GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES
A SCORE BASED SENTIMENT ANALYSIS OF ONLINE FOOD REVIEWS

Sasikala P[*1] **& L.Mary Immaculate Sheela**[2]
[*1]Research Scholar, Department of Computer Science, Mother Teresa Women's University, India
[2]Supervisor, Dilla University, Ethiopia, Africa

ABSTRACT

The growth of web contributes a huge quantity of user created content such as customer feedback, opinions and reviews. Sentiment analysis in web embraces the problem of aggregating data in the web and extraction about opinions. Studying the opinions of customers helps to determine the people feeling about a product and how it is received in the market. Various commercial tools are available for sentiment analysis. In this paper, we propose a system which classifies the reviews on a scale of 1 to 5 based on the sentiments in the words. The groups of words used to make a decision to rate the reviews are displayed as word cloud.

Keywords- *Customer reviews, Machine Learning, Opinion mining, Natural Language Processing, Sentiment Analysis.*

I.   INTRODUCTION

"What others think?" is always important information in a decision-making process. Every day people discuss various products on social media sites. Web and its associated distribution services provides information services such as online services where data objects are linked together to facilitate interactive access. Web pages may not have a predefined schema or pattern and it is difficult for computers to understand the semantic meaning. Companies want a piece of that pie to determine how their audience communicates to find the important information that drives business. Sentiment analysis is the robotic mining of opinions and feelings from content through Natural Language Processing (NLP). Sentiment analysis is nothing but categorizing opinions in the given content or documents into "positive" or "negative" or "neutral"[1].

Sentiment analysis can be processed in three levels: aspect-level, document-level, and sentence-level. Sentiment analysis in document level considers the entire document as a single topic and classifies positive or negative sentiment. The sentiment expressed in each and every sentence is classified in sentence level. Since sentences are part of the documents does not make much difference between sentence and document level. To get the detail opinion or sentiments have to process the document to aspect level. Aspect level sentiment analysis is to determine the features of the sentiment conveyed towards each aspect and the given target entities. This proposed model handles sentiment polarity classification which is a fundamental problem of sentiment classification.The online data have several drawbacks to hinder the sentiment analysis task. The first defect is anybody can post their own contents and quality or impact of their comment is not assured. Online spammers may post their fake opinions [4]. The next fault is the polarity of reviews cannot be ascertained or unavailable. The dataset used in this paper is around 3000 food reviews collected from amazon. Each post in amazon is inspected and verified by the company before it gets posted. Each review has a rating scale from 1 to 5 stars which can be used to identify the sentiment polarity.

II.   BACKGROUND AND LITERATURE SURVEY

Sentiment analysis techniques are machine learning, lexicon based and hybrid techniques. Machine learning techniques are implemented in supervised classification. Lexicon based approach depends on the collection of opinion terms. Combined approach of lexicon and machine learning is hybrid approach. The computational techniques involving languages are called as natural language processing. Most of the existing researches focus on the mining of data available online in web. Before World Wide Web, information or opinion about the products is collected based on surveys manually.

Xing et al. [8] had proposed a work on product reviews collected from amazon to identify the negation phrases. Sentence level and review level classification of data is performed for the data collected from February to April 2014. Aashutosh Bhatt et al. [3] used reviews of iPhone 5 extracted from Amazon website and suggested a rule based extraction of product feature sentiment analysis. POS technique is implemented to each and every sentence level and the results are shown in charts. Ahmad Kamal [2] used supervised and rule based techniques to mine the opinions from online product reviews.

III. RESEARCHDESIGNANDMETHODOLOGY

Sentiment classification is to select and extract the text features. Feature selection in sentiment analysis is collecting the information from reviews in web and performing the following steps.

DataPreparation: The data preparation step will pre-process the data and removes all the non-textual information and tags. Data pre-processing performs cleaning of data by removing the information like review date and name of the reviewer which is not required for sentiment analysis.

ReviewAnalysis: finding parts of speech (POS) adjectives and counting the presence and frequency.

SentimentClassification*:* Classifies the extracted words as positive or negative.
The architectural view of the system is given in fig.1.

The reviews are extracted from the amazon website and stored in SQLite database. An organized machine learning technique is planned to group the subjectivity words in the review sentence. The subjectivity detection is implemented using part-of-speech (POS) tagger [6]. The POS tagger is used to filter the word which does not have any sentiments.

A collection of nouns (dog, restaurant, cat, etc), verbs (is, are, would, etc) and adjectives (hot, beautiful, fragrant, etc) are used to decide the subjectivity of the sentence. Frequently used features are identified and stored as vector of text. These features are processed and converted to lower case alphabets. Further all the punctuations and stop words are removed.
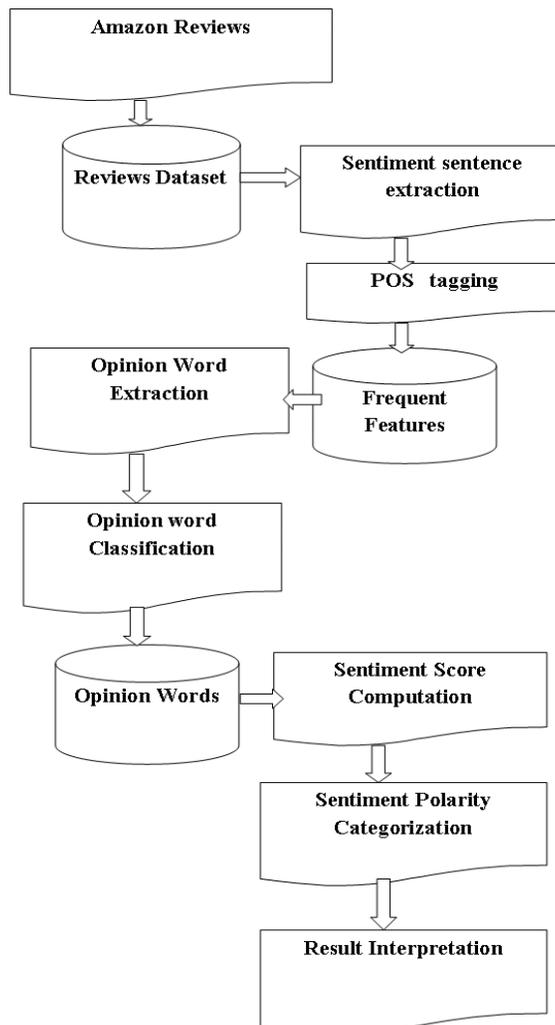
Fig .1 Architecture of the proposed system.

The word frequency matrix is created in the next step. Opinion words with sentiment polarity as positive, negative, neutral are extracted and a score value is provided. Opinion words are mined as summary and displayed as word cloud. The best reviews are those with score of 5 stars and 4 stars, average reviews are those with score of 3 stars and worst reviews are those with score of 2 stars and 1 star.

1. Procedure **ScoreOrientation** (score *t, s*ummary s_i)
2. begin
3. for each opinion word *op* in summary s_i
4. if (*t* > 3)
5. s_i orientation = positive
6. else
7. s_i orientation = negative
8. endfor;
9. end

Fig .2 Prediction of score orientation.

The above mentioned procedure in Fig.2 is used to compute the scores and sentiment polarity. The results are displayed as word clouds with scores rating from 1 to 5. R is a popular tool for statistical analysis using data frames [7]. It is an open source tool and suitable for various operating systems.

IV. EXPERIMENTAL SETUP AND RESULT EVALUATION

In this division we assess some of our design preferences explained in the preceding sections. The dataset consists of 50000 fine food reviews posted by more than 200,000 customers about 70,000 and over food products[5]. For the experimental usage latest 3000 fine food reviews is used and implemented.

Sentiment Computation using scores
The rating of the products is in five star scales shown in fig.3.The score value 1 indicates very negative, score value 2 is negative, 3 is neutral, 4 is positive and 5 is very positive.
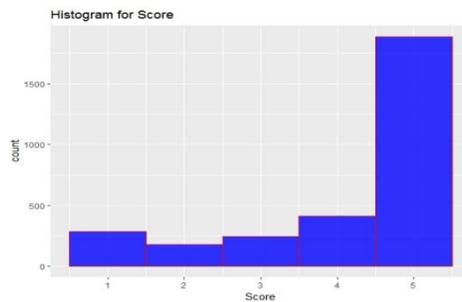
Fig.3 Histogram of score

Table 1. Summary of User Provided Score

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	4.000	5.000	4.146	5.000	5.000

The above table depicts the summary of user score mean value is 4.146.
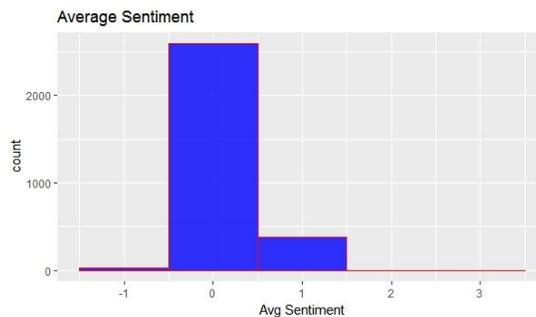
Fig.4 Histogram of average sentiment

Table2.Summary of Average Sentiment

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.06600	0.04845	0.20960	0.21800	0.37710	2.60300

The histogram of average sentiment is represented in fig. 2. The top 3000 fine food reviews are used to find the average sentiment. The average mean is 0.21800 shown in table 2. There is a huge difference between score and average sentiment.

Table3. Top Best Reviews by Sentiment Score

S.no	Summary
1	*"I liked getting this one for my twin 10-month olds because it has such interesting ingredients, particularly the zucchini and the garbanzo beans. I like the fact that it's vegetarian yet I believe it has a complete protein since it also has brown rice. This list of ingredients looks excellent and it's a green color naturally. The babies liked it. They have liked everything that has carrots. I have the subscription for this one as I really like the ingredients and the fact that I'm giving the babies food that is naturally green. Ingredients per the label: Water, organic corn, organic zucchini, organic green beans, organic brown rice, organic carrots, organic garbanzo beans, organic canola oil. This has 70 calories, 8% protein, 40% vitamin A, 10% vitamin C, 2% calcium and 2% iron. I've been informed that the percentage of daily needs met is based on an adult."*
2	*"This is the very best marinade you can buy as far as I am concerned---of course, everyone to his or her own tastes, but all in all, I think anyone will like it---I give it five stars..."*
3	*"My partner is very happy with the tea, and is feeling much better since starting to drink it. She has been drinking it both hot (normal) and iced (chilled) and likes the refreshing nature of it."*
4	*"I ordered this for a friend at South Carolina for Christmas, i want her to try some Italian chocolates, lol! The item was arrived earlier than expected, but, nevertheless, it was well packed and only melts on her mouth, she totally loves it. Three different taste and flavors of Ferreros guarantees to give three different yummy joy, Mmmmm...*
5	*"Excellent shipment and product as well , looking forward to purchase again , thanks Excellent shipment and product as well , looking forward to purchase again , thanks Excellent shipment and product as well , looking forward to purchase again , thanks"*

Top best review sentiment score is listed in the above table. The text in the summary field is best reviews fetched from the dataset. To extract the opinion word, the above text has to be pre-processed; punctuation, special symbols like <br> and stop words have to be removed. Top worst review comment is presented in the below Table 4. *"Being a salt-free product is why I purchased this, but the chips are quite greasy."* The word "greasy" is used as opinion word to decide the sentiment polarity. The sentiment or opinion of each of every term in the summary is analysed.

11

Table 4. Top Wosrt Review Comments by Sentiment Score

S.no	Summary
1	*"Being a salt-free product is why I purchased this, but the chips are quite greasy."*
2	*"My husband and I were very disappointed in this coffee, very weak, watery cup of coffee. A definite waste of $13.00."*
3	*"Here is a nutritional breakdown: Calories: 100 Calories from Fat: 30 Total Fat: 3g Saturated Fat, Trans Fat, Poly and Mono: 0g Cholesterol: 0mg Sodium: 160mg - 7% Potassium 180mg - 5% Total Carbohydrates 15g - 5% Dietary Fiber 1g - 4% Sugars 2g - 4% Protein 1g Calcium 2% Iron 2% For Weightwatcher folks this breaksdown to 2 Points......."*
4	*"If you are ordering cookies, beware, the 'cookies' arrived in crumbles, so, if you want a box of stale crumbs order away :)"*
5	*"No doubt about it this is just the right combination of Lobster, scampi, prawns with a hint of Dry white wine & a dash of brandy finish this soup to perfection. Ingredients Water, Lobster (4%), Cod (3.5%), Scampi (2.5%), Concentrated Tomato Paste, Modified Cornflour, White Wine, Prawns (1.5%), Skimmed Milk Powder, Butterfat, Double Cream, Salt, Yeast Extract, Sugar, Shrimp Powder, Fish Powder, Vegetable Oil, Vegetable Extracts, Stabiliser (Polyphosphates), Brandy, Concentrated Lemon Juice, Spices, Herb and Spice Extracts with celery. *No Artificial Colours *No Artificial Flavours *No Artificial Preservatives Information Gluten Free Contains: MILK, FISH, SHELLFISH, CELERY."*

Word Clouds

The visualization of opinion words and its counts with sizes relative to their counts is displayed in Fig.5. This word cloud is created by R software using text mining packages. Frequency matrix is constructed before generating the word cloud.

Fig. 5 Word Cloud of Sentiment words

The word cloud shows the scores in a 5 point scale ranging from very negative to very positive.

Fig.6 Word clouds of scores

Sentiment Score Prediction
This prediction model focuses on sentiment polarity like positive or negative instead of scores. Predictive techniques like Naïve Bayes, Regression can be used to test the data. The below table is the representation of split data into train and test set.

Table 5.Test and Training of Data Set using Scores

	Score	Summary
0	5	Good Quality Dog Food
1	1	Not as Advertised
2	4	"Delight" says it all
3	2	Cough Medicine
4	5	Great taffy
5	4	Nice Taffy
6	5	Great! Just as good as the expensive brands!
7	5	Wonderful, tasty taffy

8	5	Yay Barley
9	5	Healthy Dog Food
10	5	The Best Hot Sauce in the World

The score mentioned in the second field in Table 5 is converted into positive or negative using score value. All the values above score 3 are predicted as positive and the values less than 3 are predicted as negative. We can see the result in Table 6.

Table 6. Test and Training of Data Set Using Sentiment Polarity

	Score	**Summary**
0	Positive	Good Quality Dog Food
1	Negative	Not as Advertised
2	Positive	"Delight" says it all
3	Negative	Cough Medicine
4	Positive	Great taffy
5	Positive	Nice Taffy
6	Positive	Great! Just as good as the expensive brands!
7	Positive	Wonderful, tasty taffy
8	Positive	Yay Barley
9	Positive	Healthy Dog Food
10	Positive	The Best Hot Sauce in the World

After performing stemming and pruning of the data, further machine learning algorithms can be applied to predict the expected results.

V.   CONCLUSION AND FUTURE WORK

Sentiment analysis is the process of identifying the feeling expressed in the text or document. We proposed a methodology for mining the food reviews based on score combined with existing text analysing packages. The proposed system has produced a very good result using the score ratings. The limitation of this system is, it works better only for the open sentiments like rating or scores. The results were not promising for hidden sentiments. In Future work, prediction based methods will be implemented with existing approach. More features will be extracted to handle the implicit sentiment analysis.

REFERENCES
1. *Anjali Ganesh Jivani, "A Comparative Study of Stemming Algorithms", International. Journal. Computer. Technology. Applications., Vol 2 (6), 1930-38, ISSN:2229-6093.*
2. *Ahmad Kamal , "Subjectivity Classification using Machine Learning Techniques for Mining Feature - Opinion Pairs from Web Opinion Sources", International Journal of Computer Science Issues (IJCSI), Volume 10 Issue 5, 2013, pp 191-200.*
3. *Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande, "Amazon Review Classification and Sentiment Analysis ",IJCSIT, Vol. 6 (6) , 2015, 5107-5110.*

4. Liu B (2014) ,"The science of detecting fake reviews", http://content26.com/blog/bing-liu-the-science-of-detecting-fake-reviews/.

5. J. McAuley and J. Leskovec, "From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews." ,WWW, 2013.

6. Roth D, Zelenko D (1998), "Part of speech tagging using a network of linear separators" In: Coling-Acl, The 17th International Conference on Computational Linguistics, 1136–1142.

7. Tierney, L. (2005), "Some notes on the past and future of Lisp-Stat", Journal of Statistical Software, 13 (9), 1–15.

8. Xing Fang and Justin Zhan , "Sentiment analysis using product review data ", Journal of Big Data 2015 : DOI: 10.1186/s40537-015-0015-2© Fang and Zhan; licensee Springer. 2015.