used to collect data for developers. The text mining methods are used to analyze the large amount of twitter datasets which affords a better approach for decision making. Bayesian method uses multiple- measures information to calculate approximately the significant determinants of consumers' preferences for competing banks in the market. Feature Selection represents the behavioral characteristics along with measures of liquidity, safety, profitability and efficiency of the banking institutions under consideration [13]. The discriminant analysis is essential tool to be used by any bank, because it answers to various concerns that the managers and the people cooperating with a bank may encounter. ROC analysis and accuracy parameters can be applied for the business analysts to easily observe the expected downstream issues and benefits of negative and positive test results based on these test modalities, and directly link the classification accuracy to important decision making.

## V. TWITTER DEFINITION

In micro blogging, the word 'micro' specifies the constraint of view of content conveyed. A twitter user will produce at most of one hundred forty characters for every tweet. A tweet is not only a basic instant message but it's a combination of Meta data and text data connected with the tweet. These are the highlights of tweets. They communicate the content of the tweet or what's that tweet concerning. The Meta data can also discover the location of the tweet. The information of tweet are many entities and spots. These entities incorporate hash tags, user mentions, media Users, URLs and, Twitter user ID. RT meant for retweet, '@' trailed by a user identifier report to client, and '#' trailed by a word portrays a hash tag.

## VI. FEATURES OF TWITTER

Some of the features are useful for Opinion Retrieval Process:

### Twitter Specific Features

#### URL

Several tweets carve up a link beside with the commencement to the links. The sharing link is started as URL. Existence of URL, contribute its characteristic value as 1, otherwise 0.

#### Mention

When user must advert to a different one user in a tweet he will compose his name starting with @symbol. it's referred as Mention and it's in addition expressed as "@username". If tweet encloses mention the binary feature representing it'll have worth one, else is 0.

#### Recency

Recency is used at the point when the question is let go to retrieve a tweet, it is smarter to get latest tweet about that issue. Thus, Recency feature finds the age of a tweet in few seconds after its creation.

#### Hash tag

It is a word beginning with # image. It refers to a word concerning the content of text or indicating the subject of tweet. The binary feature value provides the solution, if the tweet contains hash tag or not.

#### Emotions

The Emotions are facial representations pictorially represented utilizing letters and punctuation; they convey the user's temperament.

#### Retweet

A tweet are often solely a comment created by a user or is also a solution to a different tweet Retweets are set apart with either "rt" trailed by '@user id' or "by suggests that of @user id". Retweet is viewed because the feature that has created Twitter, Associate in Nursing another medium of knowledge diffusion and additionally direct communication.

#### Singleton

In the event that a tweet has no answer, at that point the tweet is called as singleton.

## **SENTIMENT ANALYSIS IN TWITTER**

Extracting opinion from the content is referred to as Sentiment Analysis [10]. There are different perspectives, reasons, introduction of emotional extraction as per explanation for the analysis. Location detection, Event detection, and so on assignments should be possible on tweets.

At a particular point this task is practiced on twitter data, the design or framework to try to do sentimental analysis differs as indicated by what kind of result one have to be compelled to accomplish from the tweets. Another imperative issue behind the ever-changing plan of stream of twitter sentiment analysis is utilization of assorted approaches and techniques. Several scientists confirm their own explicit system or stream to try to do sentiment analysis to reinforce the potency of the end result. A number of basic strides in twitter sentiment analysis and also the keywords in it are given below:

### **Preprocessing**

In the introduction of the design of sentiment analysis, we are able to define the topic into the concomitant work method. Therefore the final steps concerned during this structure are as follows:

Prior to the start of the sentiment analysis, data preprocessing should be done.

### **Removal of Non-English Tweets**

The tweets are free from large datasets like Clue web or TREC dataset, it includes English and additionally non-English tweets. During this manner, we'd like to run language identification on every tweet, and wish to delete from our assortment of all tweets that are parceled out a 0-likelihood of being English.

### **Removal of Re-tweets**

We need to erase any content that took once a RT token (and to boot the RT token itself), since such content frequently corresponds to the retweeted material.

### **Conversion to ASCII**

Numerous tweets contain abnormal characters, which can be hazardous for down-stream development. To address these problems, we need to utilize a mix of BeautifulSoup5 and Unidecode6 to change over and reproduce all tweets to ASCII.

### **Removal of Empty Tweets**

After finishing the majority of the other pre-processing, we need to erase any empty tweets.

### **Restoration of Abbreviations**

We can reestablish famous truncations utilized in the tweets, to their comparing unique structures utilizing a lexicon of abbreviations[11]. For example, "wknd" to "week-end"). Punctuations are kept since individuals regularly express feeling with tokens, for example, ":)", ":- )". These feelings can likewise be utilized for sentiment classification.

## **FEATURE SELECTION**

### **Lexicon Features**

we can categorize the words into negative, positive and neutral lexicons based on the originality of the term We have to correlate each word with word net libraries.

### **Part-of-speech Features**

Parts-of speech features i.e. adverbs, nouns, adjectives, etc. are tagged in each tweets.

### **Micro-blogging Features**

we can detect the presence of negative, positive and neutral emotions by creating the binary features. Tweets can categorize in neutral, positive, and negative by the occurrence of abbreviations and intensifiers. Online existing slang dictionaries can be used for emotions and abbreviations.

## **STEPS TO EXTRACT FEATURES**

### **Case Normalization**

Entire document is transfers into lowercase for case normalization.

### **Tokenization**

Tokenization is dividing up the frameworks of text into individual terms or tokens. This strategy will take enormous varieties, in keeping with the term being analyzed [12]. For English, viable tokenization technique is to utilize blank space and accentuation as token delimiters.

### **Stemming**

Stemming is that the technique of decreasing relevant tokens into one variety of token. this technique contains the feeling and elimination of unsuitable pluralizations, suffixes, and prefixes.

### **Steps To Retrieve Tweets**

[13] Retrieve tweets from Twitter use library (twitterR) & library(ROAuth)

## Twitter authentication

Setup consumer\_key k , consumer\_secret s, access\_token akt,access\_secret as)

Set up n value

Get tweets by User timeline

readRDS fn: Load tweets L

### **Text Cleaning**

Load tm lib

build a corpus c, specify the source to be character vectors char c1;

load corpus vector

Map corpus to content\_transformer

Remove URL: remove anything other than English letters or space from L

remove stop words

remove extra whitespace

### **Stemming And Stem Completion**

## refer card data mine r1 to provide link package cran package

## map reduce hadoop ad

Stem Completion x,

Execute lapply fn from corpus data set



It is assessed that more than 70% of possibly usable business data are unstructured, frequently as text data. Text mining gives an accumulation of techniques that enable us to get significant experiences from these data utilizing the word cloud.

## VIII. CONCLUSION

Twitter knowledge is extremely helpful in higher cognitive process as a result of it's providing a range of opinions on numerous topics. The text mining can accomplish on twitter knowledge and that area unit consumed visualizing techniques. it's possible to use text mining tasks and sentiment analysis for twitter knowledge to research user contributed reviews for product or services offered by banking establishments. it'll give a competitive advantage for banking sectors, business retailers and provides services to suppliers to research their end users' views relating to their product or service exploitation social media knowledge. It is generalized that, the Banking establishments will utilize their end users opinions generated from social media pursuit and analysis by adapting their promoting plans, product and business intelligence severally. Moreover, it's conjointly appreciated to use un-supervised techniques in sentiment analysis and opinion mining for enlightening the business economical price and therefore the client relationship management.

## REFERENCES

- [1] Vineet Richhariya, Vivek Richhariya and Shilpy Gandharv, "Twitter Dataset, Text Mining in real time", International Journal of Computer Applications, Volume 178, 2017.
- [2] Abdullah AlBannay, "Twitter Analysis Using Text Mining Tools", The Manchester School of University, Computer Science department, 2014.
- [3] Isti Surjandari, Muthia Szami Naffisah et al., "Twitter Data for Public Sentiment Analysis of Staple using text mining", Foods Price Changes of Industrial and Intelligent Information Vol. 3, 2015.
- [4] Dr. Marco Spruit, Dr. Matthieu Brinkhuis, Text mining in Monetary industry: implementation of text mining methods on bank policies, master's thesis, Utrecht University Utrecht, 2017.
- [5] Salloum, Mostafa-Emran, Khaled Shaalan, et al., "A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives", Advances in Science, Technology and Engineering Systems Journal Volume: 2, 2017.
- [6] R. R. Deshmukh, Kishori K. Pawar, "Twitter Sentiment Classification of Sanders Data uses Hybrid Approach", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 17, 2015.
- [7] German Gemar Jose Antonio Jimenez-Quintero, "Text mining social media for competitive analysis", ISCTE – University Institute of Lisbon, Volume 11, 2015.
- [8] Dr. K. Chitra, B. Subashini "Data Mining Techniques and its Applications in Banking Sector", International Journal of Emerging Technology and Advanced Engineering Certified Journal, Volume 3, 2013.
- [9] Mousa Albashrawi, "Detecting Financial Fraud Using Data Mining Techniques", Journal of Data Science 2016.
- [10] Pukharaj, P Shrishrimal, R.R. Deshmukh, Kiashore K. Pawar, "Twitter Sentimental Analysis: A Review", Engineering Research of International Journal and Scientific, Volume 6, 2015
- [11] Yanchang Zhao, "Twitter Data determination with R Text Mining & Social Network Analysis", University of Canberra, 2016.
- [12] Haldorai, A. Ramu, and S. Murugan, "Social Aware Cognitive Radio Networks," Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010
- [13] Chien-Chung Chan, Kathy J. Liszka, et al., "Sentiment Analysis of Twitter Data a. in Text Mining", The University of Akron, Department of Computer Science.
- [14] Fakeeha Fatimax, Shaeela Ayeshaz, et al., "Text Mining: Applications & Issues and Techniques", International Journal of Advanced Computer Science and Applications (IJACSA), Volume: 7, 2016.

- [15] Prof. Dr. Paulo Rita, Dr. Sergio Moro (2015), "Sentiment Analysis In Hospitality Using Text Mining: The Case Of A Portuguese Eco-Hotel" ISTE Business school.
- [16] Hejr, H., Paymard, A., Sataie, J., Roozitalab, M. A three-year study of causes of mortality from poisoning in referral bodies to the yasuj forensic medicine faculty in 2012-2014(2018) International Journal of Pharmaceutical Research, 10 (1), pp. 299-304. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059620565&partnerID=40&md5=c04207e737acd8a2ade4d0815487aa21>