International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

# E-Commerce Product Classification using Lexical Based Hybrid Feature Extraction and SVM

**V.Saravanan, Sathya Charanya.C,**

***Abstract:*** ***At present, online shopping has become a growing process, in which the profit statistics are posted by familiar e-commerce corporations like Amazon, Flipkart, Snapdeal, etc. However, this kind of online shopping unkindly omits the touch and feel of the products that can be used to estimate the product quality as the main factor while buying the commodities from the shops. The estimation of product quality is more significant during the purchasing of online products. Therefore, many opinion mining and sentiment classification methods were introduced to purchase the best products through online shopping. But, these classification methods haven't attained the effective product classification with best reviews and ratings. In this paper, we propose a hybrid feature extraction method PCA (Principle Component Analysis) and t-SNE (t-Distributed Stochastic Neighbor Embedding ) with SVM (Support Vector Machine) using lexicon-based method to classify and separate the products from the large set of different products depending on their features, best product ratings and positive reviews. In this process, the online products will be isolated and listed according to their high positive reviews. The data preprocessing is applied to the dataset to get the data accuracy before the execution of feature extraction and classification. The dimensionality reduction and best visualization of large data set are executed by applying the PCA and t-SNE method. The sentiments are also been extracted by this hybrid feature extraction method to acquire the best neighboring product ratings. The polarity of words is discovered using a lexical based approach to extract positive reviews for obtaining the best products. Finally, the SVM is exploited to the classification of products. The performance of the proposed method is estimated with precision, recall, accuracy and complexity that can provide the entire accurateness of the system.***

***Keywords:*** ***Product classification, Feature Extraction, Dimensionality reduction, PCA, t-SNE, Lexical based approach***

## I. INTRODUCTION

Online shopping can be a type of E-commerce (electronic commerce) that can permit users to purchase products from e-commerce companies through the Internet. The interesting products are searched by the customers searching through the online seller websites applying a shopping search engine that will show the available same products and cost of products at various e-retailers. Since 2016, customers using online shopping via various computers and mobile devices contain laptops, desktop computers, smartphones, and tablet computers. An extensive range of products has been sold on online e-commerce containing electronic products, clothes, and household things, etc.

**Revised Manuscript Received on November 30, 2019.**
***** Correspondence Author
**Dr.V.Saravanan*,** Dean - Computer Studies, Dr.SNS Rajalakshmi College of Arts and Science (an Autonomous Institution), Coimbatore.
**Sathya Charanya.C,** Assistant professor, Department of computer science, Salem sowdeswari college, Self-financing courses wing, Salem-



Nowadays, the companies employing effective methods to recognize the purchasing patterns of their buyers to enhance the experience of the client with rapid development in this online shopping. In the E-commerce sites, many competitions have happened in the way of to provide their products, offer discounts of products, promotions to the customer. These providing can be depending on the wide range of market research and analytics accomplished by a specialist within these E-commerce organizations. These companies have to maintain their customer reviews and product ratings on their e-commerce websites. The customer reviews not only utilized by the organization other than the customer can use these reviews to decide the quality of the product and make a decision whether to purchase or not. Thus, the recognition of reviews of the customer about products will help both the shoppers and E-commerce corporations. The E-commerce companies and individuals can make their platforms that will be utilized to compare the products according to the customer reviews and ratings using the accessibility of efficient machine learning methods and apparatus. E-Commerce sites have to compare the recommended Products to customers to make decisions on pricing at products and goods.

Thus, the decision-making process has been improved by the mining reviews and it can be valuable to the customers. In general, customer reviews mining contains automated reviews and rating extraction. Mining techniques are used to evaluating the ratings and to evaluating the customer reviews using the sentiment analysis or opinion mining. The entire score of a specific product can assist the customer to distinguish the various products depending on the customer reviews. At present, various methods were presented to mining customer reviews and presented some suggestions. In these mining methods, reviews are located from various viewpoints other than what might be needed can be a holistic scheme that can obtain multiple phases into valuable insights and it used to the customer to make decisions about the product to purchase. In this paper, a new approach is proposed with t-SNE and SVM for the product classification based on neighbor's ratings. For the review based feature extraction, the text reviews, star ratings, review score, the date of the reviews are taken in this proposed work. We use the lexicon-based approach to decide the reviews' polarity. Negative, positive and neutral reviews are also analyzed and utilized these analyzed ratings to compare the two or more products on e-commerce websites and recommend the best product and classify the product to the consumer.

## II. RELATED WORKS

A deep learning framework was presented in [1] to sentiment classification about a product review that has employed the obtainable ratings as weak supervision signals.

Two phases have been applied such as studying a high-level demonstration and a classification layer was added on top of the embedding layer. A product aspect ranking structure has been proposed in [2] to recognize the aspects of the significant product from the reviews. Three major components were utilized that is recognition of product aspect, categorization of aspect sentiment, and ranking of product aspect. A Two-Message method was presented to explain the review assessment of customers with online reviews [3]. The complementary aspect-based opinion mining has been used with asymmetric collections. The CAMEL method has attained information complementarity using both specific and ordinary aspects of crossways collections, as maintaining every equivalent reviews about products for contrastive examination [4]. The influences were analyzed amongst product involvement, product attributes, word-of-mouth, and buy purpose to medical apparatus in online shopping [5].

A semi-supervised topic-adaptive sentiment classification (TASC) model was used in [6] that have begun with a classifier constructed on ordinary features and assorted labeled data from several focuses. The hinge loss was reduced to adjust to features and unlabeled data containing topic-related sentiment words, sentiment connections and authors' sentiments. The product attributes have been examined for the process of extraction and classification from online reviews. a word-level text categorization technique was applied depending on the word vectors that have employed the semantic and grammatical features of word vectors to categorize product attribute words [7]. The SVM classification method was applied to categorize the sentiment sand texts from smartphone product consumer reviews that have recognized the several datasets employed for the categorization of texts and sentiments [8]. The Naive Bayes algorithm has been exploited in [9] for the classification of emotions of consumers and the emotion was classified using the attained features from sentence context analysis and emotion lexicons. The sentiment challenges were examined that can be a domain-dependence. Furthermore, the negation issue has become accepted in every kind of review prearranged just varied in explicit or implicit meaning [10].

## III. PROPOSED METHODOLOGY

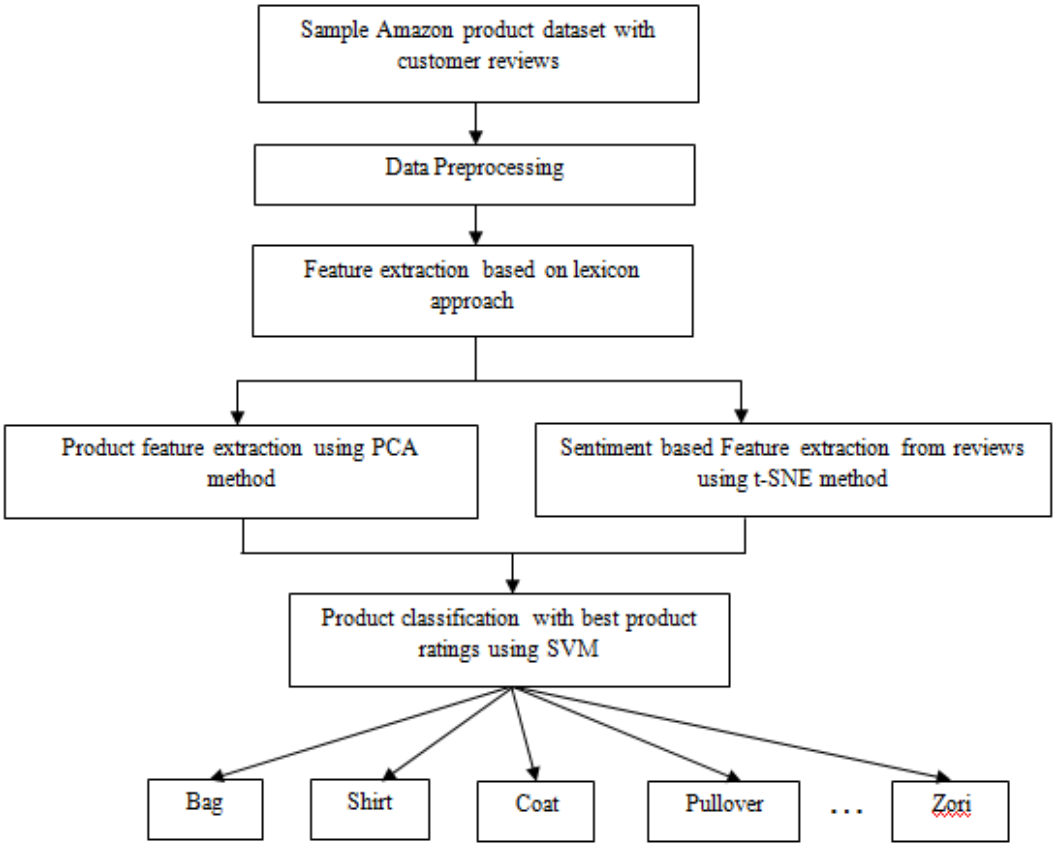


**Figure 1 System Architecture**

In this paper, three key steps are suggested to online product classification with the best ratings. The Amazon sample product data set is taken for this work. This data set contains ten fashion products like jersey, trousers, pullover, Kurtis, coat, zori, shirt, sneaker, bag, ankle_boot. Figure 1 shows the system architecture of the proposed work. In the first step, the data preprocessing is performed to enhance data quality. In the second step, feature extraction is computed based on a lexicon-based approach using PCA and t-SNE methods to obtain valuable features of products and reviews. The PCA method will be exploited to extract the product features and the t-SNE can be used to extract the sentiment-based features for obtaining of efficient neighbor's product ratings. The t-SNE is also being employed to get the best data visualization with low dimensions. Finally, we classify the products using the SVM classifier with the help of extracted product features and neighbors' ratings. The proposed method contains the following phases:

- Data Preprocessing
- Product Feature extraction using PCA

- Sentiment based feature extraction using t-SNE
- Product classification using SVM with best ratings

**1. Data Preprocessing**

In this method, the given sample Amazon product dataset can be preprocessed for further processes such as feature extraction and product classification. The sample product dataset contains details of obtainable fashion products and their reviews. Here, three steps are done to improve the data quality such as irrelevant data removal, text cleaning and stemming.

***Irrelevant Data Removal***

The irrelevant and unnecessary data will be removed from the sample Amazon product dataset. Only the product data are taken for this work. So, if there is any irrelevant data then it should be removed from the given dataset automatically using filtering methods.

***Text cleaning***

In this process, the special characters, questions, URLs and repeated words are removed from the customer review database and the uppercase letters are converted into lowercase ones that will be used to find the sentiment features easily.

***Stemming***

This stemming method can be acted as a heuristic technique and it will eliminate the affixes to obtain separated word form. The noun will be in the form of singular or plural by applying the –s or –es suffix. Here, the adjectives can be superlative shape employing –est suffix or relative structure exploiting the –er suffix. The verb will be in the form of present or past participle employing the –ing and –ed correspondingly. Finally, the given dataset has been preprocessed to obtain data quality.

**2. Product Feature extraction using PCA**

The product features are extracted from the given Amazon product data set by applying the PCA method after the execution of data preprocessing. It can reduce the high dimensionality of product data. It can be acted as an eigenvector or eigenvalue-based method that will be applied in multivariate data's dimensionality reduction.

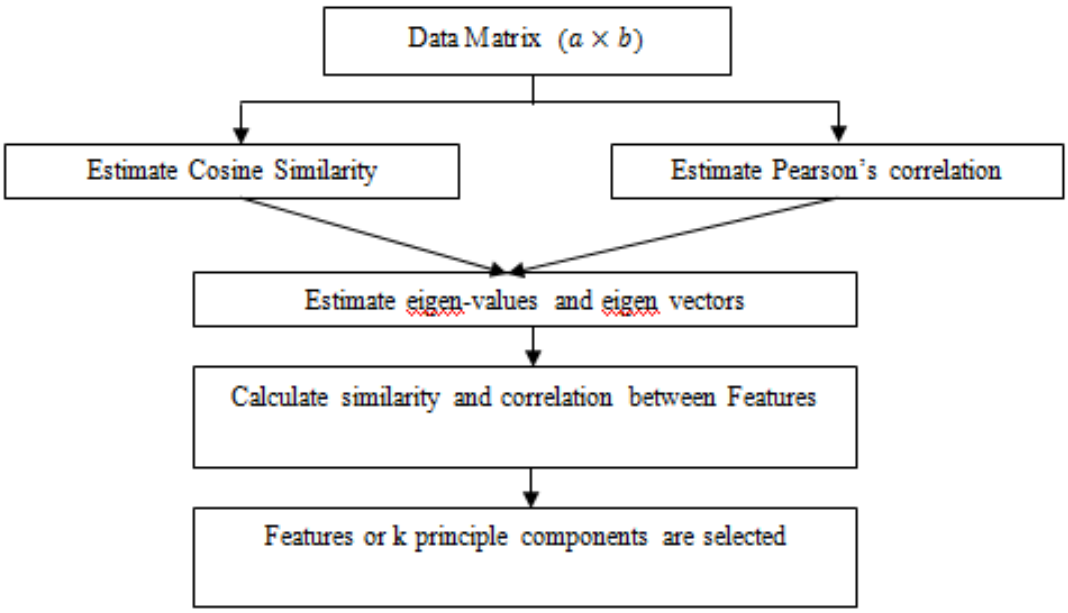


**Figure 2 Steps of PCA method**

Each kind of product features is found by this method and extracts the features of products according to their distinctions and similarities. The features of data can be easy to find each type of product with low dimensions such as bag, shirt, sneaker, etc. The PCA method is employed as one of the powerful tools to recognize the features of every individual product on Amazon. The steps for product feature extraction by the PCA method has been demonstrated in figure 2. From figure 2, cosine similarity and correlation matrix are combined to estimate the similarity between product features. For the product feature extraction, the PCA method uses the following steps:

Step 1: To take the data set matrix $(a \times b)$ matrix P and to take m represents the number of sample data and n represents the number of product features in product dataset

Step 2: To set the mean value of each feature is zero and to convey the matrix P to a mean detached matrix Q

Step 3: To calculate the cosine similarity

Two product feature vectors are $p = \{p_1, p_2, \ldots, p_n\}$ and $q = \{q_1, q_2, \ldots, q_n\}$ and then the cosine similarity is estimated by using the equation (1)

$$CosineSim(P,Q) = \frac{\sum_{i=1}^{n} p_i . q_i}{\sqrt{\sum_{i=1}^{n} p_i^2} \sum_{i=1}^{n} q_i^2} \quad (1)$$

Step 4: To estimate the correlation matrix to evaluate the power of a linear connection between two product features and it is given by

$$Corr(P,Q) = \frac{\sum_{i=1}^{n}(p_i-\bar{p})(q_i-\bar{q})}{\sqrt{\sum_{i=1}^{n}(p_i-\bar{p})^2}\sqrt{\sum_{i=1}^{n}(q_i-\bar{q})^2}} \quad (2)$$

In the above equation (2), $\bar{p}$ denotes the mean value of p and $\bar{q}$ represents the mean value of q

Step 5: To estimated eigen-values and eigenvectors of cosine similarity and correlation matrix.

Step 6: To select the eigenvectors such as principal components in the descending order of eigenvalues

Step 7: To extract and select the number of top principal components or features, to maintain applying the various criterion and an n×k projection matrix P is generated with every column vector of eigenvalues that can maintain the principal components to extract the product features based on their similarities and differences.

Step 8: Product Features are selected for product classification

In this type of feature extraction, cosine similarity and correlation equations are used by the PCA method to evaluate the similarity and correlation between every two vectors of product features in the training Amazon product dataset. The training product data's correlation matrix and similarity matrix has been made that can be utilized to extract the features. This will be defined by similarities between features for the feature extraction.

Finally, the number of principal components is maintained or the number of product features is chosen such as the value of k and it can be used to improve the classification accuracy on the training given product data set. Hence, each kind of product features is chosen to support the SVM classifier for the classification of online products present on the Amazon website.

## IV. SENTIMENT BASED FEATURE EXTRACTION USING T-SNE

The sentiment features are extracted from the review database using the t-SNE method. This method can also be applied to reduce the multi-dimensionality for the visualization of data during the product classification. t-SNE is executed as a non-linear method for the procedure of dimensionality reduction that is employed to discover the high-dimensional data and to decrease the difference between two contributions such as one contribution that will estimate the pairwise input product's similarities and another contribution that will estimate the equivalent low-dimensional points' pair-wise similarities in the embedding procedure. In this method, the similarity estimation is computed between the sample pairs present in the low and high dimensional space. Consequently, these two similarity estimation is optimized by applying the cost function. For the review feature extraction, the t-SNE algorithm is executed through the three values like perplexity, learning rate and supervise for the dynamic evolution.

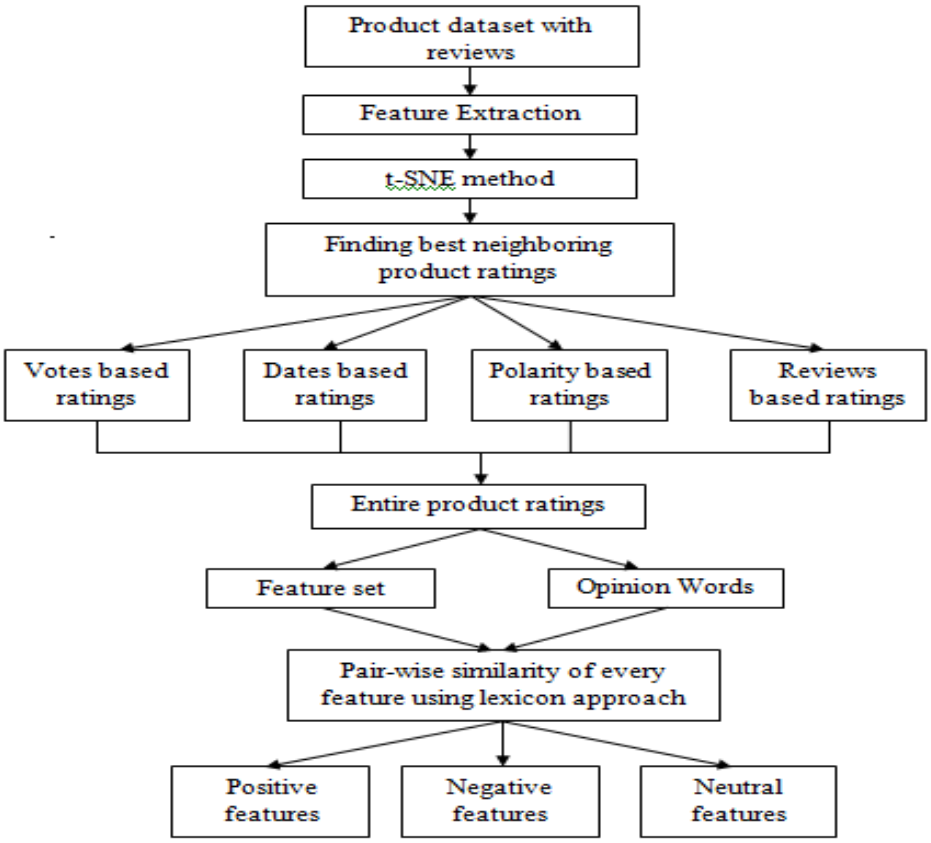


**Figure 3 Sentiment feature extraction using t-SNE**

The steps for feature extraction using the t-SNE method have been illustrated in figure 3. The product review dataset is used by the t-SNE method to extract the sentiment words for the discovering of best neighboring ratings about each product in the Amazon dataset. The t-SNE can be exploited to find the best neighboring product ratings with dimensionality reduction. The ratings about products are collected based on the votes, dates, polarity, and reviews. Sometimes, the reviews are in the form of polarity words. For this reason, the t-SNE method employs the lexicon-based approach to extract the features from the review database. Every word can correspond to one dimension and equal words to a similar dimension.

Therefore, t-SNE applies the dictionary-based method with the help of an online dictionary that including the antonyms and synonyms for every word. This lexicon approach contains a set of well-lexicons called WordNet dictionary process in the English language. WordNet dictionary has been maintained the set of the lexical dataset for online product review words and also maintains the sentiment relationship of each word in the dataset. The pair-wise similarity of every feature is extracted by t-SNE using a lexicon approach. The features are in the form of positive, negative or neutral words.

The t-SNE uses the perplexity parameter to balance the consideration between global and local review features of the given Amazon product dataset. This perplexity parameter will measure how many numbers of close best neighboring ratings are available in the product review database. The perplexity uses the values between 5 and 50 to find and extract the neighboring reviews or ratings. The t-SNE applies the learning rate in the range of [10.0, 1000.0] to find the best reviews about every product

For the sentiment-based feature extraction, the uniqueness of the reviews can be described concerning dual probabilities in the original space that can denote the pairwise similarities between ratings and it is given by

$$P(x,y) = \frac{\exp\left(-(d_{xy}^{o})^2/2\sigma_x^2\right)}{\sum_{k\neq r}\exp\left(-(d_{kr}^{o})^2/2\sigma_x^2\right)} \quad (3)$$

In the above equation (3), $d_{xy}^{o} = \left\|a_x - a_y\right\|_2$ , $\sigma_x$ can be decided by the perplexity parameter that has returned the efficient number of local neighbors ratings. This t-SNE method is utilized in the feature space to support the division of distinct review features. The pairwise similarities can be calculated by using the following equation (4).

$$q(x,y) = \frac{(1+(d_{xy}^{F})^2)^{-1}}{\sum_{k\neq r}(1+(d_{xy}^{F})^2)^{-1}} \quad (4)$$

In the equation (4), $d_{xy}^{F} = \left\|b_x - b_y\right\|_2$

The local structure of review features are maintained and extracted by the objective function during matching $P(x,y)$ and $q(x,y)$ and it is given by:

$$F(x) = \sum_{xy} P(x,y) log \frac{P(x,y)}{q(x,y)} \quad (5)$$

Hence, by using the above equations (3-4) and lexicon approach, the sentiment features are extracted by the t-SNE from the product review dataset to obtain the best positive reviews based product.

## V. PRODUCT CLASSIFICATION USING SVM

After the product feature extraction and sentiment feature extraction, the product classification is computed using the SVM method. The extracted product and review features are given to the SVM classifier to classify and separate the recommended product from the huge set of product dataset. The SVM classifier can classify the Amazon product based on product features and the best neighboring ratings.

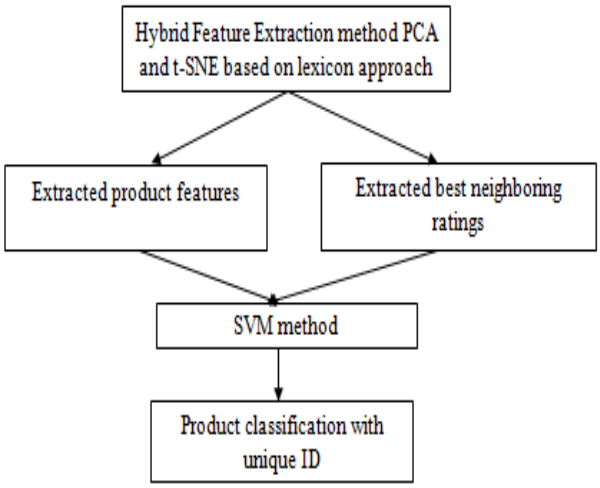


**Figure 4 Steps of product classification**

Figure 4 shows the product classification steps using the SVM method. In this figure 4, the PCA method has provided the extracted product features and the t-SNE method has provided the best neighboring ratings to the SVM classifier to acquire the best product classification. If anyone customer asks their required product, then this presented work can separate the asked product with best reviews from the given dataset and list the best rating products. Thus, the customer can get the best product with high positive reviews. For each case, we have to fix the component variance of v1, v2, and v3 to select a unique index ID. When the Iteration sequentially running, we can randomly stop the iteration process and looking for all jersey products and categorize using selected unique index ID and isolates the individual product.

In this type of classification, the classification line is discovered by the SVM method to generate the classification space the biggest. The best surface of classification can be found using some spatial form by isolating and extending the two kinds of samples accurately. The SVM classifier can reduce the multi-dimensionality space with the help of the t-SNE and PCA method. Here, the t-SNE method is applied to reduce the multi-dimensionality for the product image data visualization. The PCA method is used for the reduction of dimensionality for a large set of fashion product images such as shirts, coat, bag, sneaker, etc. The SVM method classifies the products according to their selected features and neighboring ratings using the following algorithm.

Step 1: Assume training data set $T = \{(a_1, b_1), \dots, (a_i, b_i) \in (A \times B)^{'}\}$, $a_i \in A = R^n$,

$b_i \in B = \{1, -1\}, i = 1,2, \dots, n$

Step 2: To compute optimal solution $p^* = (p_1^*, p_2^*, \dots, p_n^*)$ for the resolving of optimal problem

Step 3: To select the positive component $p_j^*$ of $p^*$ using following equation (6)

$$X^* = \sum_{i=1}^{n} p_i^* \; b_i a_i \quad (6)$$

And to estimate $q^* = b_j - \sum_{i=1}^{n} b_i \, p_i^* \, (a_i \, . \, a_j)$ (7)

Where, $p_i^*$ represents the Lagrange multiplier for every example, $q^*$ denotes the threshold of classification.

Step 4: To use equation (6) into hyper plane equation to obtain the best solution and it is called as decision function and it is calculated by using the following equation (8)

$$g(a) = sgn\{X^*a) + q\} = sgn\{\sum_{i=1}^{n} \alpha_i^* b_i(a_i * a) + q^*\} \quad (8)$$

Where, $sgn()$ denotes the symbol function.

Step 5: To recognize the optimal surface of classification by applying the inner product function $I(a_i, a_j)$ and it is given by

$$Y(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j \, b_i b_j \, I(a_i, a_j) \quad (9)$$

Step 6: Finally, to execute the classification process using the following equation (10)

$$g(a) = sgn\{\alpha_i^* b_i I(a_i, a_j) + q* \quad (10)$$

By using the above algorithm, the SVM method can classify the products and separate the individual products according to their neighboring ratings by selecting the product's unique index ID. Hence, the products have been classified and isolated depending on their unique ID from the large set of Amazon product dataset and recommend the best products to consumer.

## VI. RESULTS AND DISCUSSIONS

In this section, the performance of the proposed method is estimated and compared with existing methods in terms of parameters such as precision, recall, F-measure, accuracy and time complexity. We have taken the Amazon product data set to the estimation of the presented method's performance. This Amazon dataset contains details of obtainable products and consumer reviews. The parameters can expose the effectiveness of proposed and existing methods. This proposed product classification method will give very effectual and accurate results depending on the Amazon product dataset.

### 1. Precision, Recall and F-Measure

**Precision**

In these results, the precision of this proposed and existing methods are evaluated by using the values of correctly classified products with best ratings out of all classified products and it is evaluated by using the following equation (11)

$$recision = \frac{Values\ of\ correctly\ classified\ products\ based\ on\ the\ neighbor's\ ratings}{Values\ of\ all\ classified\ products\ based\ on\ the\ neighbor's\ ratings} \quad (11)$$

**Recall**

The recall of the presented and existing works is estimated by identifying the values of correctly classified products out of the actual product dataset and it is given by

$$recall = \frac{Values\ of\ correctly\ classified\ products\ based\ on\ the\ neighbor's\ ratings}{Values\ of\ actual\ products\ based\ on\ the\ neighbor's\ ratings} \quad (12)$$

**F-Measure**

The F-measure of this method is estimated with precision and recall that has been applied in the task of best product classification based on best ratings out of all products present in a given dataset.

$$F - Measure = \frac{2(precision \times recall)}{precision + recall} \quad (6)$$

**Table 1 Precision, Recall and F1-Measure of the presented and existing methods**

| Metrics | Proposed Method (%) | Standard PCA (%) | Linear SVM (%) |
|---|---|---|---|
| Precision | 93.9 | 87.5 | 86.7 |
| Recall | 92.7 | 89.3 | 84.2 |
| F-Measure | 93.3 | 88.84 | 85.43 |

The precision, recall and F-Measure of presented product classification method and existing methods are given in table 1. From the above table 1, the proposed method has taken a high percentage of precision, recall and F-measure compared to standard PCA and Linear SVM. As a result, this presented work has given the best performance to the sentiments extraction and product classifications based on ratings.

### 2. Classification Accuracy

The product classification accuracy is depending on the correctly classified products with best ratings. The classification accuracy of presented and existing methods has been illustrated in figure 5. The method's accuracy has been evaluated according to how many numbers of corrected classified best ratings based products are there out of many classified products based on consumer reviews.

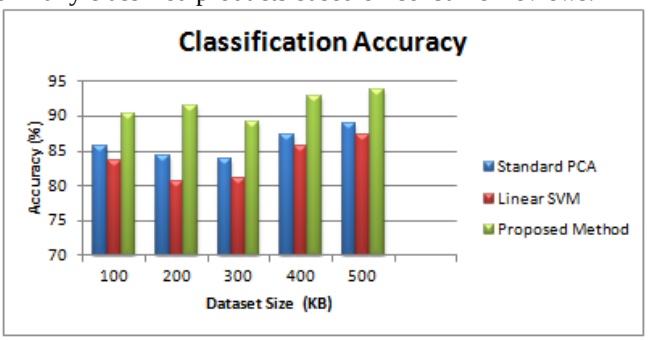


**Figure 5 Accuracy of proposed and existing methods**

The above figure 5 has clearly demonstrated that the proposed method has given high accuracy than the existing methods such as standard PCA and Linear SVM. From this comparison accuracy result, it is proved that the proposed system will provide the high values of corrected classified best ratings based products.

### 3. Time Complexity

The comparison chart of execution time is demonstrated in figure 6. In this kind of estimation, the product feature and sentiments feature extraction, and product classification time is evaluated depending on the total execution time on a given dataset. The execution time has been computed in milliseconds. The below chart 6 has clearly said that the proposed hybrid feature extraction and lexical based approach with SVM have taken only a small amount of time to execute the individual product classification based on the best ratings. As a result, this presented work will take low time complexity to classify and separate the individual products according to their ratings than the existing methods standard PCA and Linear SVM.

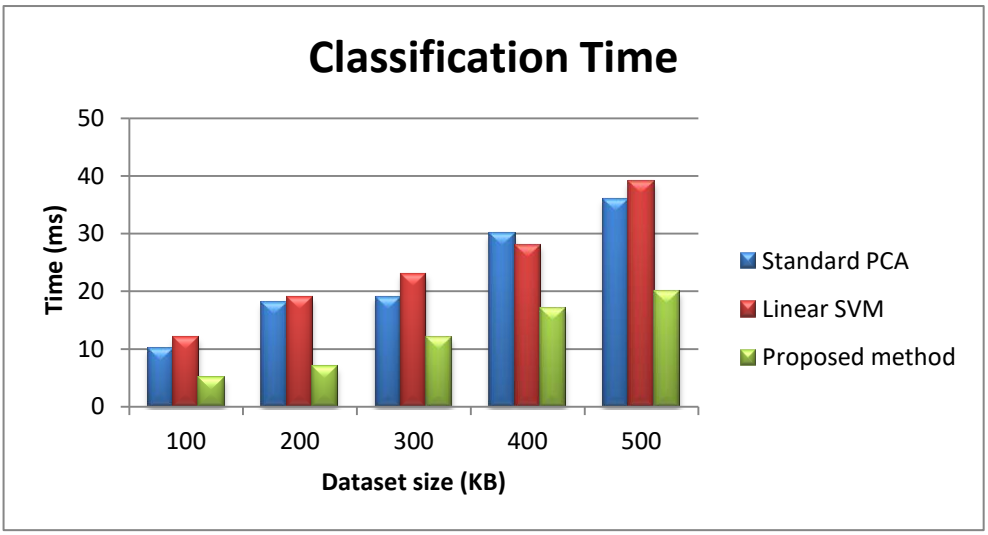


**Figure 6 Execution time of proposed and existing methods**

## VII. CONCLUSION

In this paper, an opinion based E-commerce product classification has been presented with hybrid feature extraction methods PCA and t-SNE, lexical approach and SVM. The amazon product dataset has taken in this work. The given dataset quality was improved by removing the irrelevant data from the dataset. The product features were extracted from the dataset by employing the PCA method and the sentiment features were extracted using t-SNE. The visualization of data has also been executed by the t-SNE method with dimensionality reduction. The sentiment features are extracted according to their positive, negative and neutral features. The best neighboring ratings were found and extracted in this method t-SNE. The SVM method has taken the extracted product and sentiment based features to the product classification. In this process, the individual products have been categorized according to the best fit neighboring product ratings with unique index ID. This presented work can recommend the best products to the consumers. Thus, the customer can take decision easily using this presented method whether the product can be bought or not.

## REFERENCES

1. Wei Zhao, Ziyu Guan, Long Chen, Xiaofei He, Deng Cai, Beidou Wang and Quan Wang, "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis", IEEE Transactions on Knowledge and Data Engineering", vol.30, no.1, pp. 185 – 197, 2018.
2. Jingbo Zhu, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou and Matthew Ma, "Aspect-Based Opinion Polling from Customer Reviews", IEEE Transactions on Affective Computing, vol.2, no.2, pp. 1211 – 1224, 2011.
3. Jia Chen and Gang Kou, "How Online Review Valance Affect on Consumer Opinion Evaluation?", Procedia Computer Science, vol.91, pp. 635-641, 2016.
4. Yuan Zuo, unjie Wu, Hui Zhang, Deqing Wang and Ke Xu, "Complementary Aspect-Based Opinion Mining", IEEE Transactions on Knowledge and Data Engineering, vol.30, no.2, pp. 249 – 262, 2018.
5. Wan-I. Lee, Shan-Yin Cheng and Yu-Ta Shih, "Effects among product attributes, involvement, word-of-mouth, and purchase intention in online shopping", Asia Pacific Management Review, vol.22, no.4, pp. 223-229, 2017.
6. Shenghua Liu, Xueqi Cheng, Fuxin Li and Fangtao Li, "TASC:Topic-Adaptive Sentiment Classification on Dynamic Tweets", IEEE Transactions on Knowledge and Data Engineering, vol.27, no.6, pp. 1696 – 1709, 2015.
7. Lin Sun, "Research on Product Attribute Extraction and Classification Method for Online Review", International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp.117-121, 2017.
8. Upma Kumari, Arvind K Sharma and Dinesh Soni, "Sentiment analysis of smart phone product review using SVM classification technique", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017.
9. Tejasvini Patil and Sachin Patil, "Automatic generation of emotions for social networking websites using text mining", Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Vol.4, pp.1-6, 2013.
10. Doaa Mohey El-Din Mohamed Hussein, "A survey on sentiment analysis challenges", Journal of King Saud University - Engineering Sciences, vol.30, no.4, pp. 330-338, 2018.
11. Zhang, Z, Ye, Q, Zhang, Z and Li, Y (2011), "Sentiment classification of Internet restaurant reviews written in Cantonese," Expert Systems with Applications, vol. 38, no. 6, pp. 7674–7682,
12. Chen, J, Huang, H, Tian, S and Qu, Y (2009), "Feature selection for text classification with Naïve Bayes," Expert Systems with Applications, vol. 36, no. 3, pp. 5432–5435.
13. Dan Li, Jiang Qian (2016), "Text Sentiment Analysis Based on Long Short-Term Memory", First IEEE International Conference on Computer Communication and the Internet, pp. 471 – 475.
14. Zhang H, Yu Z, XU M and Shi Y (2011), "Feature-level Sentiment Analysis for Chinese Product Reviews", IEEE International Conference on Computer Research and Development, Vol.2, pp. 135 – 140.
15. Xing Mou and Yajun Du (2016), "Sentiment Classification of Chinese Movie Reviews in Micro-Blog Based on Context", IEEE International Conference on Cloud Computing and Big Data Analysis, pp.313-318.