A Text Mining Application Of Emotion Classification Of Twitter's Users

R. SonyKrishna[1], L. Priyanka[2] , K. VijayaLakshmi[3] , M. Sowmya[4] .

Assistant Professor[1] ,CSE,Andhra Loyola Institute of Engineering And Technology,Vijayawada.

Final Year B.Tech [2,3,4],CSE, Andhra Loyola Institute of Engineering And Technology,Vijayawada.

ABSTRACT—*Application of text mining procedures on social networking sites can extra reveal results regarding human pondering patterns group identification and suggestion and likewise opinion about any specified issues of pursuits. Many researches labeled emotions into 2 categories that positive and negative. Consequently this excludes twitter's users different general feelings like anger, sad, disguist and surprise. It is noticed that those present strategies of emotion prediction in text do not mainly target the datasets in social networking websites and they rarely point out the preprocessing segment that's primary to simplify the text mining procedure. In this paper,the mentioned dis-advantages are being resolved.This is mainly focusing on six different emotions,they are happy, sad, surprise, disguist,angry and fear. Classification is mainly done in three phases (text collection, preprocessing and processing). Steps performed in the preprocessing section had been case folding, cleaning, discontinue-word elimination, emoticons conversion, negation conversion, and tokenization. Within the processing phase, it carried out weighting and classification utilising the Naive Bayes algorithm.The datasets size is also increased and the prediction of emotions are more accurate.*

Key words—Text mining, Emotion detection, Naïve Bayes method, Preprocessing, Processing, Datasets.

I. INTRODUCTION

Social networking web pages like twitter and facebook create enormous possibilities for users to be in contact with one yet another while not having to worry about variations in moral and social values. Additionally they allow mutual studying and sharing of useful competencies with no regard to geographical distance time barrier and language talents. Users thus become a member and engage in more than a few communities and discussions corporations that great suit their wants.

Twitter has atleast millions users and tweets posted in its web page everyday.Tweets are written messages in the form of texts that have many opinions, expressions and feelings of users.Information in Twitter's web site is unstructured because users don't care about Spelling and grammatical mistakes when they are posting their tweets.This is difficult to identify the emotions from the unstructured data. The every tweet posted by user may contain maximum of 140 letter [1] These tweets have many hidden emotions.A written tweets have more than one emotion or may not have any emotion.In this paper, the tweets are collected and some methods are applied to classify the emotions.

a)Text Mining:

Text mining, additionally known as textual content data mining, roughly similar to text analytics, is the approach of deriving important and uesful data from large data. High quality data is most commonly derived through the devising of pattern corresponding to statistical pattern studying. Text mining more commonly entails the method of structuring the enter text, deriving patterns within the structured information, and ultimately analysis and interpretation of the output. High quality in text mining refers to a combinations of relevance, novelty, and interestingness[3]. Common text mining have methods like textual content categorization, textual content clustering, thought/entity extraction, construction of granular taxonomies, sentiment evaluation, report summarization, and entity relation modeling.
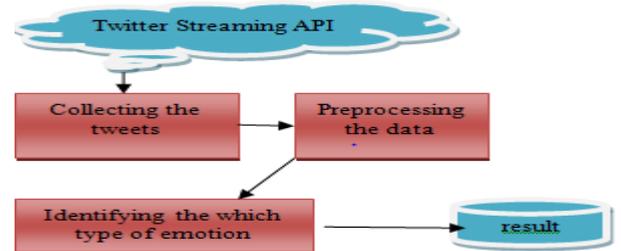

Figure1: Text Mining

II. PROPOSE SYSTEM

Text mining application is used to discover twitter users feelings. Software of text mining system on social networking web sites can extra reveal the human pondering patterns.Text mining is used to overcome this suitation as its supplies computational intelligence. A text mining software of emotion classification of twitter's

users which will predict the twitter users emotions involves six types of feelings. Emotions of Twitter's users which can be classified into six emotions, particularly happy, sad, anger, disgust, fear, and surprise[11][13].

Three important phases of the text mining utilized on this application had been text collection, preprocessing, and processing. Activities carried out within the preprocessing section had been case folding, cleaning, stop-word removal, emoticons conversion, negation conversion, and tokenization to the learning information and the test data established on the sentiment analysis that carried out morphological evaluation to construct a number of models. Within the processing section, it performed weighting and classification utilizing the Naive Bayes algorithm.Text mining application makes use of Naive Bayesian methods which is used to foretell the twitter person feelings[9].

It could actually extract the data from twitter site which is unstructured, tremendous and dynamic.To organize the accrued knowledge into pre-outlined categories that can be used for performing text analysis by way of preprocessing. Assemble the compatible units centered on the information set through processing. Then validate the emotions of tweets within the information set.

III. EMOTIONS

The categorization of emotions has often been studied from two principal techniques: basic emotions and core influence.

1. Basic Emotions:

Basic emotion theorists think that people have a small set of normal feelings, that are discrete [20]. More than a few researchers have attempted to establish a number of general emotions which are universal amongst all people and vary one from an additional in important ways. A trendy example is a go-cultural study of 1972 by means of Paul Ekman and his colleagues, where they concluded that the six common emotions are anger, disgust, fear, happy, sad, and surprise[17].

2. Core Affect Model:

Core influence model of emotion characterizes human feelings by defining their positions along two or three dimensions. That's, most dimensional units incorporate valence and arousal dimensions[20].

a)Emotion Analysis in Text:

Effort for emotion evaluation on Twitter knowledge entire by Bollen and his colleagues . They tried to find a relationship between overall public mood and social, fiscal and other principal pursuits[14]. They extracted six dimensions of mood (anxiety, depression, anger, vigor, fa- tigue, confusion) utilizing an multiplied variant of POMS (pro- file of temper States), a psychometric instrument. They located that social, political, cultural and fiscal pursuits have a enormous, and immediate outcomes on the various dimensions of public mood[15].

IV. STOPWORDS

Discontinue phrases are on the whole probably the most ordinary phrases together with articles (a, an, the), auxiliary verbs (be, am, is, are), prepositions (in, on, of,

at), conjunctions (and, or, nor, when, even as), and it record together with bad verbs (now not, is just not, does now not, don't, must now not, and many others.), auxiliary verbs (be, am, is, are), prepositions (in, on, of, at).In addition, we changed the phrase "very" with blank and the word "clean no longer clean" is replaced via "clean not". That do not provide additional growth for engines like google however broaden the computational complexity by means of growing the size of the dictionary[16].

Example:

For instance, "i'm happy." => "i'm happy." => "i'm happy."
"i'm not very happy." => "i'm not happy." => "i'mNOThappy."

In this example the phrase "happy" and "not happy" is used to create new words "happy" and "NOThappy". On this means, we are able to discriminate the phrase "happy" having positive meaning and it is classified as the word belongs to happy class. In the identical means, the new phrase "NOThappy" has a negative meaning and it is classified as the word belongs to sad class.

V. ARCHITECTURE

The architecture of proposed system is shown in figure 2. This paper makes use of machine learning based methodologies for predicting the emotins of twitter's users.Tweets are collected from the Twitter API. The words had been extracted and saved in a feature vector.The text is collected from the tweets and divided into 2 sets.They are training set and testing set.From these sets the feature are extracted from each and every word respectively. From the sets the features are extractes and stored.And also naïve bayes methods are being implemented[5].

Naive Bayes does not recollect the relationships between aspects equivalent to emotional keyword phrases and emoticons. This is best for sentiment analysis as regularly these facets do not invariably relate to one an additional corresponding to in the use of a smiley emoticon on the finish of a negative tweet[20].The features extracted are based on the techniques applied in the preprocessing phase.The Naive Bayes methods will classify the emotions present in the tweets.

Both training and testing sets are important for the classification of emotions.The classified emotions are then label whether it belongs to happy or sad or anger or disguist or fear or suprise.
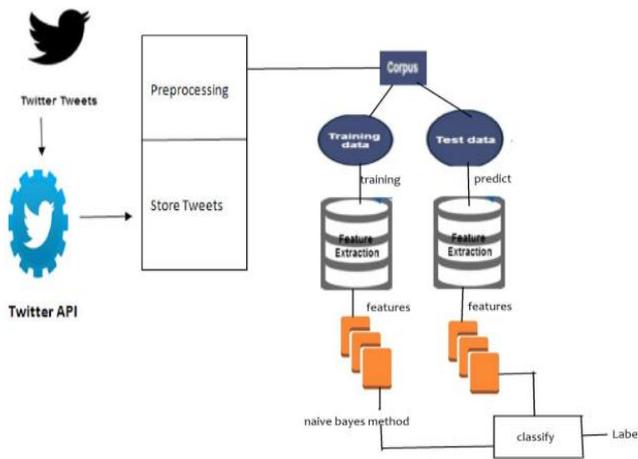
Figure 2: Architecture

VI. PHASES IN THE PROPOSED SYSTEM

Text mining Application makes use of three main phases which is used to classify the emotions. The Phases are:

• Text Collection
• Preprocessing
• Processing

(a) Text Collection

Text collected is completed using the streaming API. Twitter Search with extra filters based on username and key words. We colleting the tweets from twitter_site either by means of utilising the username [1].And then performing the preprocessing segment and after elimination of all stop phrases we colleted the preprocessed tweets again.These tweets are being used and in addition some systems like naïve bayes approach is implemented.And sooner or later emotions are categorized[6][7].

(b) Preprocessing

In this preprocessing phase,the techniques used are case folding, stop word elimination, emoticons conversion, tokenization, Conversion to slash case, Removing URL, Eliminating mention from the tweets, Delete a character other than a to z, etc.And there some alogrithm for these above mentioned techniques such as porter stemmer algorithm,etc[3].

Figure 3: Tweet collection and Preprocessed tweets

(c) Processing

For classifying the tweets in this paper we're using Naive Bayes algorithm in preprocessing section. Naive Bayes classification on each tweet represented in a pair of attributes.They are training set and testing set.From these sets the feature are extracted from each and every word respectively. From the sets the features are extractes and stored.And also naïve bayes methods are being implemented.This algorithm is for sentiment analysis as commonly these facets don't invariably relate to at least one an extra similar to in using a smiley emoticon on the finish of a poor tweet. Naive Bayes classifier for classification which is a machine learning algorithm is then utilized to the model classifier and a label is produced [1][2]. The Naïve Bayes procedure for classification is in most cases utilized in text classification as a result of its pace and ease. It makes the belief that phraseare generated independently of word position. The classifier then returns the category with the best chance giventherecord.

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Figure 4: Naive Bayes method

VII. GENERAL APPROACH TO NAÏVE BAYES METHOD

In the textual content classification literature, two specific units established on the naïve Bayes assumption have been proposed: the multinomial mannequin and the multivariate Bernoulli mannequin . In this paper we have applied the multinomial model. The multinomial model specifies that a document be represented through the frequencies of time period within the record by means of representing a record with the bag of words[2][4].

The records in every category can then be modeled as samples drawn from a multinomial word distribution. Thus, the conditional likelihood of a file given a class is effortlessly a manufactured from the chance of every found phrase within the corresponding class. Fig1 shows the final process for constructing a Naïve Bayes Classification algorithm[8].
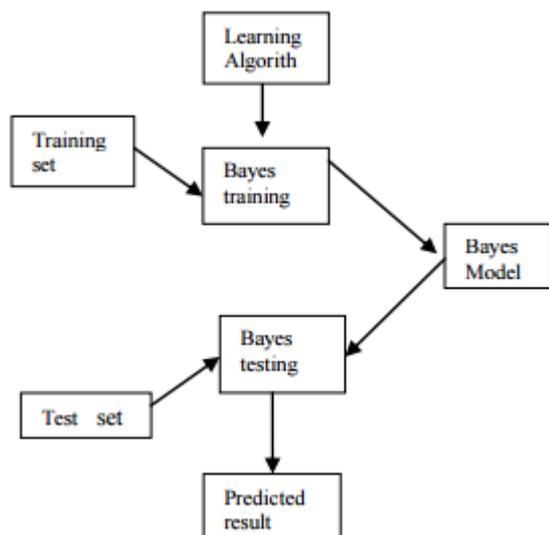
Figure 5: General Approach to naïve bayes method

mining software to discover emotions of Twitter customers which can be classified into six emotions, specifically happiness, sad, anger, disgust, fear, and shock. This paper can get 75 % accuracy which used to be observed on the one hundred fifty tweets.

X. FUTURE ENHANCEMENT

There are several possible instructional materials for future study on this field. Essentially the most promising one we believe is a model where in more security is furnished to the end users information and the passwords are saved in encrypted format. And this can be applied to other sectors like agriculture based,etc.And likewise dataset dimension should be elevated and by using increasing the dimensions, the accuracy of the prediction additionally will also be increased[18][19].And in addition voice based software can also be completed with the aid of mining inspiration.Other types of algorithm and techniques can implemented in the text mining phases.

VIII. RESULTS

This is the result obtained after preprocessing and processing phases are applied and in figure 5 the actual tweet is " I am proud of you. I was in a state of joy and shock with your results". After preprocessing phase is completed the tweet contain only proud, joy, shock, results.Where tweet consists of two happy words and one fear word,the tweet is considered as "happy".
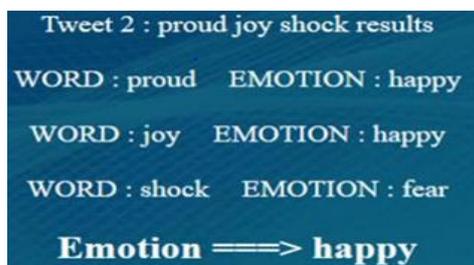

Figure 6: Emotion predicted in the tweet

IX. CONCLUSION

Digital textual documents are particularly got from the social websites. Huge numbers of technologies are developed for the extraction of meaningful knowledge from enormous collections of textual data utilising extraordinary textual content mining tactics[12]. However, textual content pre-processing becomes more challenging when the textual understanding shouldn't be structured in line with the grammatical conference. This overview presents a thorough figuring out of specific textual content classifiers in the social networking internet sites. From our evaluate we concluded that specific algorithms perform differently depending on information collections [2][16]. The text

REFERRANCES:

[1].D.Farber"Twitter Hits 400 Million Tweets Per day,Mostly Mobile,"06th june2012.[online].Available:www.c net.com/news/twitter-hits-400-million-tweets-per-day-mostly mobile/.[Accessed 10th September 2015].

[2]. VandanaKorde, C NamrataMahender "TEXT CLASSIFICATION AND CLASSIFIERS: A SURVEY" International Journal of Artificial Intelligence &Applications (IJAIA), Vol.3, No.2, March 2012.

[3]. RizwanaIrfan, Christine K. King, Daniel Grages, Sam Ewen, Samee U. Khan,Sajjada. Madani, Joanna Kolodziej, Lizhe Wang, Dan Chen, Amma R Rayes,NikolaosTziritas, Cheng - ZhongXu, Albert Y. Zomaya, Ahmed SaeedAlzahrani,AndHongxiang Li "A Survey on Text Mining in Social Networks, " The Knowledge Engineering Review, United Kingdom, (2004) pp.1-24.

[4]. Susan Dumais John Platt David Heckerman, "Inductive Learning Algorithms and Representations for Text Categorization", Published by ACM, 1998.

[5]. MoromiGogoi, Shikhar Kumar Sarma"Document Classification of Assamese Text Using Naïve Bayes Approach", IJCTT) – volume 30 Number 4 – December 2015.

[6]. Ananthi.J "A Survey Web Content Mining Methods and Applications for Information Extraction from Online Shopping Sites", International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, pp. 4091-4094.

[7]. Xin Chen, MihaelaVorvoreanu, and Krishna Madhavan "Mining Social Media Data for Understanding Students' Learning Experiences," IEEE transactions on learning technologies, manuscript id 1, (2013), pp. 1-14.

[8]. Ms.S.Valarmathi, Mr.P.Purusothaman "A Survey on Web Content Mining Techniques and Tools", IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014.

[9]. "#Emotional Tweets",Saif M. Mohammad, June 7-8, 2012.

[10]."EMOTEX: Detecting Emotions in Twitter Messages", Maryam Hasan, ElkeRundensteiner, Emmanuel Agu.

[11]. Shahheidari, S.; Hai Dong; Bin Daud, M.N.R., "Twitter Sentiment Mining: AMulti Domain Analysis," Complex, Intelligent, and Software Intensive Systems(CISIS), pp.144-149, 3-5 July 2013.

[12]. Khan, K., B. Baharudin, and A.Khan. Mining opinion from text documents:A survey. In Digital Ecosystems and Technologies, 3rd IEEE InternationalConference on. 2009: IEEE: pp. 217–222.

[13]. Chaumartin, F., A knowledge-based system for headline.

[14]. EMOTEX: Detecting Emotions in Twitter Messages :Paul Ekman, "Basic emotions," Handbook of cognition and emotion, vol. 98, pp. 45–60, 1999.

[15]. EMOTEX: Detecting Emotions in Twitter Messages :[1]Johan Bollen, Huina Mao, and Alberto Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena.," in International AAAI Conference on Weblogs and Social Media (ICWSM'11),

[16]. E. Boiy, P. Hens, K. Deschacht, and M.-F. Moens, (2007) Automatic Sentiment Analysis in On-line Text, ELPUB2007. Openness in Digital Publishing: Awareness, Discovery and Access In : Proceedings of the 11th International Conference on Electronic Publishing,Vienna, Austria 13-15, June (2007). 011.

[17].Paul Ekman, "Basic emotions,"Handbook of cognition and emotion vol. 98, pp. 45–60, 1999.

[18]. Learning to Classify Emotional Content in Crisis-Related Tweets: [15] K. Glass and R. Colbaugh,

"Estimating the sentiment of social media content for security informatics applications," Security Informatics, vol. 1, no. 3, 2012.

[19] SA-E: Sentiment Analysis for Education :[30] Bhargavi, P., and S. Jyothi. "Applying Naive Bayes data mining technique for classification of agricul-tural land soils." International Journal of Computer Science and Network Security 9.8 (2009): 117-122.

[20]SA-E: Sentiment Analysis for Education: [17] Prabowo, Rudy, and Mike Thelwall. "Sentiment analysis: A combined approach." Journal of Informet-rics 3.2 (2009): 143-157.

AUTHORS PROFILE:

Ms. R. Sony Krishna ReddyM.Tech degree in computer Science and Technology from V R Siddartha Engineer college. Working as assistant professor in department of CSE in at Andhra Loyola Institute of Engineering and Technology (ALIET). She has published many papers in the international journals and international conferences. Her area of interest includes data mining, computer networks.

Ms. L. Priyanka currently pursuing her B. Tech degree in Computer Science and Engineering at Andhra Loyola Institute of Engineering and Technology (ALIET) at Vijayawada, India,2017.Her research includes data mining.

Ms.K.VijayaLakshmi currently pursuing her B. Tech degree in Computer Science and Engineering at Andhra Loyola Institute of Engineering and Technology (ALIET) at Vijayawada, India,2017.Her research includes data mining.

Ms.M.Sowmya currently pursuing her B. Tech degree in Computer Science and Engineering at Andhra Loyola Institute of Engineering and Technology (ALIET) at Vijayawada, India,2017.Her research includes data mining.