iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6042

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

Review Analysis with Domain Ontology Sentiment Score Measurement System within an Online Application

Reena Cheriyan[1], Nidhiya Sany[2], Angitha Thomas[3], Geethumol M S[4]
[1]*Asst. Professor,* [3, 4]*Department of computer science, DiST*
[2]*Dual Degree MCA, Depaul Institute of Science And Technology, Angamaly*

Abstract: ***The proposed system provides a sentiment score pattern of products(textile) by calculating the sentiment score using a domain ontology sentiment score measurement method by analyzing the reviews provided by the customers. The system helps to easily analyze the reviews and provide accurate emotional polarity which can represent the emotional intensity of customers on a product. The emotional score shows the increase or decrease of demand and quality of a product and therefore it helps in decision making. As the product reviews are available widely the study on these reviews make an insight about the product that is the attitude and emotion towards it. Therefore the sentiment analysis is becoming more important. There are various tools and techniques available for sentiment analysis.***

Keywords: Sentiment Measurement, Domain Ontology, Data Analysis, Feature Extraction, Data Mining.

I. INTRODUCTION

Sentiment measurement in the sense to analyze the natural language text review by determining Sentiment terms and finds whether it is positive, negative or neutral based on Domain Ontology [1]. Data mining is the core process where a number of complex and intelligent methods are applied to extract patterns from data [2]. DOSSMS is implemented using Eclipse IDE. It has 3 important phases such as, user interface, sentiment Analysis phase and output processing phase. An online textile application is used as the user interface and through this front end application customer purchases the product and provide reviews. Sentiment Analysis is a kind of text classification task and it also calculate a score using sentiment measurement algorithm. Finally in the processing phase, these sentiment scores of products will be depicted on graph patterns. According to these graphs, product quality and demand for future are evaluated and therefore decision making is done about the product improvements. Customer shows their opinion through online reviews and ratings. This paper emphasis on product review analysis using sentiment measurement technique.

II. PROPOSED SYSTEM

The main contributions of this paper include a method using Domain Ontology to Analyze and Calculate the Review Sentiment, It introduce a simple data analysis, polarity analysis and sentiment score measurement algorithm and Provide Statistical representation as well as graphical representation and it also proposes lexicon dictionary approach for sentiment analysis. Working of proposed system shows on the Fig 1.

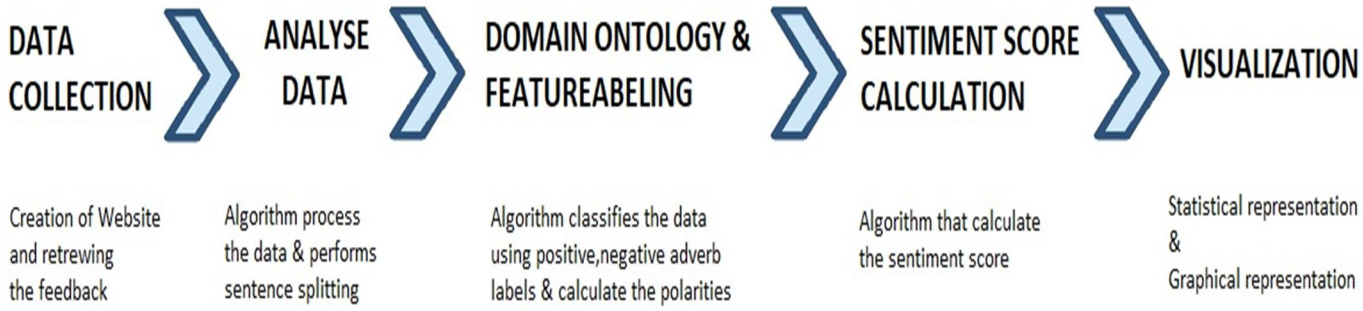


Fig. 1 working outline of Proposed System

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

III. MODEL BUILDING

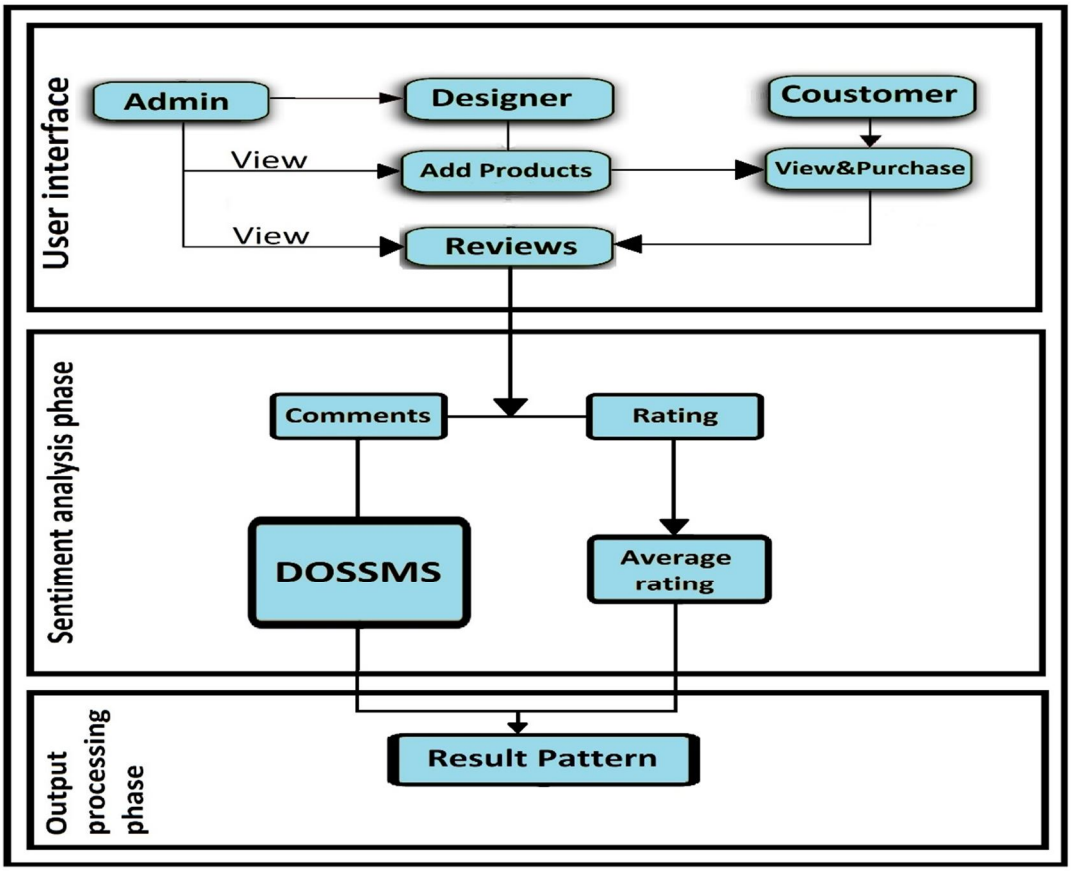


Fig. 2 System Architecture

Proposed system include 3 phases that are user interface, sentiment analysis phase, output processing phase. User Interface is an online textile application which helps to purchase different designer's products and the customer can provide reviews on the products. The sentiment analysis phase is the core of the application which include the Domain Ontology Sentiment Score measurement System. And the final phase is the output processing phase in which the result analysis is done using data patterns. Fig. 2 shows System Architecture.

A. User Interface level

It is a textile application which include the user module, designer module and admin module. Designer module can add the products and view the purchase details. User module can purchase the products then user can provide ratings and comments. Admin module deals with every other modules and further processing. The main purpose of the user interface is to collect reviews of products from the customers to analyze the comments.

B. Sentiment Analysis Phase

This phase is the core of our application where reviews are extracted, analyzed, classified and sentiment score is calculated. These are achieved by Domain Ontology Sentiment Score Measurement System(DOSSMS). Sentiment analysis is contextual mining of text which identify and extract the emotional information in the data source. Sentiment analysis is the most common text classification tool that analyzes an incoming message and tells whether the underlying sentiment is positive negative or neutral.

C. Output processing Phase

The output of the data mining are patterns which helps in analyzing the data easily [3]. The proposed system depicts the results as emotional intensity rating graph and emotional intensity DOSSM. Rating graph shows the data using the data of sentiment measurement.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

IV. DOMAIN ONTOLOGY SENTIMENT SCORE MEASUREMENT SYSTEM

It is a step by step process which result in a numerical score which indicates the product quality and demand of the product. The major steps are

1) Analyze Data (Data preprocessing).
2) Domain Ontology construction.
3) Feature Extraction and Intensity calculation.
4) Sentiment score calculation.

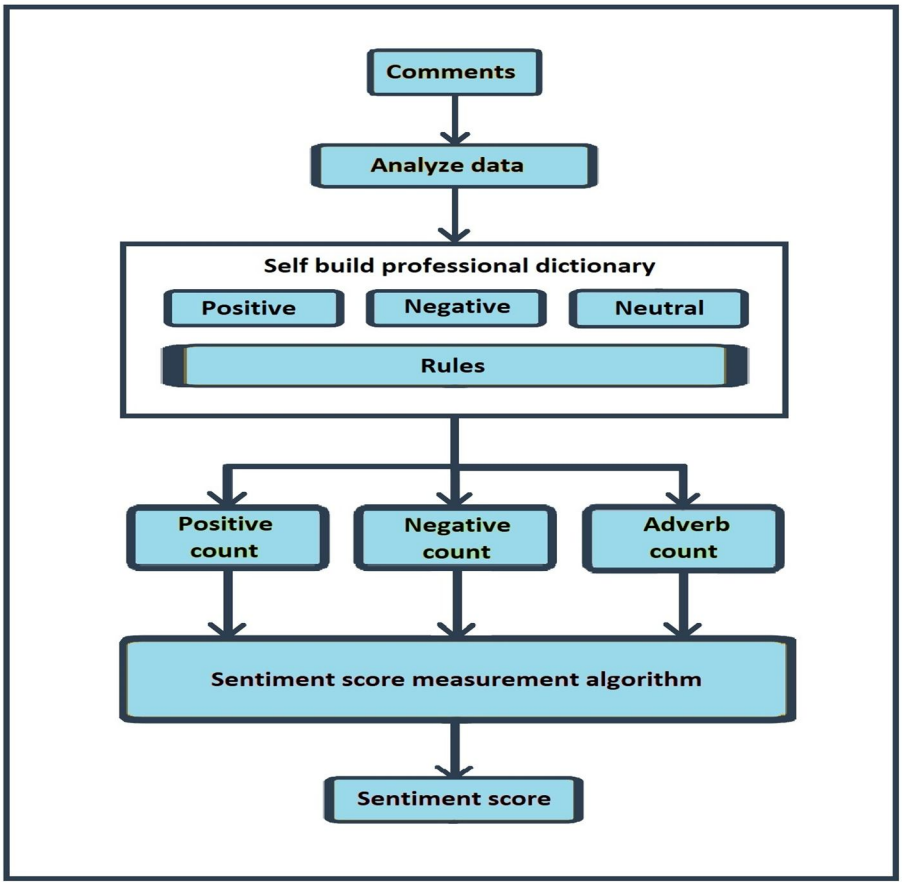


Fig.3 Domain Ontology Sentiment Score Measurement System

A. Analyze Data

Data preprocessing is an important step in the data mining process. Analyzing data that has not been carefully screened can produce misleading results. Thus the representation and quality of data is first and foremost before running analysis. Here the review data is extracted [4]. It is data preprocessing step which include transforming the row data into understandable format by eliminating incomplete, noisy and inconsistent data. The product review text need more sophisticated method to clean the noise in the raw data to perform sentiment analysis [5]. It uses sentence splitting to refine the data. It helps to make the data usefull or refined for the further processes. Each comment is extracted and splitted.It includes removal of special characters, stop word removal, word frequency counting and sentence splitting.It uses sentence splitting algorithm to refine the data. The major tasks of data preprocessing is data cleansing, data editing, data reduction and data wrangling.

Training set

$x^{(1)}, x^{(2)}, \ldots\ldots\ldots x^{(m)}$

Preprocessing (mean normalization)

$$\mu_j = 1/m \sum_{i=1}^{m} xj^{(1)}$$

replace each $x_j^{(i)}$ with $x_j - \mu_j$

B. Domain Ontology Construction.

Construction of domain ontology include the construction of dictionaries. The features words are classified using these dictionaries. These are three different sets of data that are added in the form of textual files. These three different sets are emotional dictionary, adverb dictionary, and negative dictionary. These data is used to classify each word and then sentiment contained sentences. The words which give true positive impression are included in emotional dictionary and the word which can be prefixed or postfixed to emotional words are included in adverb dictionary and negative words are included in the negative dictionary. This domain ontology helps to classify featured words using "K-mean" algorithm. Fig.4 shows an example of emotional, negative and adverb dictionaries.

Fig. 4 Example of Emotional, Negative and Adverb Dictionaries

C. Feature Extraction And Intensity Calculation.

Classification is a data mining function that assigns items in a collection to Target categories or classes. For feature extraction we use the domain ontology which contain three dictionary files. The refined data is compared with these files. Using a feature extracting and intensity calculation algorithm the weight of the total sentiment of refined data is calculated .First the word level calculation, then sentence level calculation and finally total weight of sentiments is calculated. K-mean cluster is used to classify the retrieved data set through a certain number of cluster. The number of clusters taken are three. After clustering we get the clusters of adverbs, positive and negative words [6].

In this section first each word in the sentence is extracted and then pattern matching bis performed. As a result it is classified in to adverb,negative or positive classes. After this pattern matching will be performed in sentence level. The data will take the frequency of negative, neutral or positive data and convert it as weight of sentiment of the review. The following algorithm shows the outline of this process.

Initialize sentiment word number = (w….0) as parameter
Sentiment number = (s…..0)
For x=1, 2 ……
Received instance: t(x) € sen wd
Retrieve csv files: E(w) as Emotional dictionary
A(w) as Adverb dictionary
N(w) as Negative dictionary
Pattern matching: t(x) € {E(w) or A(w) or N(w)}
Update:
Polarity (f is) = ∑ sentiment positive number
Intensity (adv) = ∑ sentiment adverb number
Polarity (neg) = ∑ sentiment negative number
Polarity (e) = polarity (f is) + intensity (adv)
W(s) = polarity (emo) – polarity (neg)

Where (senwd) is sentiment word, (f,s) is feature sentiment, (adv) is adverbs, (neg) is negative words and W(s) is weight of sentiment.

D. Sentiment Score Calculation

As the customer ratings are the data for the project there will be huge amount of data and it becomes difficult to make decisions. The purpose of this work is to calculate the emotional intensity of complex products based on the online comment text. Calculating the emotional intensity score provide more valuable insight for consumers and enterprises. Here, not only sentiment analysis but also calculation of sentiment is of done using a simple sentiment score measurement algorithm a value is calculated. Based on the polarity calculation there would be available of different polarities. The initial result will be a huge value therefore it is reduced to understand the data easily. A value ranging from zero to five is obtained as a result of this step. This sentiment score generated in this step defines the polarity of sentiment in customer reviews. If the value is above three then the polarity will be truly positive or, if the value is less than three then it is truly negative and if it is 3 then it will be neutral.

$$f(x) = \int_{ws}^{0} weight\ (s) modulus\ of\ 10$$

Where f(x) is frequency of featured word

$$T_{(SC)} = \sum_{ws}^{0} sc + freq(fw)$$

Where $T_{(sc)}$ is Total sentiment count

sc is sentiment count

$$T_{(WS)} = \int_{ws}^{0} weight\ (s)/10$$

Where $T_{(ws)}$ is weight of word score

Sentiment count is reduced to a value below 5 for

SS = sc/5

Where sc is sentiment count

ss is sentiment score

$$\text{If } g(ss) = \begin{cases} > 3 \text{ if polarity of sentiment is more positive} \\ = 3\ if\ polarity\ of\ sentiment\ is\ neutral \\ < 3 \text{ if polarity of sentiment is negative} \end{cases}$$

V. RESULT ANALYSIS

Data visualization is the graphical representation of the data and information extracted from data mining using the visual elements like graph, chart etc… Data visualization tool and techniques helps in analyzing massive amount of information and make decision on top of it [8].

1) *Data Set:* Data set is the product reviews of customers, for obtaining the data set we build an online textile shopping website. Thought this website customer can view the products and can purchase it. After the purchase they may provide the review about the product as rating or comments. Those comments and ratings are consider as the data set for our product. Figure 5 and figure 6 shows two example, graphs which shows the emotional intensity rating and emotional intensity DOSSM of each product that are calculated by data analysis, feature extraction intensity calculation and sentiment score measurement. By comparing both the graphs DOSSM graph shows 92% of accuracy. These graphs helps to analyze the sentiment over the product by the customer more easily, hence it helps in decision support.

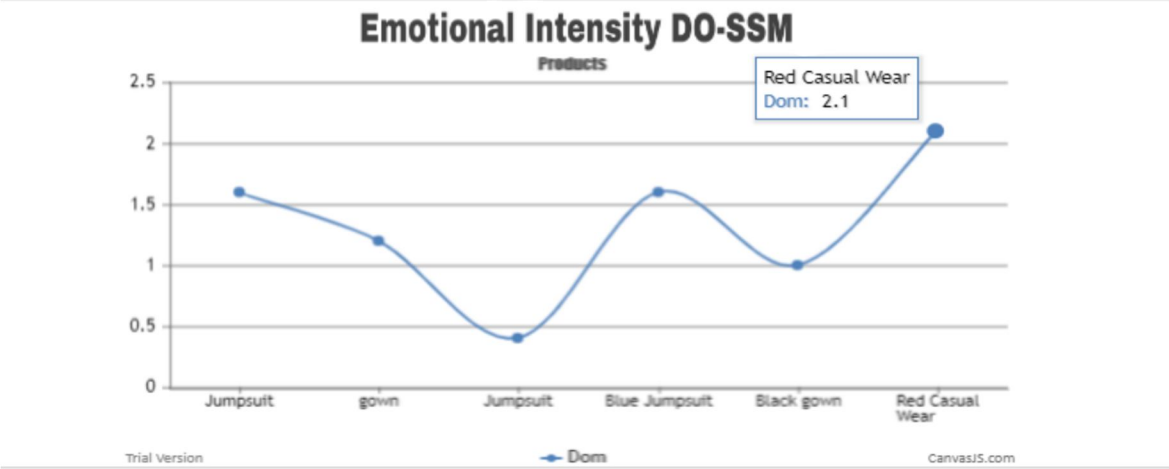


Fig. 5 Emotional Intensity DO-SSM Graph

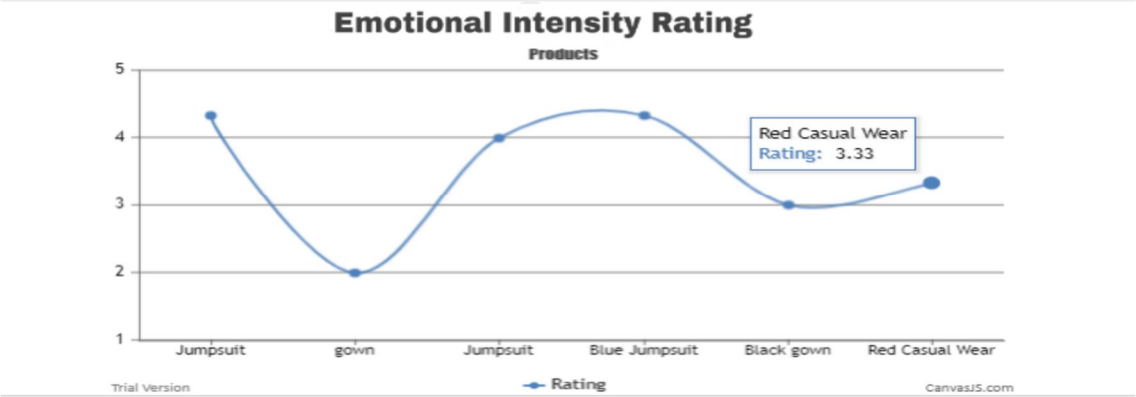


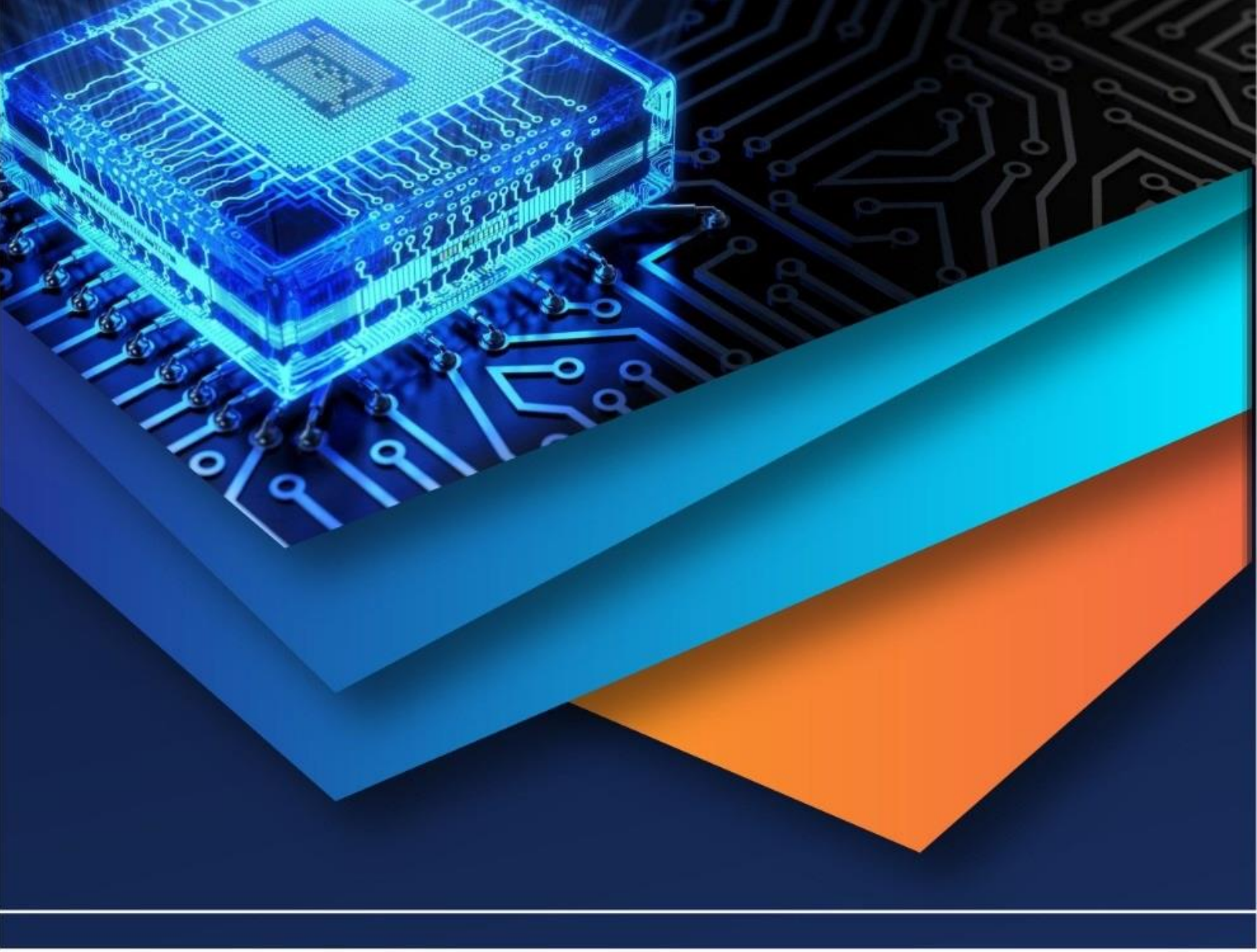
Fig. 6 Emotional Intensity Rating Graph

VI. CONCLUSION

This paper depicts the evaluation sentiment measure on reviews obtained by the customer by using DOSSM. The data used for this study are obtained from our online website product reviews of the customer. Data preprocessing is done by sentence splitting algorithm and k-mean cluster is used to classify the data into positive, negative or neutral. Features extraction and intensity calculation algorithm is used to extract the feature and convert into a numeric value to get better result. Sentiment score calculation algorithm helps in generating a value which shows the polarity of comments and it is used to analyze the quality and demand of online product. Even if there are limitations to emotion analysis, it can provide high accuracy in review assessment. The system provide decision support to user and the enterprise can use the results for upgrade the product and product sales.

REFERENCE

[1] Y. Du and L. Yang, "A Sentiment Measurement Model for Online Reviews of Complex Products", 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), Haikou, China, 2019, pp. 199-202.
[2] Jiawei Han, Micheline Kamber, Jian Pei “Data Mining: Concepts and Techniques”, Morgan kaufmann publishers, Losaltos, California 2000
[3] Alessia D’Andrea, Fernando Ferri, Patrizia Grifoni, Tiziana Guzzo, “Approaches, Tools and Applications for Sentiment Analysis Implementation”, International Journal of Computer Applications, September 2015 , pp.26-33
[4] Pyle.D, 1999,Data preparation for data mining. Morgan kaufmann publishers, Losaltos, California
[5] D. Mali, M. Abhyankar, P. Bhavarthi, K. Gaidhar, M. Bangare “Sentiment Analysis Of Product Reviews For Ecommerce Recommendation “International Journal of Management and Applied Science, 2016 , pp.127-131
[6] Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal, ” SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEW “,2017, International Research Journal of Engineering and Technology(IRJET), 2017, pp.2381-2388
[7] Georges G Grintein, “Data Mining and Data Visualization”, conference : Database Issues for Data Visualization, IEEE visualization ’95 Workshop, Atlanta, Georgia, USA, October 28, 1995



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY