IMPROVING PERFORMANCE OF TWITTER BASED ON SENTIMENT ANALYSIS

Nikhil Kalraiya
Infinity Management & Engineering College Sagar (M.P.)
RGPV,Bhopal
Sagar (M.P.), India
kalraiya.nikhil@gmail.com

Asst.Prof Sarvesh Rai
Infinity Management & Engineering College
RGPV, Bhopal
Sagar (M.P.), India
sarvesh.s51@gmail.com

Abstract

Social networking websites, application such as Twitter and Facebook, Instagram are important spaces for discussion regarding anything like product, events, election etc. Now a day the channels for shows opinions seem to increase daily. When these opinions are applicable to a company, These opinions are important sources of business insight, whether they represent critical intelligence about a customer's defection risk, the impact of an strong reviewer on other people's obtain decisions, or early feedback on product releases, company news or competitors. The importance of controlling the opinion is growing as Customer use technologies such as Twitter to express their views directly to other customer. This was the main encouragement behind this work. It is decided to develop a system that cans analyses about Demonetisation in India 2016.

Keywords: Sentiment Analysis, Opinion Mining, Natural Language Processing

(NLP), Twitter, support vector machine (SVM).

1. Introduction

Today Twitter has very popular communication way among micro bloggers and Internet users to share opinions on different aspects of daily life. Twitter communication generated content such as consumer satisfaction, opinion extraction, ratings and sentiment analysis. Use of sentiment analysis techniques such as natural language processing, computational linguistics, and fundamental text analysis. Dataset use to form collected messages from Twitter. Twitter contains huge amount of short and unstructured messages created by the users of this micro blogging sites. Sentiment analysis is a type of text classification that classifies texts based on the sentimental orientation of opinions. Sentiment Analysis is a big challenge

because ample of the sentiment that is available on the internet is in highly unstructured format. It is tedious for computers to process it and extract significant information from it. This work provides information about public's sentiment towards demonetization in India.
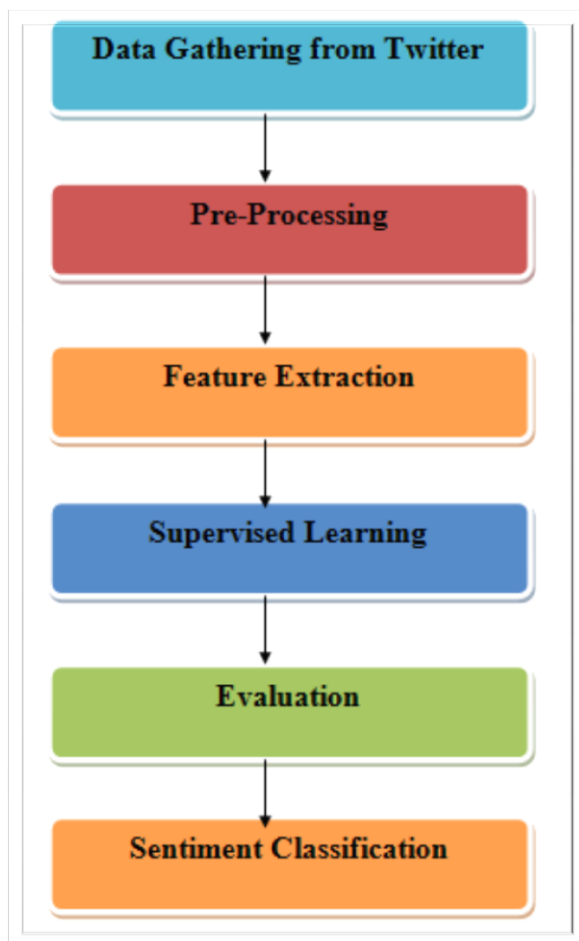
2. Proposed Methodology

Fig: 1. Flowchart of Proposed approach for sentiment analysis

Above figure.1 shows the flowchart of implementation, showing the steps needed to

be performed from collecting raw data for classification, using the proposed methodology. In the machine learning approach 'Extremely Randomized Tree' include only supervised learning to classify the sentiment. After collecting the data then preprocessing it. A list of words had been selected as features of a text and removing a number of words that do not contribute to the text's sentiment. We use an N-grams feature for opinion classification of Demonetization tweets. If N=1, 2, 3 or 4 is defined as unigram, bigram, trigram and tetra gram respectively. Finally, result is analyzed and compared with other classification algorithms.

Preprocessing

Tweet data is:

1. Unstructured
2. Misspelling,non-traditional grammar
3. Change corpus to lower-case
4. Change corpus to Plain Text
5. Eliminate Punctuations

The data cleaning procedure was done to eliminate any irrelevant content from the training data and the input tweets. The term irrelevant content is used to detail any piece of information within the tweet that will not be valuable for the machine learning

algorithm to allot a class to that tweet. Data cleaning process not only simplified the classification task for the machine learning techniques but in training phase it also provided to greatly reduce processing cost.

R' Script Used For Sentiment

'R' Package	Function
Tm	Load R package for text mining and load the texts into 'R'
SnowballC	A 'R' interface to the C libstemmer library that implements Porter's word stemming algorithm for collapsing words
CaTools	For splitting the dataset into testing and training sets
Word cloud	It allows us to highlight the most frequently used keywords
Rpart	It implements recursive partitioning

Classification

Table 1: 'R' Packages Used in this project

Writing Twitter client in R is an easy programming task. Required is to

import the TwitteR package
require(twitteR);
require(plyr);
require(stringr);
tweets=searchTwitter(Subj,
n=NumOfTweets, lang='en');
#transform tweets into pure text
tweets=laply(tweets,function(t) t$getText());
write(tweets, file=outFile);

Extremely Randomized Tree

The **sklearn.ensemble** module includes two balancing algorithms based on randomized decision trees: the Random Forest algorithm and the Extra-Trees method. Both algorithms are perturb-and-combine techniques specifically designed for trees. This means a diverse set of classifiers is created by introducing randomness in the classifier construction. The prediction of the ensemble is given as the averaged prediction of the individual classifiers. A Decision Tree obtains as input a situation or an object or described by a set of properties and outputs a YES/NO "decision".
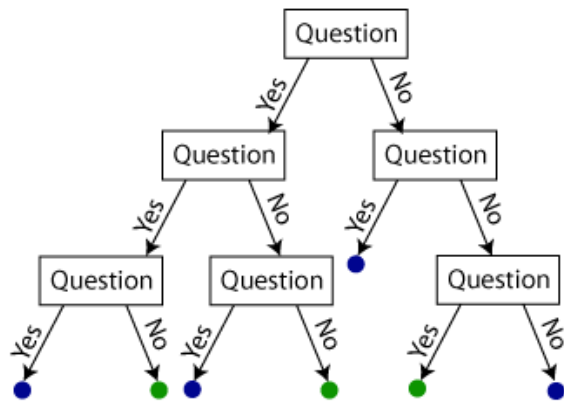
Fig: 2 Process of Random Tree

Random Forest chooses best from random subset of features as the splitting rule. This mostly permits to reduce the variance of the model a bit more, at the cost of a somewhat greater increase. This approach necessarily consists of randomizing strongly both cut-point choice and attribute while splitting a tree node.Apart from accuracy the important factor of the proposed technique is computational competence. The execution time of Extra trees is faster than Random Forest.

Random Forest algorithm can be used for both classification and regression applications. Random Forest algorithm is built in **randomForest** package of R and same name function allows us to use the Random Forest in R.
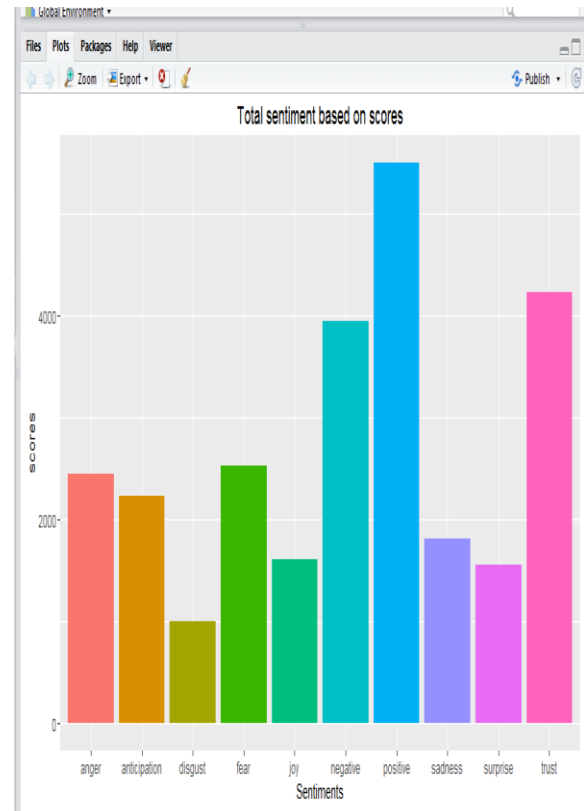
Graph: 1. Shows graph of Sentiment Analysis towards Demonetization in India 2016

Total sentiment analysis on 'Demonetization in India 2016' graph: 1 show above in respect of anger, anticipation, digust,far,joy , negative,positive,sadness,surprise,trust.

According to this graph of sentiment we conclude that views towards Demonetization are positive.

3. Conclusion and Future work

Sentiment Analysis provides huge number of opportunities by uncovering the opinions

and views from the unstructured Twitter data set. To conclude, this research has explained that an efficient sentiment analysis can be performed on an event, Demonetization in India 2016. Throughout the continuation of this research various data analysis tools were applied to gather, clean, mine and determine review from the dataset. This analysis can help them to spot a positive turn in viewer's opinion of their brand image. Uncovering positive trends early on can permit them to make educated decisions.

REFERENCES

[1]Sanket Sahu, Suraj Kumar Rout, Debasmit Mohanty "Twitter Sentiment Analysis" 2015 IEEE International Symposium on Nanoelectronic and Information Systems.

[2]Walaa Medhat , Ahmed Hassan , Hoda Korashy "Sentiment analysis algorithms and applications:A survey"Ain Shams Engineering Journal (2014) 5, 1093–1113 2090-4479 _ 2014 Production and hosting by Elsevier B.V. on behalf of Ain Shams University.
http://dx.doi.org/10.1016/j.asej.2014.04.011.

[3] Dr Nirmala C R, Roopa G M, Naveen Kumar K R "Twitter Data Analysis for Unemployment Crisis"978-1-4673-9223-5/15/$31.00_c 2015 IEEE

[4] Lorenzo Gatti, Marco Guerini, and Marco Turchi " SentiWords: Deriving a High Precision and High Coverage Lexicon for Sentiment Analysis"This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2015.2476456, IEEE Transactions On Affective Computing Journal Of Latex Class Files, VOL. 6, NO. 1, JANUARY 2007

[5] Neha Raghuvanshi, Prof. J.M. Patil "A Brief Review on Sentiment Analysis" International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016 978-1-4673-9939-5/16/$31.00 ©2016 IEEE

[6] Mr. S. M. Vohra, Prof. J. B. Teraiya "A Comparative Study Of Sentiment Analysis Techniques" Journal Of Information, Knowledge And Research In Computer Engineering Issn: 0975 – 6760| Nov 12 To Oct 13 | Volume – 02, Issue – 02

[7] Prajwala T R "A Comparative Study on Decision Tree and Random Forest Using R Tool"International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 1, January 2015