International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 17th Sept 2021. Link :http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-09

DOI: 10.48047/IJIEMR/V10/I09/17

Title **BUYER RATING AND SENTIMENT ANALYSIS USING DISCRIMINATORY SUPPORT VECTOR MACHINE (SVM)**

Volume 10, Issue 09, Pages: 121-134

Paper Authors
P. G. V. Suresh Kumar, Haftom Gebregziabher, Sisay Wayu

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

BUYER RATING AND SENTIMENT ANALYSIS USING DISCRIMINATORY SUPPORT VECTOR MACHINE (SVM)

P. G. V. Suresh Kumar[1], Haftom Gebregziabher[2], Sisay Wayu[3]
[1]Professor, Department of ICT, Ethiopian Technical University, Addis Ababa, Ethiopia
E-mail: pendemsuresh@gmail.com
[2] Deputy Director for Academics, Ethiopian Technical University, Addis Ababa, Ethiopia
E-mail: habtilo@gmail.com
[3]Head of Department, Ethiopian Technical University, Addis Ababa, Ethiopia
E-mail: sisw24@gmail.com

Abstract: This substantial issue is increasingly important in business and culture. It presents many challenging research scenarios but guarantees a relevant insight for everybody interested in view evaluation and social networking analysis. This paper's key aim is to detect sentiment polarity such as positive, negative, and emoji representation with customer feedback on various products. Opinion mining from e-commerce sites has a significant part in making purchase decisions and founders to boost their product and marketing strategies. But, it becomes very difficult for the clients to understand and assess the product's actual view manually. Because of this, we need an automated way. The majority of the researchers used machine learning algorithms to do an automated representation of phrase embedding. Among the popular techniques in machine learning has been used the support vector machine (SVM). The weighted support vector machine (WSVM) is the improved version for the standard SVM to grow the outlier sensitivity issue. In this paper, the word2Vec version uses to extract the attributes from customer reviews in WSVM based on opinion analysis of product reviews in E-commerce websites. The experiment result shows that the suggested WSVM can works better on the opinion classification job doing any version applied.

Keywords: Sentiment analysis,Opinion mining, feature extraction, Weighted support vector machine.

I. INTRODUCTION

Opinion, expressions and associated concepts such as evaluation, attitude and emotion, affect emotion and mood; they refer to our own emotions and beliefs. They are of value to human psychology and are the main influences in our behaviour. Our beliefs and perceptions of facts and the choices we make are largely conditioned by how others view and perceive sand. For this reason, our views on the square are largely driven by others' opinions, and whenever we need to make a decision, we often try to find out the opinions of others. It is no longer the most practical for people, but also companies. From a software point of view, we naturally need to map human evaluations and feelings toward any hobby-based anxiety, the emotion assessment project. More precisely, sentiment evaluation, also known as opinion mining, is an examination system that seeks to extract criticism and feelings from the textual content of natural language through computational methods [1].

The initiation of opinion analysis and rapid expansion coincides with investigations of social networks on the world wide web, including reviews and talks in discussion forums, blogs, and microblogs, because for the first time in human history, we've got a huge volume of view information recorded in the electronic bureaucracy. This caused opinion analysis or opinion mining issues because this info is filled with opinion. Having these records filled with evaluations isn't always surprising since the number one reason people post messages on social networking systems is to ascertain their perspectives and evaluations, so the sentiment score is in the center of a social networking analysis. Since the early 2000s, opinion analysis has grown into one of the most important research areas in processing herbal language. It has also been extensively studied in data mining, net mining, and document retrieval. Studies have been developed from technological understanding of computers to control technological knowledge and socio-technological knowledge on account of their significance to the business firm and society. In recent decades, the industrial game surrounding belief analysis has also thrived. Sentiment evaluation structures noted programs in just about any industrial venture, fitness, government, and societal sphere [2].

On e-commerce websites, buyers often face trade-offs in buying decisions. Online product/service reviews act as sources of documents connected with products/sellers. Meanwhile, the modern generation has brought with them a wealth of the content material, which makes it prohibitively costly to exhaust all available records (if that is possible in any case). Consumers must specify a subset of facts which are applied. Star ratings are unusual keys for this choice, as they provide a succinct indication of the tone of this evaluation. Sentiment analysis, textual content analysis methods that routinely identify polarity in a text, can assist in those situations with a more accurate appraisal. With this note, we examine the consequences of stargazing sentiment in 3 odd domains to discover this technique guarantee [3].

A) Social Network Analysis

Social network analysis is a method specially developed by sociologists and social psychology researchers. Social network analysis considers social relationships in terms of the network concept. Simultaneously, the individual actor appears as a knot, and the courtship between each node is presented as an aspect. The social community evaluation was defined in [4] as an assumption of the importance of relationships between interacting units, and family members identified through interconnections between devices are an essential component of network theories. . The analysis of social society has emerged as a major focus of avant-garde sociology. Also, it has received extensive follow-up in anthropology, biology, communication studies, economics, geography, data technology, organizational research, social psychology, and sociolinguistics1. In 1954, Barnes [5] began to use the period conclusively. Systematically to indicate patterns. From relationships traditionally include concepts. After that, many students expanded the use of systematic analysis of social networks. Due to the rise of online social media sites, online social media evaluation will become a hot topic of study these days.

II. LITERATURE REVEW

The aims of this research method to extract reviews and opinions from the customer reviews. Opinion mining is also known as sentiment analysis. It has been one of the most explored areas in data science in recent years. Despite great research, the answers cited and the regulations in place now did not satisfy the discontinued customer's opinions. The main problem is the new conceptual algorithms that dictate emotions. There is much unique evidence (perhaps countless) that can turn those thoughts into verbal conversations with people. Some of the existed works for the sentiment analysis from the customer reviews are given below

Abdalgader et al. (2020) As a result of its principal role in setting the general semantic management of natural language expressions, it's one of the most troublesome issues facing these research areas. This report introduces a special program for the dictionary-based word willpower polarity method over several sets of people's opinion information. In this guide, they used a variation of the entire linguistic word polarity devotion method, which calculated that the semantic connection between the contextual growth organization of the target phrase and the institutional synonym expansion composed of the synonyms of all of the sentences surrounding the target from the person. Text content. The polarity of this self-declaration is determined when the semantic courtship involving these very important agencies is final.

Compared to strategies for specifying multipolar sentences depending on the most lexicon, the strategy used is no longer based on estimating marital relationship within a period score. Additionally, it depends upon the size of the semantic relationship within the duration. It allowed for significant research and application of degree in semantic and abstract data. Likewise, you can enhance normal performance by integrating an initial step. The relative negation assortment of words within the management's exact textual content component changes even while ascertaining its psychological orientation. Presentation of the execution results performed by the entire lexical polarity dedication method factor definitely contrary to the test strategies, which can be assessed in the a variety of reference information units through separate and continuous evaluation versions.

D. Kwak et al. (2019) total digital reviews play an important role in fostering international communications between customers and influencing customer-style purchases. E-commerce giants like Amazon, Flipkart, etc., provide a platform for customers to provide their entertainment and provide real insight into the overall product performance for customers in the future. It is required to rank reviews of high quality and poor review to extract valuable information from a large pool of reviews. Sentiment analysis is a mathematical analysis of extracting personal records from the textual content. More than 4000.00 opinions were categorized into massive and vulnerable feelings through sentiment analysis in the proposed work. Naïve Bayes, Support Vector Machine (SVM), and Decision Tree were used for type notes among the outfits of the different classes. The fashions assessment is completed using the ten pleat validation.

Zhao et al. (2018) Product reviews are a treasure trove of target customers, helping them make choices. For this termination, several opinion polling strategies have been proposed, with the evaluation sentence's trajectory judged (e.g., maximal or dire) as one of the most demanding situations. Recently, in-depth knowledge has emerged as an effective way to solve emotion classification problems. The nervous community, by nature, learns an instrumentally useful illustration without human effort. However, achieving deep mastery depends greatly on the widely available education data. We direct a unique framework for the in-depth study of opinion in a product review that uses the most available scores as sensitive follow-up indicators. The framework includes steps: (1) Mastering a formidable example (domain of understanding) that captures the general distribution of emotions in a sentence by recording information; And (2) include a text content layer at the top of the critical element layer and use categorized strings for over-moderation under supervision. We explored the types of low-stage community forms for modelling evaluation statements, particularly abstracts of convolutional functions and rapid long-time period memory. To look at the proposed framework, we created a dataset containing 1.1 million poorly rated rating sentences and eleven, 754 rating sentences from Amazon. Empirical results demonstrate the effectiveness of the proposed framework and its superiority to baselines.

Santhosh Kumar et al. (2016) Opinion mining is very important in e-commerce sites, and it is beneficial to the users as well. An increasing amount of results are saved online, more people can purchase items from the web, and as a result, customer reviews or posts increase daily. Comments on messengers' websites express their feelings. Any organization, for example, web forums, speech companies, blogs, etc., can have a lot of stats. Logs are diagnosed with devices on the network, which can be useful for all manufacturers and customers. The method for finding someone's opinion about a topic, product, or nuisance is called opinion mining. It can also be defined because the extraction of automatic records through criticism expressed through the person is currently called polling to use the product in an almost small group of products. Sentiment analysis of received evaluations is described as sentiment analysis. The purpose of opinion exploration and

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

sentiment evaluation is to make the computer understand and express feelings. This panel specializes in extracting reviews from websites like Amazon, allowing the consumer to write the offer freely. It routinely pulls reviews from its website. Also, it uses the algorithm along with the Naïve Bayes classifier, logistic regression, and SentiWordNet algorithm to classify the rating as higher and weak. At the station, we use satisfactory metric parameters to classify the performance of each set of bases.

ShanshanGaoet al. (2015) The popularity of social computing and sentiment analysis has attracted increasing interest from tourism companies and academia. The analysis of the feelings of citizens and tourists plays an essential role in improving tourism. Its goal is to perceive and examine the opinions and emotions in the opinions expressed by citizens or travelers. However, a challenging project, many groups and think tanks are ramping up their net offerings to provide public access and cost-effective responses to an inconvenience. However, to the satisfaction of our knowledge, there is very little research on applying and evaluating these internet services in tourism. So in these panels, we pre-test and explore a group of 3 internet services and compare their emotion rating skills using the popular TripAdvisor stats set. The results of the experiment yielded some interesting conclusions.

Huttoet al.(2014) The inherent nature of social media content places heavy demands on factual sentiment analysis packages. Introducing VADER, a simple rule-based version of the well-known sentiment analysis, theyanalysed its effectiveness on 11 common standards for the exercise, including LIWC, ANEW, General Inquirer, SentiWordNet, Naive Bayes, and device-oriented learning techniques and algorithms. Maximum entropy, vector machine support (SVM). Using a combination of qualitative and quantitative strategies, we first compile and validate a golden list of linguistic skills (along with measures of depth of associated feelings) that can be specifically tuned to match sentiments in microblogging contexts. We then intently incorporate these lexical features into five favourable polices made up of grammatical and grammatical rules to define and emphasize emotion intensity. Interestingly, while using our full version of rule-based banalities primarily to assess tweet sentiments, we found that VADER outperforms individual human reviewers (F1 rank accuracy = 0.96 and zero; 84,

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

respectively) generalizes better in all contexts in which it does. It has another standard.

III. PROPOSED MODEL

In this paper, we suggest learning sentiment-specific word embedding model for sentiment analysis from customer reviews. The word2Vec model is used to extract the features from the customer reviews in WSVM based sentiment analysis of product reviews in E-commerce sites. The proposed model general architecture for the sentiment analysis shown in figure.1. This model divided into three main parts of pre-processing, feature extraction using Word2Vec model, and sentiment classification using WSVM algorithm.

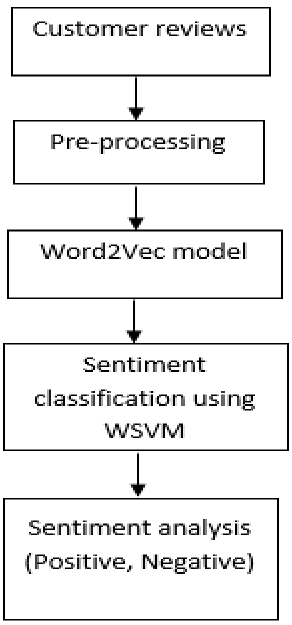

Fig.1 System architecture

A detailed description of all the steps in the proposed model flow chart is given below.

a) Pre-processing

Pre-processing is done before the start of the main form. Some actions in this stage are performed using tokenization, cleaning and case folding. In tokenization, every review is divided into smaller units known as tokens or phrases. One of the steps in preprocessing is case folding is a commitment to convert all characters in the revision text to lowercase. When, in the cleaning process, the outside alphabets are ignored, including punctuation, numbers, and HTML. In this research, we are not applying stemming and filtering operations due to its inefficiency in the previous studies in sentiment analysis.

b) Feature extraction using Word2Vec model

After the pre-processing phase is complete, we build a vector representation of the words using Word2Vec. First, the Word2Vec version creates a vocabulary of training information. Then learn and define the vector illustration for each word. There are learning algorithms in word2vec, that is. The term continuous bag of words (CBOW) and skip-gram [46]. In this analysis, CBOW is used. In CBOW, a word vector is constructed by predicting each phrase's match based on adjacent words. The resulting word vector can be contracted as writing properties. Word2Vec model can generally help improve overall class performance because, in Wor2Vec, the same phrases contain comparable vectors.

Algorithm 1:

Step1: Initialize P(pos) − num − popozitii(positive)/ num_total_propozitii

Step2: P(neg) − num − popozitii(negative)/ num_total_propozitii

Step3: Convert sentences into words

for each class of {pos, neg}:

for each word in {phrase}

P(word | class) <

num_apartii(word | class) 1|

num_cuv(class) + num_total _cuvinte

P(class) − P(class) ∗ P(word | class)

Return max ({P(pos), P(neg)}

Sentiment classificationusing WSVM:

Starting with the development of the cost function, WSVM wants to increase the division margin and decrease classification errors to complete precise generalization. When evaluating the usual SVM penalty period, in which the C-value is determined, and all academic statistic elements are handled similarly across the entire school, the WSVM weighs sentence length to minimize the influence of less important statistical factors (including outliers and noise). The confined optimization problem is formulated as follows.

$$Minimize\ \Phi(w) = \frac{1}{2}w^T w + C\sum_{i=1}^{l} W_i \mathcal{E}_i \quad (1)$$

Subject to

$$y_i(\langle w, \emptyset(x_i)\rangle + b) \geq 1 - \mathcal{E}_i, i = 1, \dots\dots, l$$
$$\mathcal{E}_i \geq 0, \quad i = 1, \dots\dots, l \quad (2)$$

Notice that we assign the weight W_i to the point from the xi records inside the above formula. Hence, the formula becomes duplex.

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3)$$

Subject to

$$\sum_{i=1}^{l} y_i \alpha_i = 0 \quad 0 \leq \alpha_i \leq CW_i \quad i = 1, \dots\dots, l \quad (4)$$

And the KT conditions of WSVM become

$$\boldsymbol{\alpha_i[y_i(\langle w, \emptyset(x_i)\rangle + b) - 1 + \mathcal{E}_i] = 0} \quad i = 1, \dots\dots, l \quad (5)$$

$$(CW_i - \boldsymbol{\alpha_i})\mathcal{E}_i = \boldsymbol{0}, i = 1, \dots\dots, l \quad (6)$$

The simplest distinction between the proposed SVM and WSVM is the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

maximum certainty of the Lagrange multiples α_i within the double problem.

Algorithm 2 steps for WSVM

$\boldsymbol{Input}: reviews\ r = \{r_1, r_2, r_3 \dots \dots r_n\},$

$\boldsymbol{Database}: Customer\ Review\ Table\ R_T$

$\boldsymbol{Output}: Positive\ reviews\ p = \{p_1, p_2, ..\},$

$Negative\ reviews\ n = \{n_1, n_2, n_3, \dots\},$

$Review\ R = \{r_1, r_2, r_3 \dots \dots r_n\}$

$\boldsymbol{Step1}: Divide\ customer\ reviews\ into\ words$

$r_i = \{w_1, w_2, w_3 \dots\}, i = 1, 2, \dots n$

$\boldsymbol{Step2}: if\ w_i \in R_T\ return + ve\ polarity\ and -$

$\boldsymbol{Step3}: Calculate\ overall\ polarity\ of\ a\ word$

$= \log(+ve\ polarity) - \log\ (-ve\ polarity)$

$\boldsymbol{Step\ 4}: Repeat\ step2\ until\ end\ of\ the\ words$

$\boldsymbol{Step\ 5}: Add\ the\ polarity\ of\ all\ words\ of\ revi$

$i.e. total\ polarity\ of\ a\ reviews$

$\boldsymbol{Step\ 6}: Based\ on\ that\ polarity, review$

$can\ be\ positive\ or\ negative$

$\boldsymbol{Step\ 7}: Repeat\ step1\ until\ R \in NULL$

IV. RESULTS AND DISCUSSIONS

A. Experimental setup

Experimental setup conducted one various product reviews on various domains such as car, book, dresses, kitchen applications, software products, hardware products, and so on. To perform operations we are using JAVA as a development tool. It is an open source object orient programming language. It uses NetBeans as an integrated development environment(IDE) tool to design e-commerce websites.

Fig.2 Add new domains

Figure2 shows to add various domain in e-commerce website such as book, cars, kitchen applications, and etc.

International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

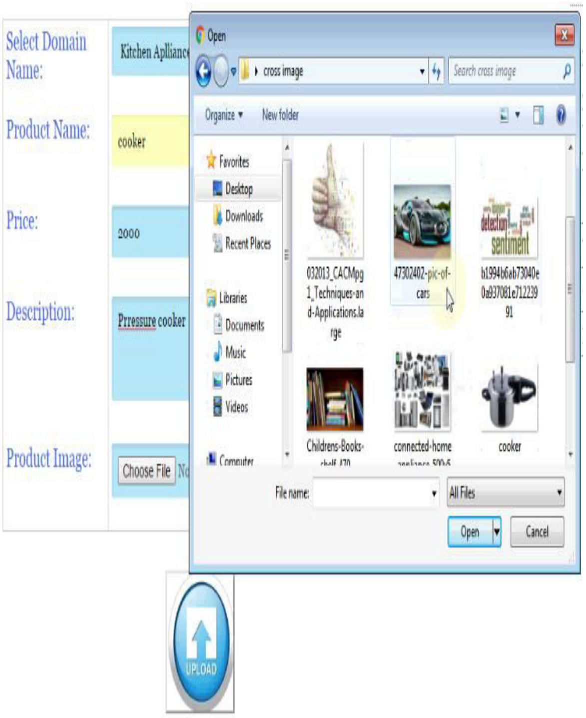

Fig.3 Adding new products in kitchen application domains

Fig4 Example products

Fig.5 User search results for the query of cooker

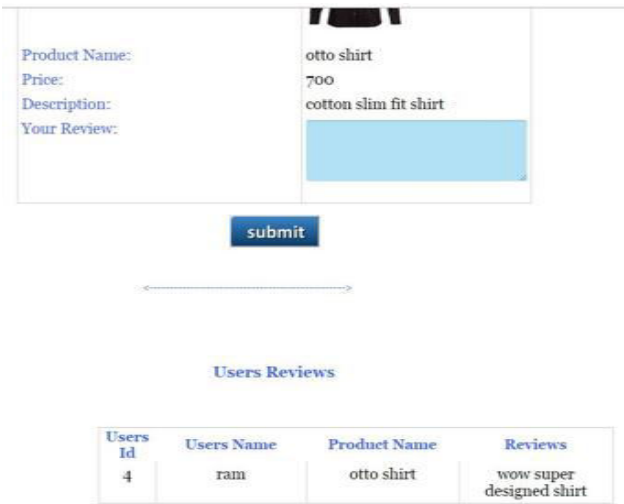

Fig.6 User review on the product

Fig.7 User providing review for cooker product

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Users Reviews

Users Id	Users Name	Product Name	Reviews
1	suresh	cooker	bad product..
3	Pavithra	cooker	worst product to cook rice
3	Pavithra	cooker	good style like royal
4	ram	cooker	Amazing product ..

Fig.8 various users reviews

Kitchen Aplliances Positive Reviews

6	fridge	*super cool freezing*
6	fridge	*super*
5	cooker	*good style like royal*
6	fridge	*amazing*

Kitchen Aplliances Negative Reviews

5	cooker	*bad product..*
5	cooker	*worst product to cook rice*

Emoji Reviews

5	cooker->	
6	fridge->	

Fig.9 Positive and negative reviews with emoji expression

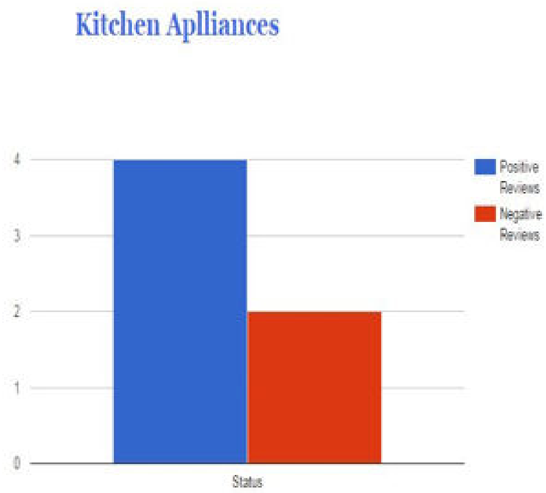

Fig.10 Positive and Negative reviews for product of kitchen applications

B. Performance measures

To evaluate and estimated proposed model performance for sentiment analysis for customer reviews on the product, we are using two types of performance metrics such as precision and recall. Here every metric have its own operation for the estimation of proposed work performance.

Precision:

A Precision is a measure of how much detailed information is given, and is the degree to which exactness is applied.

$$Precision = \frac{TP}{TP+FP} (7)$$

Recall:

A recall specifies that the system is only able to retrieve relevant features and opinions

$$Recall = \frac{TP}{TP+FN} (8)$$

C. Experimental results

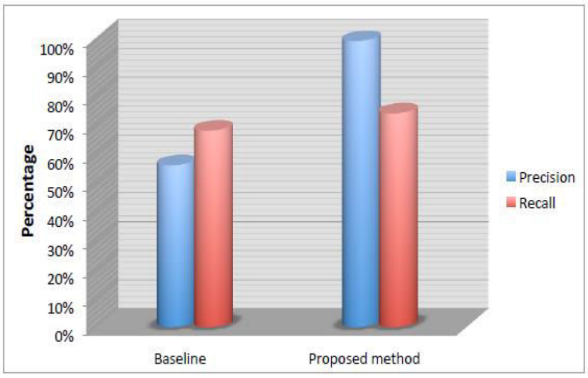

Fig.11 Feature extraction performancefor proposed method.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

As shown in the figure 11, we have taken experiment on different products. And the precision and recall calculated for both baseline model (Existed) and the proposed model. The graph shows that the proposed model provides better precision and recall rates in percentage when compared with baseline model.

Table.1 Feature extraction results for the 10 products using DR (Dependency relation)

Products	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	AVG(Mean)
Precision Baseline	56	62	62	63	65	66	69	71	71	74	68%
Precision Proposed	67	68	68	71	73	77	81	91	100	100	84%
Recall Baseline	59	61	64	70	75	79	81	81	83	81	76%
Recall Proposed	70	71	73	93	88	93	93	94	94	95	88%

Precision for Proposed Method Vs Baseline:

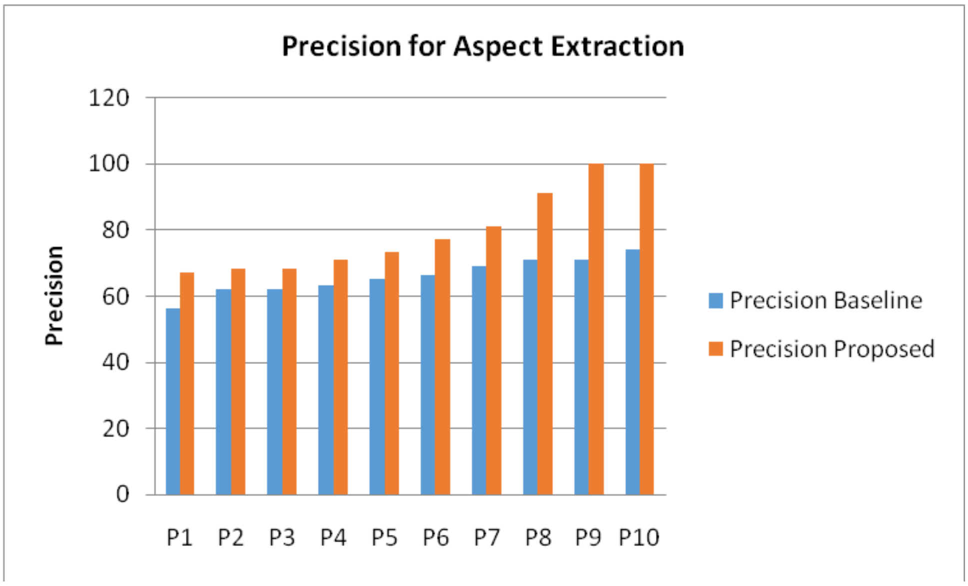

Fig.12 Precision for feature extraction of proposed method Vs Baseline

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Precision for Proposed Method Vs Baseline:

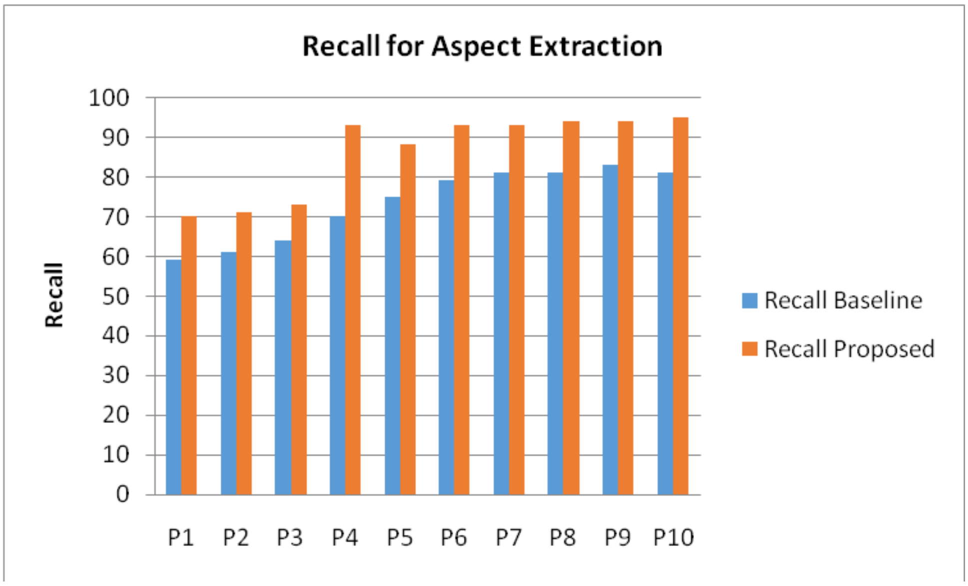

Fig.13 Recall for feature extraction of proposed method Vs Baseline

As show in the figure 12 and 13, the experimental taken on the 10 different products and performance calculated for both baseline model and proposed model. The proposed provided better precision and recall rates when compared with baseline model.

V. CONCLUSION

These studies investigated a sentiment analysis (SA) tool's ability to detect true opinions expressed in consumer reviews properly. The results of the experimental investigation of multiple SA tools, forward with ratings of products, indicate that the SA score can reflect the sentiments of an overview with an honest level of accuracy.

In this paper, the proposed sentiment level word embedding technique using machine learning-based weighted support vector machine (WSVM) increases the outlier sensitivity issue. In the proposed model Word2Vec model is used to feature extraction from the customer reviews. The experiment was conducted on 10 different customer reviews products and evaluated the performance of both the baseline model and proposed model using precision and recall measurements. With the analysis of experimental results, that proposed model provided better results.

REFERENCES

1. Dr.R.ManickaChezian and G.Angulakshmi, 2014, "An Analysis on Opinion Mining: Techniques and Tools.".
2. Stepchenkova P and A Kirilenko, 2018, "Automated sentiment analysis in tourism: Comparison of approaches", pp.1012–1025.
3. M Thelwall and R Prabowo, 2009, "Sentiment analysis: A combined approach", pp.143–157.
4. Pang B. and Lee L, 2008, "Opinion mining and sentiment analysis", in FTIR, 2(1-2):1-135.
5. B. Pang, L. Lee, and S. Vaithyanathan, 2002, "Thumbs up? Sentiment classification using machine learning techniques", pages 79-86, 2002
6. K.Abdalgader and A. AL Shibli, 2020, "Experimental Results on Customer Reviews Using Lexicon-Based Word Polarity Identification Method", IEEE Access, pp.179955-179969.
7. Kwak and S. El-Sappagh, 2019, ''Transportation sentiment analysis using word embedding and ontology-based topic modeling,'', pp. 27–42.
8. W. Zhao, and Q. Wang, 2018, ``Weakly-supervised deep embedding for product review sentiment analysis,'' IEEE Trans, pp. 185-197.
9. Santhosh Kumar K, JharnaMajumdarL, 2016, "Opinion Mining and Sentiment Analysis on Online Customer Review".
10. Hao, J and S Gao, 2015, "The Application and Comparison of Web Services for Sentiment Analysis in Tourism".
11. C. J Hutto and Gilbert, E, 2014, "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social media Text".