

OPINION ANALYSIS ON TWITTER DATA USING R- PROGRAMMING

J.DavidSukeerthikumar(Ph.D)
Asst Prof (CSE)
Santhiram Engineering College, Nandyal

Y.Suteja(16X51A0586)
Dept.Of CSE
Santhiram Engineering College,Nandyal.

S.Supriya(16X51A0585)
Dept.Of CSE
Santhiram Engineering College,Nandyal.

A.Supriya(16X51A0584)
Dept.Of CSE
Santhiram Engineering College,Nandyal.

P.Uttejitha (16X51A0589)
Dept.Of CSE
Santhiram Engineering College,Nandyal.

Abstract:

Opinion mining also termed as sentiment analysis basically utilized to implement an automated algorithm to analyze and classify opinion. If we are interested to get the idea, the extent to which the Indian citizen likes or dislikes money demonetization or for instance consider a product marketer who on his present company market image, would like to decide upon, launching of his new product. A film celebrity who would like to judge his or her present popularity so as to decide on the perfect time to launch his or her new movie, for any of such case an obvious solution would require analysis of opinion from a random sample of people. For such public opinion measurement, various survey tools and techniques are available. With the sudden increase in text based social media, lump of people simulcast their point of view and ideas on a large range of issues. We can study this data available from public to conclude population attitudes to understand the current trends of the market. This study presents a very easy, cost and time effective approach that expose the opinions of much larger public which otherwise would have been not possible. The study presents an exhaustive study on the efficiency of R language in opinion mining and how opinion data can be extracted from twitter database. Extorting the opinion of people from social media text provides

a rich and interesting context of database to analysis

Key words: Concept-level sentiment analysis, Big data, Twitter, Social data analysis using R

Introduction:

People make judgments about the world around them when they are living in the society. They make positive and negative attitudes about people, products, places and events. These types of attitudes can be considered as sentiments. Sentiment analysis is the study of automated techniques for extracting sentiments from written languages. Growth of social media has resulted in an explosion of publicly available, user generated text on the World Wide Web. These data and information can potentially be utilized to provide real-time insights into the sentiments of people [1]. Blogs, online forums, comment sections on media sites and social networking sites such as Facebook and twitter all can be considered as social media. These social media can capture millions of peoples' views or word of mouth. Communication and the availability of these real time opinions from people around the world make a revolution in computational linguistics and social network analysis. Social media is becoming an increasingly more important source of information for an enterprise. On the other

hand people are more willing and happy to share the facts about their lives, knowledge, experiences and thoughts with the entire world through social media more than ever before. They actively participate in events by expressing their opinions and stating their comments that take place in society. This way of sharing their knowledge and emotions with society and social media drives the businesses to collect more information about their companies, products and to know how reputed they are among the people and thereby take decisions to go on with their businesses effectively. Therefore it is clear that sentiment analysis is a key component of leading innovative Customer Experience Management and Customer Relationship Marketing focused enterprises. Moreover for businesses looking to market their products, identify new opportunities and manage their reputation. As businesses look to automate the process of filtering out the noise, understanding the conversations, identifying the relevant content and take appropriate action upon it. Many are now looking to the field of sentiment analysis. In the era which we live today, sometimes known as information age, knowledge society; having access to large quantities of information is no longer an issue looking at the tons of new information produced everyday on the web. In this era, information has become the main trading object for many enterprises. If we can create and employ mechanisms to search and retrieve relevant data and information and mine them to transfer it to knowledge with accuracy and timeliness, that is where we get the exact usage of this large volume of information available to us. However, in many cases these relevant data and information are not found in structured sources such as tables or databases but in unstructured documents written in human language. Human languages are ambiguous and the same sentiment can be used to express two different ideas in two different contexts. Moreover some people use different jargon, slang communications and short forms of the words for their ease. Therefore, it is difficult to gauge and measure the sentiments accurately in terms of their polarity such as positive, negative or neutral and the subjectivity of sentiments [2]. Most solutions in the market today rely on simple

Boolean terms to express sentiment about a post, tweet, Facebook wall post etc. But this is not enough to address the above mentioned problems in the area of sentiment analysis and it will not generate precise and timely knowledge for aggregate sentiments. In order to get accurate knowledge after analyzing a sentiment, it should thoroughly consider solving the issues mentioned above. Most other systems that try to give solutions for these issues are still on research level, some systems also try to analyze sentiments from multiple languages and few systems which address some of the above mentioned drawbacks are available commercially also. This paper reveals an approach which is implemented as a tool that can analyze sentiments on twitter social media addressing above issues and then develop an application to generate knowledge that can be useful for business environments using people's attitudes about their products and services.

LITERATURE REVIEW:

This section illustrates other similar work related to analyzing sentiments. Most of these approaches analyze sentiments as positive and negative while some approaches are in research level and few more are commercially available.

Adobe Social Analytics:

Adobe Social Analytics basically measures the impact of social media on businesses by understanding how conversations on social networks and online communities influence marketing performance. After capturing and understanding the conversations going on, it correlates the impact of those conversations with key business matrices such as revenue and brand value. Other than that it measures the interactions that businesses have with their customers in social media including how Facebook posts drive site visitors and purchase. Adobe Social Analytics uses a natural language processing algorithm to implement sentiment analysis.

Brandwatch Sentiment Analysis:

Brandwatch is also a sentiment analysis tool developed by a team of PhD qualifiers in the United Kingdom; this is also commercially

available currently. Through this tool they are trying to access whether a sentiment is positive, negative or neutral [4]. Sentiment140 This is an online tool for analyzing sentiments of Twitter social network. This tool allows discovering the sentiment of a brand, product or topic on Twitter. This was created by three Computer Science graduate students at Stanford University and their main focus is analyzing the languages English and Spanish. Sentiment140 basically states whether the specified brand, product or topic is positive, negative or neutral [5].

Social Mention:

Social Mention is a social media search and analysis platform which analyses user sentiments through social media. This is also an online tool that allows tracking what people are saying about a particular brand, product or topic in real time. This tool allows the user to define a time period in which to analyze user sentiments.

TweetFeel:

TweetFeel is also a web tool that analyzes sentiments of the given input through the twitter social media. This gathers real time data on Twitter, about the search items and evaluates those tweets into positive and negative categories in real time. This uses machine learning based sentiment analysis which enables to get much clearer feeling about sentiments. Determining the Semantic Orientation of Terms through Gloss Classification Sentiment classification is a recent sub discipline of text classification which is concerned not with the topic a document is about, but with the opinion it expresses. In this approach of sentiment classification it uses a method that is based on the quantitative analysis of the glosses of such terms, i.e. the definitions that these terms are given in on-line dictionaries, and on the use of the resulting term representations for semi supervised term classification [6].

Sentiment Analysis using Adjectives and Adverbs:

While most work in sentiment analysis determine its polarity using specific parts of speech such as adjectives, verbs and nouns, in this approach it uses Adverb-Adjective Combinations (AACs) to determine the strength of subjective expressions of a sentence. Instead of aggregating scores of both adverbs and adjectives using simple scoring functions, it proposes an axiomatic treatment of AACs based on the linguistic classification of adverbs. Three specific AAC scoring methods that satisfy the axioms are presented [7]. The specialty in our system is, it does not only analyze the sentiments, and instead it uses the analyzed sentiment scores to provide product profile, trend analysis and forecasting for the user.

DESIGN AND IMPLEMENTATION:

There are four main modules naming, crawler, sentiment analysis tool, data mining module and dashboard in this project. The design and implementation details of those modules are described as follows.

Crawler:

The basic purpose of the crawler is to gather social media data to a local data source for ease of analysis. For this it had to use a twitter API to get access to twitter data. There are a set of streaming APIs offered by Twitter which gives developers access to Twitters' global stream of tweet data. Twitter offers several endpoints that have been customized to certain use cases such as Public streams, User Streams and Site Streams. Out of these three, for the crawler in this project we use the Public stream which is suitable for users or topics and data mining which reads the stream and directs those data to a database.

Sentiment Analysis Tool:

This is the tool that analyzes user sentiments and gets the correct polarity of the given sentiments. Here it gets the data from the database that has been crawled; as inputs to this tool. Inside this tool, it uses different machine learning techniques to get the most accurate answers for the given sentiments through a proper classifier. The purpose of this classification task is to classify sentiments automatically basically into positive, negative and neutral categories which mean choosing the correct class label for a given input. Since it uses supervised classification, it is a prerequisite that there need to be a labeled text corpus into categories to train, test and build the classifier. The next most important task after having a labeled text corpus is to find a way to extract features out of labeled corpus to train the classifier. The entire system depends on how good this method of feature extraction is. Therefore, it uses many different feature extraction methods in sentiment analysis as follows.

- Unigrams - Take each word individually in sentences as the feature set of corresponding category. Here it does not consider any relationship between words.
- Unigrams except Stop Words – Same feature extraction as above except it does not consider stop words which is a list of words that frequently appears in almost all sentences with no meaning.
- Bigrams – Take each adjacent two words in sentences as the feature set of corresponding category.
- Bigrams except Stop Words – This is same as Bigrams feature set, except words in Stop Words list.
- Most Informative Unigrams and Bigrams – Get the feature set with unigrams and bigrams with highest informative and highest frequency. Out of these feature extraction methods, the last one which is most informative unigrams and bigrams were selected to use in our approach. Sometimes extracting too many features reduces the accuracy; therefore, in the above selected method

it does not remove stop words because when it does, it reduces the accuracy. Then, extracted features need to be trained using a supervised classifier. As supervised classifiers, it uses Naïve Bayes classifier and Maximum Entropy classifier basically [8].

Data Mining:

When implementing the product profiling, it uses the decision tree after comparing it with the clustering technique and for the trend analysis and forecasting it used Holt Winters method which is capable of analyzing seasonal data and predict proper values for the future.

Results:

In this project, we find out the view of different people on the demonetization through analyzing the tweet from twitter using the Naïve Bayes and Maximum Entropy.

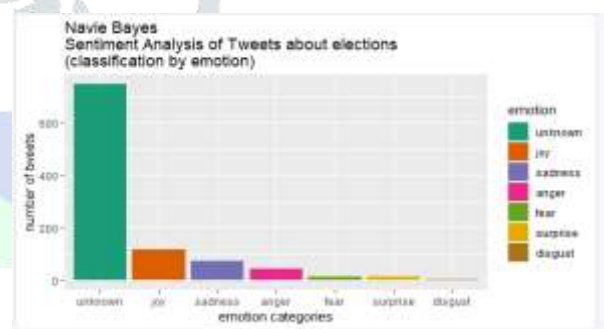


Fig.1 Naïve Bayes Classification by Emotion

Naive Bayes classification shows the emotions of Indian Citizens on demonetization 30% are joy, 10% as sad and few are fear and disgust. Here Naive Bayes classification uses the predefined features so if the tweets don't have any words those are defined then those opinions are classified as "unknown".

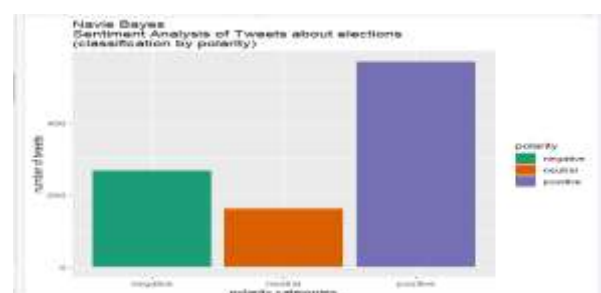


Fig.2 Naïve Bayes Classification by Polarity

