

# My Opinion: AnRGE to Enrich Your Knowledge

Shubhankar V. Mujumdar<sup>1</sup>, Manish Kumar<sup>2</sup>, Rohit More<sup>3</sup>, Sudarshan Mane<sup>4</sup>,  
Mr. Sagar Salunke<sup>5</sup>

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India

Savitribai Phule Pune University, Pune, India

---

**Abstract:** At present people are using various online platforms for discussion purpose. We can ask questions and get answers to our questions for other people using that platform. Some of these platforms are: Quora, face book pages, Google plus, etc. and these questions are from various fields like politics, science and technology, sports, entertainment, education, etc. The answers that we get on these platforms are mainly in the form of comments that include text data. To get a complete understanding of what is going on the user have to read all the comments and some of this comments can even have same content. It is wastage of time and energy and still the user doesn't get the answer that he is looking for. A way to provide better solution to these questions is to analyze the text data and generate a report of it. Report will be the summarized version of the entire conversation. We have tried to do this with the help of text mining techniques to extract information from these user responses and generate report having pie charts or graphs. The way asking question is in the form of multiple choice questions. The person posting question will post the question along with some option that he can think of as answers to the his question. The users submitting responses can select from the pre-defined options or can submit a new response. Whether to add that option or ignore it will be up to the owner of the question. Depending on few parameters report will be generated that can be referred by the user at any time. It will save the user from the headache of reading the entire conversation. It will be a free platform developed in android. It will require a working internet connection. At present it will be used as a simple report generation platform but in future depending upon the user base it can extended for survey and research purpose by individuals and many different organization.

**Keywords:** K-means; NLP; Clustering; Opinion mining.

---

## I. INTRODUCTION

In today's world there are a lot of applications for different purpose available in android. You think of something and you'll probably find an application for that purpose. Life is becoming very simple because of this. Still there are some areas where improvement is required.

One of those areas is the online discussion platforms that provide a platform to its users where they can discuss some topic or can ask questions related to different fields. These fields can be sports, education, religion, politics, entertainment, profession, etc.. The platforms that we are talking about are Quora, face book pages, Google plus, etc.. On these platforms we start with asking a question by posting it on the platform. It is categorized in the type of field to which the question belongs by using keyword generation algorithms and clustering algorithms. Once the categorization is done, other users can view it and if they like the question and have answer for that question they can reply to it. Answers are mainly in text format visible to everyone. Anyone having the same question can search for it and read the answers.

Even though these platforms do a great job in providing answers to these questions but there are some problems with these platforms. On these platforms the answers that are post are very long in size and to understand what it says one has to read the entire answers. It consumes more time. Also the data redundancy is present in many answers that means many replies can have same content and it means same. Also the numbers of replies are huge in some cases. To get complete understanding of what is going on one has to read the entire conversation and it is time consuming.

My Opinion is such a platform that will generate analyzed report on the basis of replies received for question. We will use K-means clustering algorithm and natural language processing for minimizing data redundancy. Here AnRGE stands for Analysis and Report Generation Engine.

## II. GOALS AND OBJECTIVES

There is a need to develop a platform where data redundancy is minimum and user get a more accurate answers to their question within minimum time period.

The objectives of developing this platform are:

- minimize the data redundancy in comments
- provide better answers to the questions
- save our user from wasting his precious in reading the entire thread of replies by generating an analyzed report

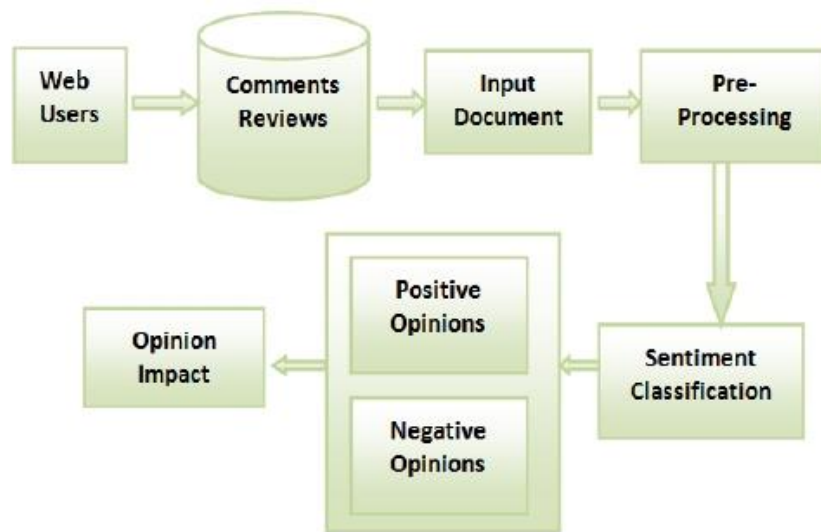
## III. LITERATURE SURVEY

In this paper we have referred "AN EXPERIMENTAL APPROACH OF K-MEANS ALGORITHM ON THE DATASET". This paper first discussed all the clustering techniques and then experiment on large dataset which is used for collecting books in different cities with price rates. They have implemented K-Means algorithm on the huge dataset using MATLAB and they have shown the result of the K- Means algorithm. We have made use of these techniques in our work. We have also referred [1]. In [1] different existing Text Work flow of opinion mining Algorithms i.e. Classification Algorithm, Association Algorithm, Clustering Algorithm were briefly reviewed, stating the merits / demerits of the algorithms. In addition some alternate implementations of the algorithms were proposed. Finally the logic of these algorithms were merged to generate an algorithm that would perform the task of Classification of a data set into some predefined classes, establish relationship between the classified date and finally cluster the data based on the association between them into groups. For the implementation of NLP we have refer the paper "A Peer Review of Feature Based Opinion Mining and Summarization". In this paper they have discussed about different techniques to perform sentiment analysis to decide whether a comment contains positive content or negative content.

## IV. IMPLEMENTATION

In this process, raw data taken and is pre-processed for in this platform, the question posted will be of the multiple choice question format. Here the options will be the different viewpoints that different users can have and these options will be provided by the owner of the question. The owner will provide as much options as he can think of. The other users who want to reply to the question can select from the options provided or can post a new view, if his view point is not already present. On these new viewpoints posted in the others category text mining algorithms will be applied to reduce the amount of text and reduce the data redundancy. The algorithms applied here are K-means clustering algorithm and NLP techniques. We have tried to implement opinion mining technique.

Opinion mining is a type of NLP which has ability to track the mood of the people about any particular product by review. Opinion Mining is a process of automatic extraction of knowledge by means of opinion of others about some particular product, topic or problem. The idea of Opinion mining and Sentiment Analysis tool is to process a set of search results for a given item based on the quality and features. Opinion mining is also called sentiment analysis due to large volume of opinion which is rich in web resources available online.



□ **Data source:** - the input from user

□ **OPINION MINING OR SENTIMENT ANALYSIS:**

Opinion mining is a technique which is used to detect and extract subjective information in text documents the evaluation of opinion can be done in two ways:-

Direct opinion gives positive or negative opinion about the object directly. For example, "The action in the movie is not good" (direct opinion.)

Comparison means to compare the object with some other similar objects. For example, "The action in movie -a is better than that of movie -b." (Comparison.)

**Steps:-**

Pre-processing In this process, raw data taken and is pre-processed for feature extraction. This Process has following sub phases:

1. Tokenization
2. Stop word Removal removes articles
3. Case Normalization

**Feature Extraction:**

The feature extraction phase deals with feature types [3] feature selection, feature weighting mechanism, reduction mechanisms. Feature Types

Types of features used for opinion mining could be:-

- 1: Term frequency
- 2: Term co-occurrence
- 3: POS
- 4: Opinion words
- 5: Negations
- 6: Syntactic dependency

**Feature Selection:**

1. Information gain

2. Odd Ratio

3. Document Frequency

#### Features weighting mechanism:

The mechanisms are of two types. They are

1. Term Presence and Term Frequency-

2. Term frequency and inverse document frequency (TFIDF)

#### □ DIFFERENT LEVELS OF OPINION MINING:

Document level Opinion Mining: - document level classification is a single review about a topic is considered.

Sentence level Opinion Mining:- the polarity of each sentence is calculated here.

Phrase level Opinion Mining: - The phrases that contain opinion words are found out and a phrase level classification is done.

#### □ TECHNIQUE:

We are going to use unsupervised learning technique. Unsupervised learning has no explicit targeted output associated with input. Class label for any instance is unknown so unsupervised learning is about to learn by observation.

## V. CONCLUSION

In this paper are going to present a platform named My Opinion that will help to analyze the user comments and generate a report on the bases of analyses done. We have used text mining techniques and natural language processing to process the comments to check whether someone commented positive content or negative content. We have implemented the techniques to perform sentimental analysis and provide our user a better solution.

## REFERENCES

- [1] A tutorial review on Text Mining Algorithms by Mrs. Sayantani Ghosh, Mr.Sudipta Roy, and Prof. Samir K. Bandyopadhyay. International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 4, June 2012.
- [2] An Analysis on Opinion Mining: Techniques and Tools by G. Angulakshmi, Dr.R.ManickaChezian International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.
- [3] A Peer Review of Feature Based Opinion Mining and Summarization by Padmapani P. Tribhuvan , S.G. Bhirud, Amrapali P. Tribhuvan International Journal of Computer Science and Information Technologies, Vol. 5 (1) ,2014, 247-250.
- [4] Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews by Peter D. Turney Proceedings of the 40<sup>th</sup>Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 417 424.
- [5] Experimental study of Data clustering using k-Means and modified algorithms by Dr. M.P.S Bhatia,Deepika Khurana.
- [6] A Survey of Text Mining Techniques and Applications by Vishal Gupta,GurpreetS. Lehal JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 1, NO. 1, AUGUST 2009.
- [7] Classification of Opinion Mining Techniques by Nidhi Mishra,C K. Jha International Journal of Computer Applications , Volume 56—No.13,October 2012.