



SENTIMENT ANALYSIS OF TEXT FEEDBACK

Miss. Prachiti V. Likhitkar
Department of Computer Science & Engineering,
Sipna College of Engineering and Technology, Amravati

Prof. Seema Rathod
Department of Computer Science & Engineering
Sipna College of Engineering and Technology, Amravati

Miss. Vaishnavi M. Daware
Department of Computer Science & Engineering,
Sipna College of Engineering and Technology, Amravati

Sipna College of Engineering and Technology, Amravati

Miss. Rutuja J. Deshmukh
Department of Computer Science & Engineering
Sipna College of Engineering and Technology, Amravati

Miss. Sakshi P. Divekar
Department of Computer Science & Engineering,
Sipna College of Engineering and Technology, Amravati

Abstract:

With the increase in the number of mobile applications in the day to day life, it is important to keep track as to which ones are safe and which ones aren't. The user cannot always get correct or true reviews about the product on the internet. We can check for user's consumer trends comments on multiple applications. The reviews may be fake or genuine. Analyzing the rating and reviews together involving both user and admins comments, we can

determine whether the app is genuine or not.

Using consumer trends analysis and data mining, the machine is able to learn and Analysis the consumer trends, emotions about reviews and other texts. The manipulation of review is one of the key aspects of App ranking fraud. By using consumer trends analysis and data mining, analyzing reviews and comments can help to determine the correct application for both Android and iOS platforms.

Miss. Neha P. Ingle
Department of Computer Science & Engineering,

Introduction :

INTRODUCTION TO consumer trends ANALYSIS

consumer trends Analysis is process of collecting and analyzing data based upon the person feelings, reviews and thoughts. consumer trends analysis often called as opinion mining as it mines the important feature from people opinions. consumer trends Analysis is done by using various machine learning techniques, statistical models and Natural Language Processing (NLP) for the extraction of feature from a large data.

consumer trends Analysis can be done at document, phrase and sentence level. In document level, summary of the entire document is taken first and then it is analyzed whether the consumer trends is positive, negative or neutral. In phrase level, analysis of phrases in a sentence is taken in account to check the polarity. In Sentence level, each sentence is classified in a particular class to provide the consumer trends.

consumer trends Analysis has various applications. It is used to generate opinions for people of social media by analyzing their feelings or thoughts which they provide in form of text. consumer trends Analysis is domain centered, i.e. results of one domain cannot be applied to other domain. consumer trends Analysis is used in many real life scenarios, to get reviews about any product or movies, to get the financial report of any company, for predictions or marketing.

Twitter is a micro blogging platform where anyone can read or write short form of message which is called tweets. The amount of data accumulated on twitter is very huge. This data is unstructured and written in natural language. Twitter consumer trends Analysis is the process of accessing tweets for a particular topic and predicts the consumer trends of these tweets as positive, negative or neutral with the help of different machine learning algorithm

Literature Review :

Trend analysis and based on that predicting public opinions. It plays important role, many researchers working on automatic technique of extraction and analysis of huge amount of twitter data. In [1] author compare six trend detection method and find that standard natural language processing technique perform well for social streams on particular topic. They conclude that n-gram give best performance other than state of-art techniques. In [4], the authors have used three different machine learning algorithms Naïve Bayes, Decision Trees and Support Vector Machine for sentiment classification of Arabic dataset which was obtained from twitter. This research has followed a framework for Arabic tweets classification in which two special sub-tasks were performed in preprocessing, Term Frequency-Inverse Document Frequency (TF-IDF) and Arabic stemming. They have used one dataset with three algorithms and performance has been evaluated on the basis three different information retrieval metrics precision, recall, and f-measure. In [6] author proposed supervised learning techniques to classify twitter trending topic for that they use text based and network-based classifier and conclude C5.0 gave best performance. In [19] author propose model which predict public opinion on political event by Applying different classifier which predict that whether mood is positive or negative. In [26], the authors proposed a way to get the pre labeled data from twitter which can be used to train SVM classifier. They used the twitter hash tags to judge the polarity of tweet. To analyze the accuracy of proposed technique, a test study on the classifier was conducted which showed the result with the accuracy of 85%. The authors in [27] introduced a new technique to classify the sentiment of tweets as positive or negative. They presented and discussed the results of machine learning algorithms for twitter sentiment analysis by using distant supervision. Training data, the authors used consisted of tweets with emotions which were used as noisy labels.

According to authors, the machine learning algorithms such as Naive Bayes, Maximum Entropy and SVM when trained with emotion tweets can have accuracy more than 80%. The study also highlighted the steps used in preprocessing stage of classification for high accuracy. In [28] sentiment analysis perform using SVM in that two pre classified datasets of tweets are used then do comparative analysis, they use measures Precision, Recall and F-Measure.

Proposed work :

Consumer trends arrangement utilizing unsupervised learning: In the unsupervised order, the content is characterized by contrasting it and given words or dictionaries. The feeling an incentive for these words or dictionaries is already characterized.

In the first step to be able to access Twitter data programmatically we need to create consumer trends and register an app on twitter developers website for authentication and thereafter we can access data by using Twitter API.

Registering App: On registering the app we will receive consumer key and consumer secret key. Next, from the configuration page of the app, we will get access_token and

access_token_secret, which will be used to get access twitter on behalf of our application. We must keep these authentication tokens private as they can be misused.

Accessing Data: Twitter provides REST API's to connect with their service. We used one python library to access twitter REST API's called tweepy. It provides wrapper methods to easily access twitter REST API. To install tweepy we used command pip install tweepy.

3. Live fetching data from tweeter API
4. Create a NPL program to read text in the statement.
5. Filter the text from each tweet from unwanted words like helping verb, Noun , number, Emoticons etc.
6. Internet interfacing for program using tweeter credential
7. Indexing all the tweet in custom format which will get the number of dataset available.
8. Uses of open-source language which is python in the form backend or frontend programming

Data Set:

This sentiment analysis dataset contains tweets. Each tweet is classified either positive, negative or neutral. The included features including Twitter ID, sentiment confidence score, sentiments, negative reasons, retweet count, name, tweet text, tweet coordinates, date and time of the tweet, and the location of the tweet.

NLP Algorithm :

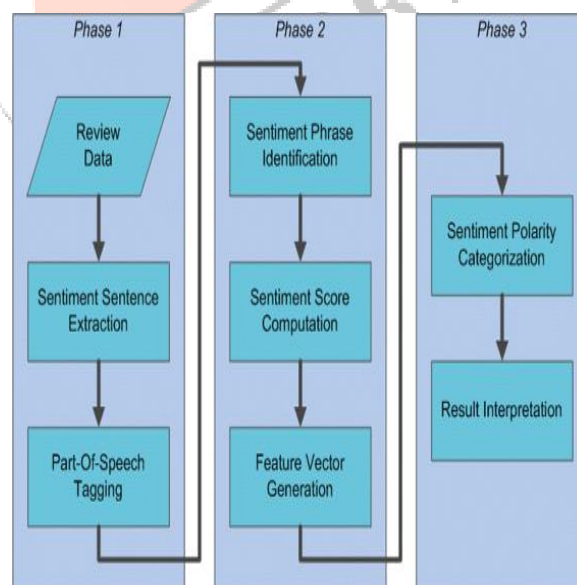


Fig 1. NLP Algorithm

Methodology:

1. To make GUI Structure for user friendly working
2. Trend analysis using graphical format of tweets response.

Storing Data: We access all tweet data from personal profile and store it for our analysis steps. Tweepy library provides simple cursor interface to iterate through all the tweets and store them in file.

Preparing Data: Before we begin to analyze the twitter data, it's important to understand the structure of the tweet as well as pre-process the data to remove non-useful terms means stop words. Preprocessing is in the simple term means to take in the data and prepare the data for optimal output considering our requirement.

Data pre-processing: Pre-processing removes stop word, handling of negation; misspell correction, positive word lists of each tweet and negative word lists of each tweet.

- **Filtering:** Filtering is a process that removes unnecessary parts or information from
- **URL:** Entire URL removed from the sentence or input file after checking the whole sentence or input file. These links are replaced by the empty space.
- **Username:** Sometimes user used any username or @ symbol before any tweet. These types of usernames or @ replaced by empty space.
- **Duplicate or repeated characters:** Users sometimes use informal language in tweets. For example, users mostly write 'baaad' in place of the bad word.

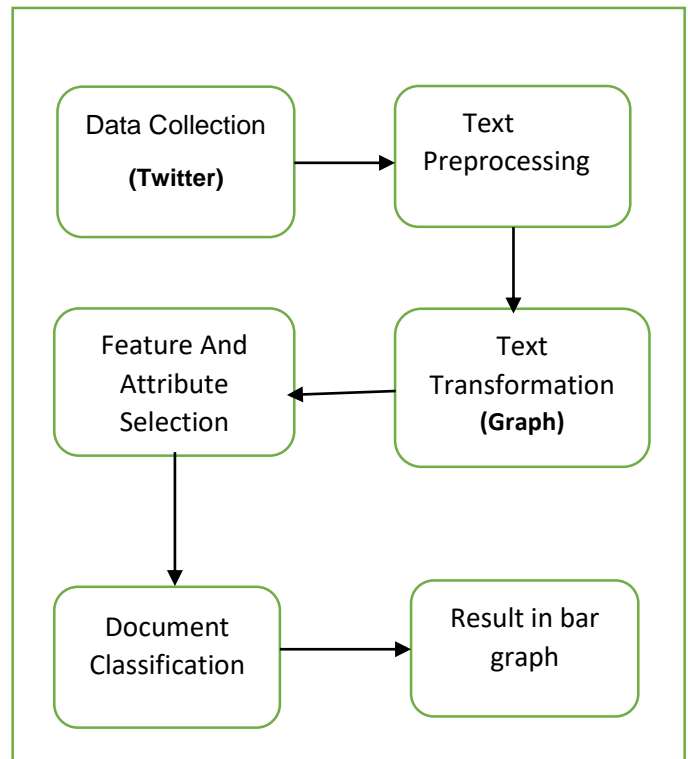


Fig. 2 Diagrammatical Representation

Required Parameters :

- To do the consumer trends analysis of tweets, the proposed system first extracts the twitter posts from twitter by user. The system can also compute the frequency of each term in tweet speech.
- We will obtain a classification of polarity of consumer trends into positive, negative or neutral and prepare a plot of the same using python module like matplotlib.
- Twitter is large source of data, which make it more attractive for performing consumer trends analysis.
- We perform analysis on around 100 tweets total, so that we analyze the results, understand the patterns and give a review on people opinion.
- We saw different people have different consumer trends results according to their progress and working procedure.
- tweets we collect with the help of keyword. It can be used for finance, marketing, reviewing and many more.



Fig. 3 Sentiment Analysis

Result :

Trends Analysis is a process of extracting feature from user's thoughts, views, feelings and opinions which they post on any social network websites. The result of trends analysis is classification of natural language text into classes such as positive, negative and neutral.

This graph is collected from twitter of text Covid 19.

By collecting the views of covid 19 from twitter we get graph in the form of positive, negative and neutral form.

consumer trends analysis is used to identifying people's opinion, attitude and emotional states. The views of the people can be positive or negative. Commonly, parts of speech are used as feature to extract the consumer trends of the text. An adjective plays a crucial role in identifying consumer trends from parts of speech. Sometimes words having adjective and adverb are used together then it is difficult to identify consumer trends and opinion.

To do the consumer trends analysis of tweets, the proposed system first extracts the twitter posts from twitter by user. The system can also compute the frequency of each term

in tweet. We will obtain a classification of polarity of consumer trends into positive, negative or neutral and prepare a plot of the same using python module like matplotlib.

Twitter is large source of data, which make it more attractive for performing consumer trends analysis. We perform analysis on around 100 tweets total, so that we analyze the results, understand the patterns and give a review on people opinion. We saw different people have different consumer trends results according to their progress and working procedure.

We also saw how any social event, speech or rally cause a fluctuation in consumer trends of people. We also get to know which policies are getting more support from people which are started by any of these parties. It can be used for any purpose based on tweets we collect with the help of keyword. It can be used for finance, marketing, reviewing and many more.

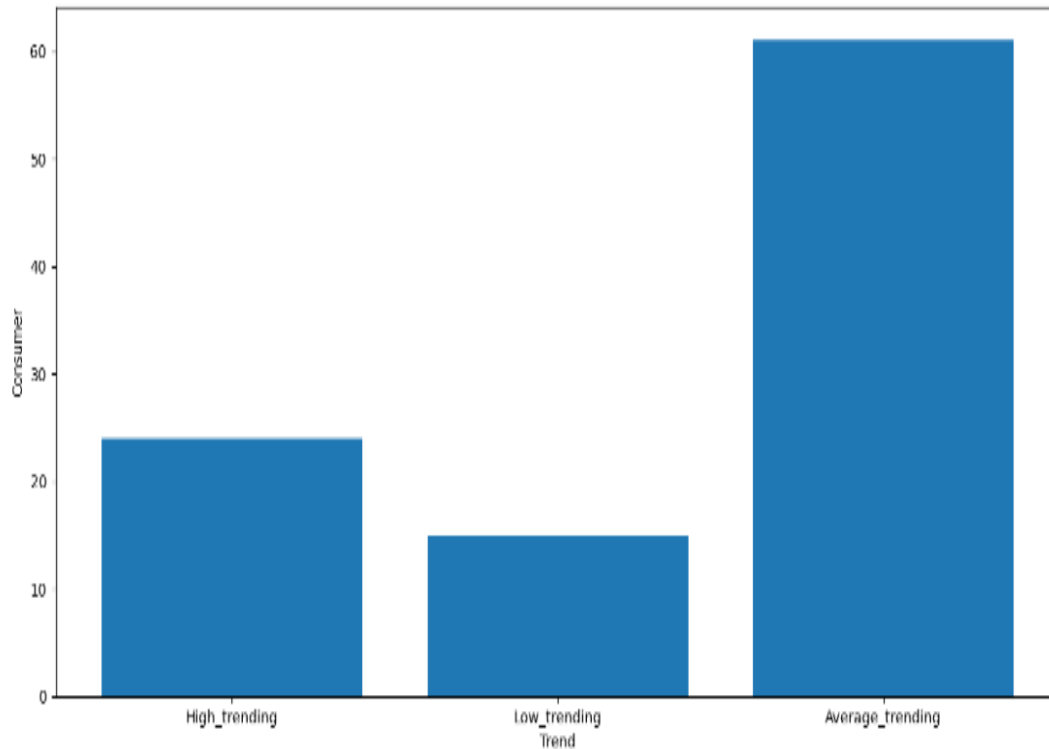


Fig. 4 Graph of Covid 19 Text

Conclusion :

Thus we first to work on trends analysis. Then classify text by overall trends , not just by topic e.g., classifying movie review either positive or negative. Then apply machine learning algorithm on movie review database which results that these algorithms out-perform human produced algorithms. The machine learning algorithms they use are Naïve-Bayes, maximum entropy, and support vector machines.

Finally get the result in the form of graph indicates positive, negative and neutral form.

Reference :

- [1]H. Zang, "The optimality of Naïve-Bayes", Proc. FLAIRS, 2004
- C.D. Manning, P. Raghavan and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, pp. 234-265, 2008
- 2) McCallum and K. Nigam, "A comparison of event models for Naive Bayes text classification", Proc. AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41-48, 1998
- 3) M. Schmidt, N. L. Roux and F. Bach, "Minimizing finite Sums with the Stochastic Average Gradient", 2002
- 4) Y. LeCun, L. Bottou, G. Orr and K. Muller, "Efficient BackProp", Proc. In Neural Networks: Tricks of the trade 1998.
- 5) T. Wu, C. Lin and R. Weng, "Probability estimates for multi-class classification by pairwise coupling", Proc. JMLR-5, pp. 975-1005, 2004

6) ss“Support Vector Machines”
[Online], <http://scikit-learn.org/stable/modules/svm.html#svm-classification>, Accessed Jan 2016

7) P. Pang, L. Lee and S. Vaithyanathan, “Thumbs up? consumer trends classification using machine learning techniques”, Proc. ACL-02 conference on Empirical methods in natural language processing, vol.10, pp. 79-86, 2002

8) P. Pang and L. Lee, “Opinion Mining and consumer trends Analysis. Foundation and Trends in Information Retrieval”, vol. 2(1-2), pp.1-135, 2008

9) E. Loper and S. Bird, “NLTK: the Natural Language Toolkit”, Proc. ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics ,vol. 1,pp. 63-70, 2002

10) H. Wang, D. Can, F. Bar and S. Narayana, “A system for real-time Twitter consumer trends analysis of 2012 U.S.presidential election cycle”, Proc. ACL 2012 System Demonstration, pp. 115-120, 2012

11) O. Almatrafi, S. Parack and B. Chavan, “Application of location-based consumer trends analysis using Twitter for identifying trends towards Indian general elections 2014”. Proc. The 9th International Conference on Ubiquitous Information Management and Communication,2015

12. L. Jiang, M. Yu, M. Zhou, X. Liu and T. Zhao, “Target-dependent twitter consumer trends classification”, Proc. The 49th Annual Meeting of the Association