Proceedings of 1st Shri Chhatrapati Shivaji Maharaj QIP Conference on Engineering Innovations
Organized by Shri. Chhatrapati Shivaji Maharaj College of Engineering, Nepti, Ahmednagar
In Association with JournalNX - A Multidisciplinary Peer Reviewed Journal, ISSN No: 2581-4230
21st - 22nd February, 2018

USE OF SOCIAL POSTS FOR DISASTER DETECTION USING NATURAL LANGUAGE PROCESSING

Prof. Anuja Phapale(Author)
Dept. of Information Technology
AISSM's Institute of Information Technology Pune, India

Onkar Dalvi(Author)
Dept. of Information Technology
AISSM's Institute of Information Technology Pune, India

Suyash Patekar(Author)
Dept. of Information Technology
AISSM's Institute of Information Technology Pune, India

Shubham Thombare(Author)
Dept. of Information Technology
AISSM's Institute of Information Technology Pune, India

Rahul Anantulwar (Author)
Dept. of Information Technology
AISSM's Institute of Information Technology Pune, India

Abstract—: A person's attitude is reflected using his behavior. If we want to know a person's behavior, we can ask to his friend about him. With the growing importance of social media, researchers made social media as a business review machine. Using sentiment analysis on reviews, product's market value, lifespan of product etc. can be predicted. Social media can also be helpful to get reaction of public on some social issue. This will help to politicians for analyzing the impact of social issue on public's mood. Sentiment analysis of reviews from different social media such as short texts are insufficient for analysis. The main idea behind the proposed system is to make use of social media which is immensely active i.e. Facebook and use the posts which are posted. Using sentiment analysis and Natural Language Processing (NLP) on posts, disasters are extracted (riots, accidents, traffic issue, natural calamities etc.) and using Naïve Bays classification technique disasters are classified. The challenges are processing of unstructured data and finding the annotated data. In this paper we find the solution for above challenges which will be beneficial for our system and provide solution to handle unstructured data easily**.**

Keywords— NLP (Natural Language Processing), Sentimental Analysis, Data Mining, Naïve Bayes Classifier

I. INTRODUCTION

Sentiment analysis is all about finding the emotions, moods, opinions etc. from the training data. Sentiment analysis can be seen as a Natural Language Processing (NLP) task. To understand the emotions from text NLP is needed. The challenge is to find out sentiments from unstructured data, this can be easily done with the help of NLP. Basic task of researchers is to make polar classification. Sentiments can be classified into Positive, Negative and Neutral classes. For example, if a review contains words like good, awesome, excellent etc. about a product then it will be classified into a positive class. Or if words like bad, disgusting, worst etc. appears in the review then it will be classified into negative class. If review contains words like quite good, not bad, etc. then it will be classified into neutral class Sentiment analysis is best platform for researchers to focus on commercial domains. Movie review can be done with help of posts. Review technique is also quite helpful for visitors. Restaurant review, hotel review, city review, product review etc. these are various commercial domains, where sentiment analysis can do the magic for researchers. The basic task is to extract the data from short text posts using NLP and finding the emotions from it. Sometimes unstructured data is difficult to handle when user write something good about product and at same instance he does the bad review rating. This challenges have to overcome, which can be done with using pre-processing and NLP. Many research has done work on sentiments that we are using in day to day life. Weather forecasting, Cricket match score prediction etc. are the few examples of it. We can do more with it. We can train a laptop device with laptop owner's activities. Like, at what time he often uses laptop or for what job on laptop he likes to hear music for example: while doing power point presentation. According to situation laptop will perform differently. Similarly, we can give our data to coffee machine and according to our emotion it will make coffee. Sometimes it may get wrong but most probably it will be surprised for owner. Sentiment analysis has many more exploratory applications, above is just an example.

II. LITERATURE SURVEY)

Hase Sudeep Kisan, et al. 2016 [1] Cloud platform is used to incorporate the sentimental analysis process of social media data using the NLP libraries. The analysis is evaluated and put into effect to handle recent affairs and activities in the world. Cloud implementation proves to be a boon in terms of improvements, efficiency and growth. Hashtags are used to express a certain feeling on twitter. Hashtags of various kinds forms group of various tweets and can be searched using the hashtags and then the calculation of sentiment is done. REST API's can be used in order to control the twitter data. It uses OAuth for identifying authorised users. OAuth is an open standard for authorization. Using OAuth, the user need not share his passwords for logging into third party websites, the user can log in using his/her social media account. The library used by twitter API is called as the twitter4j. NLP tools are provided by Stanford core NLP for sentimental analysis.

Wei Yen Chong, et al. 2014 [2] Structure of tweets is different than normal text. The major difference between tweets and normal texts are structure of sentence, length of sentence and data available. Tweets can be categorized in three sections: positive, negative and neutral. Accuracy and precision are used to evaluate performance. Tweets are converted into a language a machine can read using Txt Preprocessing. Various machine learning algorithms can be used for classification, for example, Naive Bayes, Decision tree and support vector machine.

Deebha Mumtaz, et al. 2016 [3] The main aim of this research paper is to perform sentiment analysis on movie review data. To find the polarity of a review as positive, negative or neutral, Senti-lexical algorithm is proposed. A method to handle words which have negative effect have been proposed and use of emoticons is also discussed. Bo Pang et al has used movie review to train algorithms that finds the sentiment in it. Movie reviews are best example for analysis as they mention clear opinion.

Paolo Nesi, et al 2015 [4] This paper presents a distributed framework for crawling web documents and running Natural Language Processing tasks in a parallel fashion. The system is based on the Apache Hadoop ecosystem and its parallel programming paradigm, called MapReduce. In the specific, we implemented a MapReduce adaptation of a GATE application and framework (a widely used open source tool for text engineering and NLP). A validation is also offered in using the solution for extracting keywords and key phrase from web documents in a multi-nod Hadoop cluster. Evaluation of performance scalability has been conducted against a real corpus of web pages and documents.

III. PROBLEM STATEMENT

Study Facebook posts and implement Sentiment Analysis using Natural Language Processing on it and classify them as natural calamities, accident, traffic issues, riots etc. According to classification and post verification the appropriate notification alert should be send to the domain users.

IV. OBJECTIVES

The distinct objectives of this system are to ensure:

- The Facebook comment posted by the user is preprocessed using Sentimental Analysis and Natural Language Processing.
- Using Sentimental analysis, the pre-processed comments is classified into different classes.
- The classification of the pre-processed comments is verified using Natural Language Processing (NLP). Appropriate notifications alert according to corresponding processed comments is sent to the domain users.

PROPOSED SYSTEM

In proposed system we are developing dual sentiment analysis system. In this proposed system we take two users in this system admin and multiple users. Admin can add the users and remove the users. User gives comment on particular disaster. Comment is in composite form; system first will split it into words and apply the sentiment analysis using Natural Language Processing. After applying NLP on comment text extract the opinion and targets from comment using the Stanford NLP and give it to the sentiment analysis.

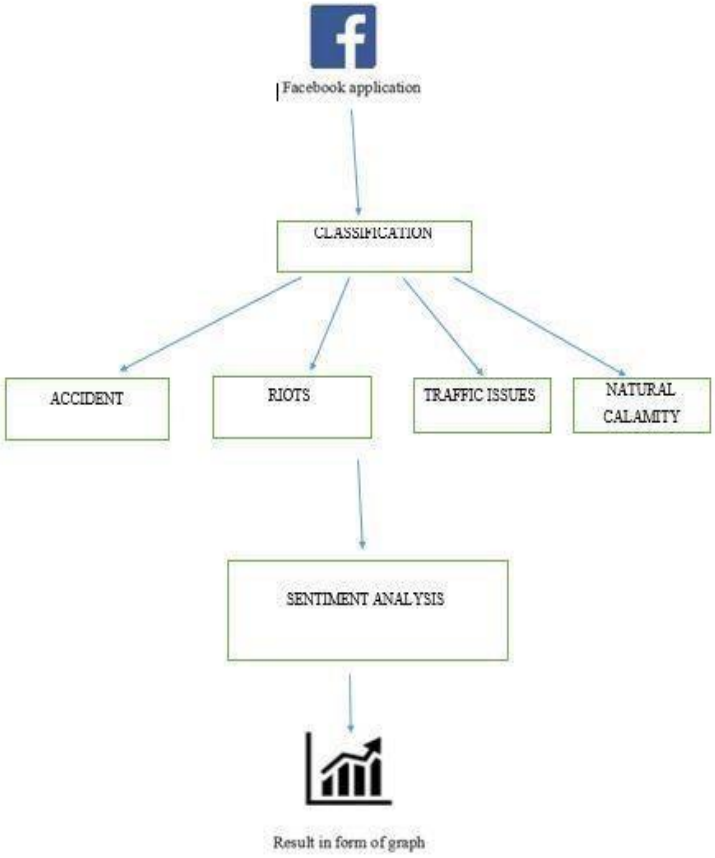

Fig.1: System Analysis Proposed Architecture

V. SCOPE

Following are some highlighted scope:

- Fetching Facebook post using developers account
- Process the post using NLP stages Tokenization, Remove Stop words, Stemming.
- Categorizations of post based on Traffic Issues, Riots, Accidents, Natural Calamity and perform sentiment Analysis.
- To find the categories of the post API will be use and for the sentiment analysis stand ford core NLP Jar will be use.

Categorization of sentiment as Positive, Negative, Neutral Graphical view of analysis.

ALGORITHM USED

NAÏVE BAYES ALGORITHM

A. A. Naïve Bayes Classifier

In machine learning, naive Bayesian classifiers are a family of simple probability classifiers based on Bayes theorem with strong (Naive Bayes is a simple method for constructing classifiers: models that assign class labels to instances of a problem represented as vectors of characteristic values where class labels are taken from of a finite set. This is not one algorithm for teaching such

classifiers, but a family of algorithms based on a general principle: all naive classifications Bayes' s assume that the value of a particular function does not depend on the value of any other function, given the class variables. For example, a fruit can be considered an apple if it is red, round and about 10 cm in diameter. The naive Bayesian classifier treats each of these functions independently of each other so that the probability that this fruit is an apple, regardless of the possible correlations between color, roundness and diameter characteristics.

For some types of probabilistic models, naive Bayesian classifiers can be trained very effectively in a controlled learning system. In many practical applications, the parameter estimation for naive Bayesian models uses the maximum likelihood method; in other words, it is possible to work with a naive Bayesian model without taking Bayesian probability or using any Bayesian methods & independent assumptions between functions.

Naive Bayes is a sort of classifier that uses the Bayes theorem. It predicts membership probabilities for each class, such as the probability that a given record or data point belongs to a particular class. A class with the highest probability is considered the most likely class. This is also known as Maximum A Posteriori (MAP).

The MAP for a hypothesis is:

MAP(H)

=max(P(H|E))

=max((P(E|H)*P(H))/P(E))

= max(P(E|H)*P(H))

P(E) is evidence probability, and it is used to normalize the result. It remains same so, removing it won't affect.

Naive Bayes classifier assumes that all the features are unrelated to each other. Presence or absence of a feature does not influence the presence or absence of any other feature.

Naive Bayes is a conditional probability model: given a problem instance to be classified, represented by a vector

$$\mathbf{x} = (x_1, \ldots, x_n)$$

representing some n features it assigns to this instance probability.

$$p(C_k \mid x_1, \ldots, x_n)$$

The problem with the above formulation is that if the number of features n is large or if a feature can take on a large number of values, then basing such a model on probability tables is infeasible.

We therefore reformulate the model to make it more tractable. Using Bayes' theorem, the conditional probability can be decomposed as

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

In plain English, using probability terminology, the above equation can be written as

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

B. B. Advantages

- The Naive Bayes algorithm is a fast, highly scalable algorithm.
- Naive Bayes can be used to classify Binary and Multiclass. It provides various types of Naive Bayes algorithms, such as GaussianNB, MultinomialNB, BernoulliNB.
- This is a simple algorithm that depends on the execution of multiple samples.
- Excellent choice for problems with text classification. This is a popular choice for the classification of spam email.
- It can easily be trained on a small set of data

VI. CONCLUSION

In Proposed system, domain users will have awareness about the disaster which has recently occurred. This system converts the comments to notifications. This system works as a social cause system with the help of social networking platforms (Facebook). Accessing data from Facebook and making the world aware about disasters, political activities etc. can be the future scope. Text notification can be the challenge for the existing system, developing own notification system will improve the efficiency of the proposed system.

REFERENCES

[1] Hase Sudeep Kisan , Hase Anand Kisan and Aher Priyanka Suresh,"Collective Intelligence and Sentimental Analysis of Twitter data by Using StandFordNLP Libraries with SaaS", IEEE International Conference on Computational Intelligence and Computing Research, 2016

[2] Wei Yen Chong, Bhawani Selvaretnam, Lay-Ki Soon" Natural Language Processing for Sentiment Analysis", 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, 2014

[3] Deebha Mumtaz; Bindiya Ahuja "Sentiment analysis of movie [6] review data using Senti-lexicon algorithm" 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016 [7]

[4] Paolo Nesi, Gianni Pantaleo, Gianmarco Sanesi"A Hadoop Based Platform for Natural Language Processing(NLP) of Web Pages and Documents" Journal of Visual Language and Computing, 2015.

[5] Martin Sarnovsky; Peter Butka; Andrea Huzvarova "Twitter data analysis and visualizations using the R language on top of the Hadoop platform" 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI) .

[6] Pankaj Kumar; Kashika Manocha; Harshita Gupta "Enterprise Analysis Through Opinion Mining" 2016

International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) Grace Park, Lawrence Chung, Latifur Khan "A Modeling Framework for Business Process Reengineering Using Big Data Analytics and A Goal-Orientation" 2017 11th International Conference on Research Challenges in Information Science (RCIS)

[7] Sudipto Shankar Dasgupta; Swaminathan Natarajan; Kiran Kumar Kaipa; Sujay Kumar Bhattacherjee; Arun Viswanathan "Sentiment Analysis of Facebook Data using Hadoop based Open Source Technologies" 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)