Prediction of Customer Review using Opinion Mining

[1]Fauzan Padvekar, [1]Faisal Sothe, [1]Priya Yadav, [2]Najmuddin Aamer
1UG Student, Theem College of Engg, (Mumbai University) India.
[2]Professor, IT Dept, Theem College of Engg, (Mumbai University) India.

Abstract: Merchants selling products on the Web often ask their customers to share their opinions and hands-on experiences on products they have purchased. Unfortunately, reading through all customer reviews is difficult, especially for popular items, the number of reviews can be up to hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision. The Opinion Miner system designed in this work aims to mine customer reviews of a product and extract high detailed product entities on which reviewers express their opinions. Opinion expressions are identified and opinion orientations for each recognized product entity are classified as positive or negative.

Keywords: Opinion Mining, Naïve Bayes, Preprocessing, Feature Extraction.

I. INTRODUCTION

Opinions are central to almost all human activities and are key influencers of our behaviours. Our beliefs and perceptions of reality, and the choices we make, are, to a considerable degree, conditioned upon how others see and evaluate the world. For this reason, when we need to make a decision, we often seek out the opinions of others. This is not only true for individuals but also true for organizations.

Opinions and its related concepts such as sentiments, evaluations, attitudes, and emotions are the subjects of study of sentiment analysis and opinion mining. The inception and rapid growth of the field coincide with those of the social media on the Web, e.g., reviews, forum discussions, blogs, micro blogs, Twitter, and social networks, because for the first time in human history, we have a huge volume of opinionated data recorded in digital forms. Since early 2000, sentiment analysis has grown to be one of the most active research areas in natural language processing. It is also widely studied in data mining, Web mining, and text mining. In fact, it has spread from computer science to management sciences and social sciences due to its importance to business and society as a whole. In recent years, industrial activities surrounding sentiment analysis have also thrived. Numerous start-ups have emerged. Many large corporations have built their own in-house capabilities. Sentiment analysis systems have found their applications in almost every business and social domain.

Opinions are central to almost all human activities because they are key influencers of our behaviours. Whenever we need to make a decision, we want to know others' opinions. In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. Individual consumers also want to know the opinions of existing users of a product before purchasing it, and others' opinions about political candidates before making a voting decision in a political election. In the past, when an individual needed opinion, he/she asked friends and family. When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Acquiring public and consumer opinions has long been a huge business itself for marketing, public relations, and political campaign companies.

With the explosive growth of social media (e.g., reviews, forum discussions, blogs, micro-blogs, Twitter, comments, and postings in social network sites) on the Web, individuals and organizations are increasingly using the content in these media for decision making. Nowadays, if one wants to buy a consumer product, one is no longer limited to asking one's friends and family for opinions because there are many user reviews and discussions in public forums on the Web about the product. For an organization, it may no longer be necessary to conduct surveys, opinion polls, and focus groups in order to gather public opinions because there is an abundance of such information publicly available. However, finding and monitoring opinion sites on the Web and distilling the information contained in them remains a formidable task because of the proliferation of diverse sites. Each site typically contains a huge volume of opinion text that is not always easily deciphered in long blogs and forum postings. The average human reader will have difficulty identifying relevant sites and extracting and summarizing the opinions in them. Automated sentiment analysis systems are thus needed.

In recent years, we have witnessed that opinionated postings in social media have helped reshape businesses, and sway public sentiments and emotions, which have profoundly impacted on our social and political systems. Such postings have also mobilized masses for political changes such as those happened in some Arab countries in 2011. It has thus become a necessity to collect and study opinions on the Web. Of course, opinionated documents not only exist on the Web (called external data), many organizations also have their internal data, e.g., customer feedback collected from emails and call centres or results from surveys conducted by the organizations.

Due to these applications, industrial activities have flourished in recent years. Sentiment analysis applications have spread to almost every possible domain, from consumer products, services, healthcare, and financial services to social events and political elections. Many big corporations have also built their own in-house capabilities, e.g., Microsoft, Google, Hewlett-Packard, SAP, and SAS. These practical applications and industrial interests have provided strong motivations for research in sentiment analysis.

II. LITERATURE SURVEY

Studies on feature-based opinion mining have exploited various methods for feature extraction and refinement, including NLP and rule-based methods [6], [2], statistical methods [4], [9], and ontology-based methods [15]. R. Nitish [2] proposed a system to extract features from review data using association rule mining. The system selects frequent terms and then extracts features by measuring the similarities between selected terms. The main problem of this method is that the system only considers the information from the term itself, for example, term frequency, which does not reflect the relationship between a feature and its related opinion information.

Zulva Eachrina [6] proposed a feature extraction method using a rule-based approach. This method extracts a relatively large number of features compared with the amount of review data. For example, it generates 263 features from 45 reviews for digital cameras. The main reason for the extraction of so many features is that terms that have the same or similar meanings are not considered as the same features. For example, 'photo,' 'picture,' and 'image' all have the same meaning; however, they are considered as different features simply because they are different words. Consequently, this system could not provide proper summary information for the product. In FEROM, we solve this problem by reducing the number of features by merging words that have similar meanings using the semantic similarity between features and then providing reliable summary information for the product based on the merged features.

Neelam Duhan [5] proposed a feature extraction method for opinion mining that uses ontology. Although this method worked well semantically, the main problem is the maintenance of the ontology to address the constant expansion of the review data. In this system, the ontology is manually constructed and must be updated when new features are added. In addition, a concept that is not defined in the ontology is not able to be classified. Thus, it is necessary to construct an automatic system to avoid continued intervention.

In summary, previous studies on feature-based opinion mining do not consider the relationship between a term and its related opinion information and also do not merge words with the same or similar meanings. We propose FEROM to solve these problems.

III. SYSTEM DESIGN

1. System Working:

The system architecture of FEROM is shown in Fig. 1. The review crawler collects customer review data from online stores, and the review cleaner removes unnecessary content such as HTML tags and then stores the review data to the review database.

The pre-processor conducts morphological analysis of the review data including POS tagging, splits a compound sentence into multiple sentences, and performs stop word removal and stemming.

The feature extractor extracts product features from pre-processed review data. Feature extraction proceeds in three phases: feature selection selects a candidate feature in a sentence by looking for a noun phrase, opinion information extraction finds an opinion phrase that is associated with the candidate feature, and opinion phrase conversion replaces an opinion phrase expressed using a negative term with its antonym.

The feature refiner reduces the number of features by merging candidate features with the same or similar meanings, defined as homogeneous features. The feature refiner recognizes homogenous features by exploiting the feature ordering process that synchronizes the word orders of the features to detect synonymous feature candidates and the feature containment checking process that examines the subset superset relationship between the features to check for similarity between them. Finally, the feature merging process merges homogeneous features into a representative feature and also prunes the feature candidates that have significantly low frequencies and very small amounts of related opinion information.

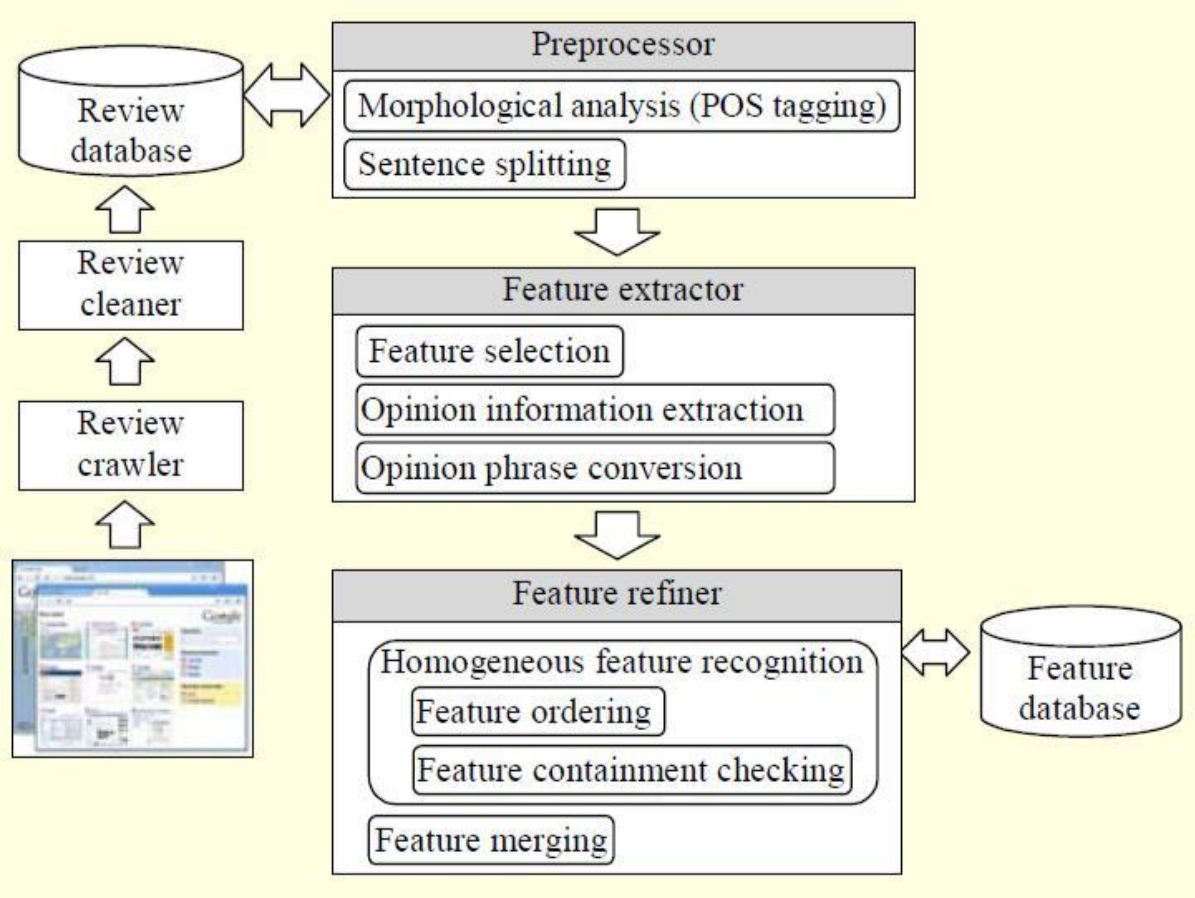

Fig 1: System Architecture of FEROM

IV. METHODOLOGY

After analysing the requirements of the task to be performed, the next step is to analyse the problem and understand its context. The first activity in the phase is studying the existing system and other is to understand the requirements and domain of the new system. Both the activities are equally important, but the first activity serves as a basis of giving the functional specifications and then successful design of the proposed system. Understanding the properties and requirements of a new system is more difficult and requires creative thinking and understanding

of the existing running system is also difficult, improper understanding of present system can lead diversion from solution.

1. Waterfall Model:

1.1 Overview:

The model that is basically being followed is the WATERFALL MODEL, which states that the phases are organised in a linear order.

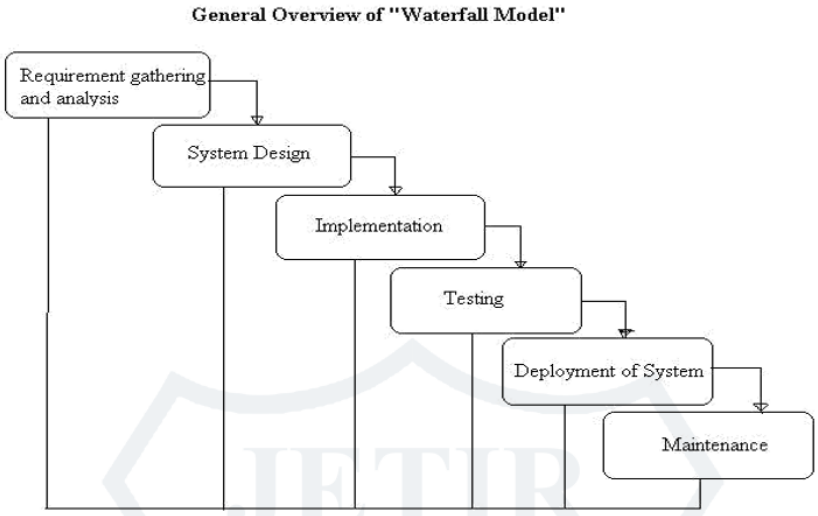

Fig 2: Waterfall Model

WATERFALL MODEL: Waterfall approach was first Process Model to be introduced and followed widely in Software Engineering to ensure success of the project. In "The Waterfall" approach, the whole process of software development is divided into separate process phases. The phases in Waterfall model are: Requirement Specifications phase, Software Design, Implementation and Testing & Maintenance. All these phases are cascaded to each other so that second phase is started as and when defined set of goals are achieved for first phase and it is signed off, so the name "Waterfall Model". All the methods and processes undertaken in Waterfall Model are more visible.

The stages of "The Waterfall Model" are:

- **Requirement Analysis & Definition:** All possible requirements of the system to be developed are captured in this phase. Requirements are set of functionalities and constraints that the end-user (who will be using the system) expects from the system. The requirements are gathered from the end-user by consultation, these requirements are analyzed for their validity and the possibility of incorporating the requirements in

the system to be development is also studied. Finally, a Requirement Specification document is created which serves the purpose of guideline for the next phase of the model.

- **System & Software Design:** Before a starting for actual coding, it is highly important to understand what we are going to create and what it should look like? The requirement specifications from first phase are studied in this phase and system design is prepared. System Design helps in specifying hardware and system requirements and also helps in defining overall system architecture. The system design specifications serve as input for the next phase of the model.
- **Implementation & Unit Testing:** On receiving system design documents, the work is divided in modules/units and actual coding is started. The system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality; this is referred to as Unit Testing. Unit testing mainly verifies if the modules/units meet their specifications.
- **Integration & System Testing:** As specified above, the system is first divided in units which are developed and tested for their functionalities. These units are integrated into a complete system during Integration phase and tested to check if all modules/units coordinate between each other and the system as a whole behaves as per the specifications. After successfully testing the software, it is delivered to the customer.
- **Operations & Maintenance:** This phase of "The Waterfall Model" is virtually never ending phase (Very long). Generally, problems with the system developed (which are not found during the development life cycle) come up after its practical use starts, so the issues related to the system are solved after deployment of the system. Not all the problems come in picture directly but they arise time to time and needs to be solved; hence this process is referred as Maintenance.

1.2 Advantages of Waterfall Model:

Waterfall model is the oldest and most widely used model in the field of software development. There are certain advantages of the waterfall model, which causes it to be the most widely used model as yet. Some of them can be listed as under.

a. Needless to mention, it is a linear model and of course, linear models are the most simple to be implemented.
b. The amount of resources required to implement this model is very minimal.
c. One great advantage of the waterfall model is that documentation is produced at every stage of the waterfall model development. This makes the understanding of the product designing procedure simpler.
d. After every major stage of software coding, testing is done to check the correct running of the code.

1.3 Drawbacks of Waterfall Model:

The question that must be bothering you now is that with so many advantages at hand, what could be the possible disadvantages of the waterfall model. Well, there are some disadvantages of this widely accepted model too. Let us look at a few of them.

a. Ironically, the biggest disadvantage of the waterfall model is one of its greatest advantages. You cannot go back, if the design phase has gone wrong, things can get very complicated in the implementation phase.

b. Many a times, it happens that the client is not very clear of what he exactly wants from the software. Any changes that he mentions in between may cause a lot of confusion.

c. Small changes or errors that arise in the completed software may cause a lot of problem.

d. The greatest disadvantage of the waterfall model is that until the final stage of the development cycle is complete, a working model of the software does not lie in the hands of the client. Thus, he is hardly in a position to mention if what has been designed is exactly what he had asked for.

2. Software implementation:

This software is been implemented in Java Language. IDE used is 'NetBeans 8.2.0'. Framework used here is 'Java Swing'. Instead of database, we are using datasets in unstructured form. Product reviews are been taken from website 'https://www.tesfreaks.co.in'. Here, we have used inbuilt libraries as well as added certain external libraries for Pre-processing as well as Tagging (Grammatical Classification). After tagging, this software creates a 'Feature Table' which displays Product, Its Feature and Opinion Sequence Wise. We had added certain files of Formal, Informal Text and using those files, this application classifies what are the total number of positive, negative and neutral words in dataset. This process is done through Frames. Finally, its summary is displayed and result is in form of 'Graph'.

V. SIMULATION AND RESULT

1. Prediction of total positive, negative and neutral reviews in Graphical Form:

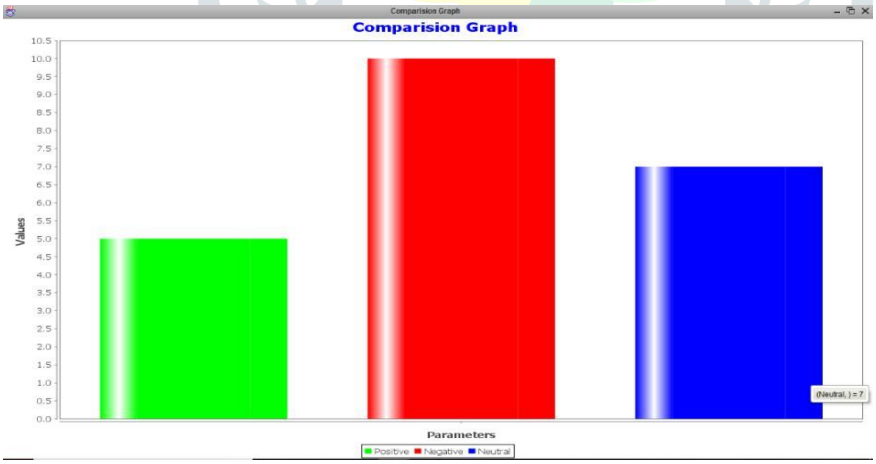

Fig 3: Graphical Analysis

Result is the 'Graph Analysis' that displays total Positive, Negative and Neutral comments. Through this graph, most of the organizations will be able to understand the statistics of their product at market (whether their product is in profit or loss.) If there are more negative opinions regarding specific product, This will also help the companies to find and fix the issue in their product and will eventually lead to best selling of their product.

2. Layout of Desktop Application:

Fig 4 : Desktop Application

Desktop application can be installed in your system This application can be used both online as well as offline. For online usage, you just have to click on given product in application and it will fetch reviews from website and load it into scroll box automatically, then you can find it's review type through application processing.

For offline usage, you must have a dataset (.csv) file that contains reviews regarding specific product. File can be in both structured and untrusted format. Once you load that file in application, then software will do its job.

3. Sequence Diagram:

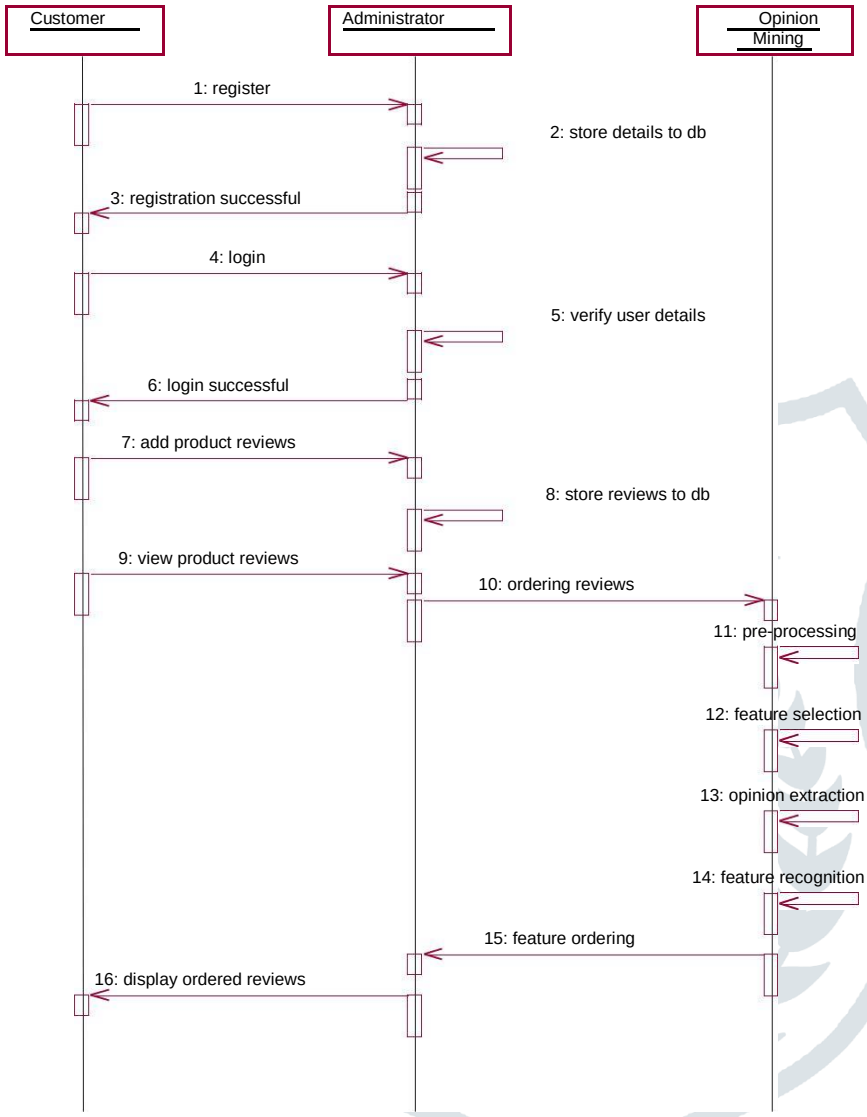

Figure 5: Sequence Diagram for FEROM

VI. CONCLUSION

We proposed an enhanced method of feature extraction and refinement for opinion mining, to analyse product review data. It extracts candidate features considering the syntactic and semantic similarities between them and reduces the number of features by merging words with similar meanings. This showed satisfactory performance results through a series of experiments conducted on real product review data. Furthermore, this system showed good performance in a virtual opinion mining framework. Based on these observations, we claim that this system is a proper method for opinion mining by employing an enhanced scheme of feature extraction and refinement to analyse customer review data.

REFERENCES

[1]. Mr. R.Nithish, S.Sabarish, M.Navaneeth Kishen , Mr. A.M.Abirami , Dr. A.Askarunisa "An Ontology based Sentiment Analysis for mobile products using tweets (2013 Fifth International Conference on Advanced Computing (ICoAC))" 16 October 2014, © IEEE Explore, INSPEC No**:** 14684544.

[2]. Ms. Ashwini Rao, Dr. Ketan Shah " An Optimized Rule based approach to extract relevant Features for Sentiment Mining (2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom))" 31 October 2016, © IEEE Explore, INSPEC No**:** 16426442.

[3]. Ms. Sujata L. Sonawane, Ms. Pallavi V. Kulkarni "Extracting Sentiments from Reviews: A Lexicon-Based Approach (2017 1st International Conference on Intelligent Systems and Information Management (ICISIM))" 01 December 2017, © IEEE Explore, INSPEC No: 17411037.

[4]. Mr. Zhibin Zhang, Hong Li Mr.Wendong Yu "Fine-grained Opinion Mining : An Application of Online Review Analysis in the Express Industry (2017 3rd IEEE International Conference on Computer and Communications(ICCC))" 26 March 2018, © IEEE Explore, INSPEC No: 17652248.

[5]. Ms. Neelam Duhan, Ms. Divya, Ms. Mamta Mittal "Opinion Mining using Ontological Spam Detection (2017 International Conference on Infocom Technologies and Unmanned Systems(Trends and Future Directions) (ICTUS))" 08 February 2018, © IEEE Explore, INSPEC No**:** 17575664.

[6]. Ms. Zulva Fachrina, Mr. Dwi H. Widyantoro "Aspect-Sentiment Classification in Opinion Mining using the Combination of Rule-Based and Machine Learning (2017 International Conference on Data and Software Engineering (ICoDSE))" 12 February 2018, © IEEE Explore, INSPEC No**:** 17578375.

[7]. Mr. Yu Mon Aye, Mr. Sint Sing Aung "Enhanced Sentiment Classification for Informal Myanmar Text of Restaurant Reviews (2018 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA))" 01 October 2018:, © IEEE Explore, INSPEC No: 18132416.

[8]. Mr. Anil Singh Parihar, Ms. Bhagyanidhi "A Study on Sentiment Analysis of Product Reviews (2018 International Conference on Soft-computing and Network Security (ICSNS))" 13 Deccember 2018, © IEEE Explore, INSPEC No: 18310228.

[9]. Mr. Sushant Kokate, Mr. Bharat Tikde "Fake Review and Brand Spam Detection using J48 Classifier (International Journal of Computer Science and Information Technologies (IJCSIT))" Date, Vol 6(4), Issue 4, ISSN: 0975-9646.

[10]. Warrant Songpan "The Analysis and Prediction of Customer Review Rating Using Opinion Mining (2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications)" 03 July 2017, © IEEE Explore, INSPEC No: 17010009.