

Sentiment Analysis of Top Colleges in India Using Twitter Data

Pallavi S, Rachana C, Ramya K V, Vidyashree, Gangadhar Immadi,
Dr. Jitendranath Mungara

ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Nowadays, peoples' reviews and opinions that are available on various websites are one of the most critical factors in formulating our views and influencing the success of a brand, product or service. With the growth of social media in the world, and because it is easily accessible, stakeholders often take to expressing their opinions on popular social media, namely Twitter. Even though twitter data is extremely informative, it presents us with a challenge for analysis because of its large and disorganized nature. Here we are trying to dive into the novel domain of performing sentiment analysis of people's opinions regarding top colleges in India. Machine learning is used to get a more accurate result.

Keywords : Sentiment Analysis, Machine Learning, Opinion Mining, Natural Language Processing, Twitter, Multilayer perceptron(MLP).

I. INTRODUCTION

Social Media has captured the attention of the entire world as it is thundering fast in sending thoughts across the globe, user friendly and free of cost requiring only a working internet connection. People are extensively using this platform to share their thoughts loud and clear. Twitter is one such well known micro-blogging site getting around 500 million tweets per day .Each user has a daily limit of 2,400 tweets and 140 characters per tweet. Twitter users post (or 'tweet') every day about various subjects like products, services, day to day activities, places, personalities etc. Hence, Twitter data is of Great germane as it can be used in various scenarios where companies or brands can utilize a direct connection to almost each of their client or user and thereby, improve upon their product. Consider a dis-satisfied costumer of a telecommunication company voicing out his/her grievances about a particular plan he/she is subscribed to. Twitter also serves as a huge platform for users to know more and get direct comments about a product or a service in which they are interested. Opinions and reviews in the form of tweets from customers, potential users and critics can easily influence the image and consequently, demand of a product/service being provided by a company. Hence, whether the

stakeholder's. Opinion is positive/negative about their offering becomes a crucial and pressing question for the organization to ask and monitor.

II. RELATED WORKS

Mining opinions and analysing sentiments from social network data help in various fields such as even prediction, analysing overall mood of public on a particular social issue. It involves analysing the mood of the society on particular news from Twitter posts. It increases the accuracy of classification by including Natural Language Processing Techniques[1]. Sentiment Analysis (SA) and summarization has recently become the focus of many researchers, because analysis of review is beneficial and demanded in many different applications. One of the application is product-based sentiment summarization of multidocuments with the purpose of informing users about advantages and disadvantages of different products. This introduces a novel solution to target-oriented sentiment summarization and Sentiment analysis of short informal texts with a main focus on Twitter posts known as "tweets" [2]. Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in source text. Social media is generating a huge amount of sentiment rich data. That will be

represented in the form of tweets, status updates, blog posts etc. In order to know the opinion of the crowd, Sentiment analysis of the user generated data is very useful. Due to the presence of slang words and misspellings, Twitter sentiment analysis is difficult compared to general sentiment analysis. The maximum limit of characters that are allowed in Twitter is 140. The two strategies used for analysing sentiments from the text are Knowledge base approach and Machine learning approach. This is used to analyze the twitter posts about electronic products like mobiles, laptops using Machine Learning approach. By doing sentiment analysis in a specific domain, it is possible to identify the effect of domain information in sentiment classification. It gives a new feature vector for classifying the tweets as positive, negative and extract peoples' opinion about products[3]. A method for sentiment classification of a text document using high-order n-grams utilizes a multilevel embedding strategy to project n-grams into a low-dimensional latent semantic space where the projection parameters are trained in a supervised fashion together with the sentiment classification task[5]. Sentiment analysis plays a major role in brand and product positioning, consumer opinion detection, market research and customer relationship management. Necessary part of information gathering for market research is to find the opinion of people about the product. The availability and popularity of like online review sites and personal blogs, more chances and challenges arise as people now can, and do use information technologies to understand others opinions[6].

III. SYSTEM DESIGN

The tweets are extracted from the twitter API directly or imported from the excel sheets. Once the user enters the keyword i.e. the college name, the data about the particular college is retrieved. The retrieved data goes through pre-processing step where all stop words are removed from the tweets. Later sentiment analysis process will take place. Each tweet is categorised either into positive, negative or neutral sentiment. Machine learning algorithm helps to categorise the data, if any of the keywords are not defined in the database.

SYSTEM ARCHITECTURE

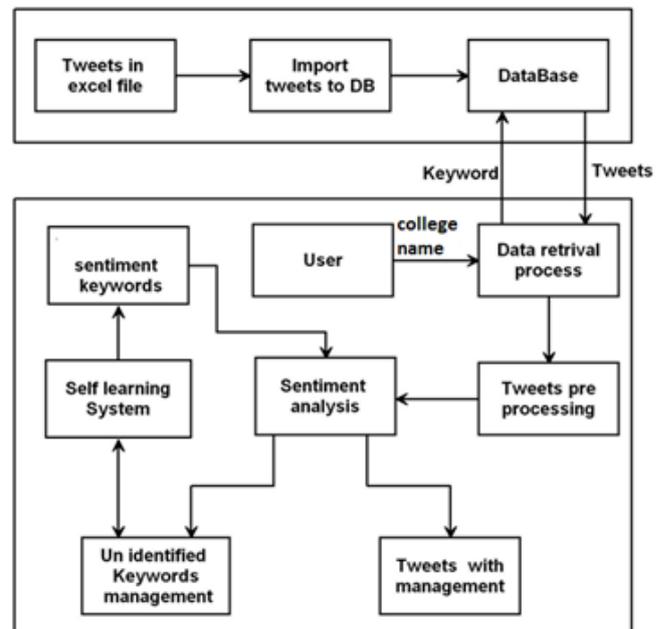


Figure 1 : System Architecture

Module used: Tweets Import Module In this module, tweets are retrieved from the twitter API dynamically based on the college name input. In developers console, first we need to create a twitter account to retrieve tweets from the twitter API account, After creating account it generates the consumers token key and access token key, with the help of these keys, we are going to communicate with twitter API to retrieve tweets. The retrieved tweets are imported into database.

Preprocessing Module

In this module, the tweets which are imported to database from the twitter API, these tweets consist of unnecessary words, whitespaces, hyperlinks and special characters. First we need to do filtering process by removing all unnecessary words, whitespaces, hyperlinks and special characters.

Machine Learning and word standardization System

In this module, first dictionary is initialized (first iteration dictionary). In the dictionary generally the positive, negative neutral and nouns are initialized. All big data and data mining projects based on the trained data (initialization of words). So initialization of the trained data is very important. In the self-learning system, word standardization is done, and not

considering past, present and future status of the words, only words will be considered.

Sentiment Analysis Module

In this module, preprocessed tweets are fetched from the database one by one. First we need check one by one keyword whether that keyword is noun are not, if noun we will remove it from the particular tweet. After that the remaining keywords checked with sentiment type, whether that keywords are positive sentiment or negative sentiment or neutral sentiment or undecided able sentiment. The remaining keywords in the tweet which does not belongs to any of the sentiment will be assigned temporary sentiment based on the more count of positive, negative and neutral. In the second iteration if the remaining word crosses the threshold of positive, negative or neutral, that keyword permanently added as expansion in the dictionary. Finally sentiment of the tweet is detected based on the positive, negative and neutral words in the particular tweet.

IV.CONCLUSION

Sentiment analysis is an effective way to classify the opinions formulated by people regarding any topic, service or product. Automation of this task makes it easier to deal with the massive amount of data being produced by social websites like twitter on real time basis. Machine learning algorithm yields promisingly accurate predictions on unseen data. So in our project we analyse the twitter data related colleges, classify them and provide the users with the best review.

V. REFERENCES

- [1]. Kanakaraj M, Guddeti R M.R. performance analysis of ensemble methods on twitter sentiment analysis using NLP techniques published by IEEE in the year 2015.
- [2]. Bespalov D, Bai B, Qi Y. Performance analysis of ensemble methods on twitter sentiment analysis using NLP techniques published by IEEE in the year 2011.
- [3]. Neethu M. S. and Rajasree R Sentiment Analysis in Twitter using Machine Learning Techniques published by IEEE in the year 2013.
- [4]. GauravBhat, Ankushmittal. Sentiment analysis of top colleges in india using twitter data published by IEEE in the year 2016.
- [5]. Bahrainian S. A, Dengel A. Sentiment classification based on supervised latent n-gram analysis published in the year 2013
- [6]. Jotheeswaran J. and Koteeswaran S. Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis Published by IEEE in the year 2015.