



# Personalized Recommendation System based Association Rule Mining and Sentiment Classification

Senthil S Sekhar, K. Satyanarayana

**Abstract:** In this emerging global economy, e-commerce is an inevitable part of the business strategy. Moreover, the business world comprises the upcoming entrepreneurs who are unaware of the current trends in marketing. Therefore, a recommendation system is very essential for them. In this paper, a fully automated recommendation system for the upcoming entrepreneurs to become successful in their business is proposed. The system works in three stages. In the first stage, the most transacted product is identified using association rule mining FP growth algorithm. This helps in extracting useful information from the previous transacted data by mining the entire set of frequent patterns. The second stage identifies the most customer preferred company based on review analysis. The multilevel clustering process with the generalization of data review is implemented to achieve an accurate review of the product. It rectifies the problems of shilling attack and gray sheep users commonly seen in single level K-means algorithm by refining the collected data. In the third stage, the reviews are sorted using a polarity shift sentiment classification algorithm. It helps in sort positive and negative reviews thereby rating a company. The top rated company would give the best product. Thus, the best product can be identified. From the experimental analysis, it is understood that the proposed system outperforms the existing recommendation methods. Moreover, this automated system helps the user to get the most accurate result within time. Hence, it would be very beneficial to the upcoming businessmen for flourishing their business in this increasing economic world.

**Keywords:** E-commerce, Recommendation System, Polarity shift sentiment classification, K-means

## I. INTRODUCTION

Today, a world without the internet cannot be imagined as it serves numerous applications to the developing society. The recent trend that exists online marketing that satisfies the needs of people much easier. However, the challenges faced them are many. One of them is the time consumption for surfing the internet. They are not aware of the customer's preferred products or repeated demand products. Even though social Medias help them to a great extent for customer review analysis, it is not reliable. There are numerous attacks

possible in such a collaborative collection of reviews namely, shilling attack and grey sheep users. When a user provides a review for their profit and false rating for the competitors, it is known as the shilling attack [1, 2]. Grey sheep attackers are those whose reviews are considered both reliable and non-reliable. Moreover, recommendation systems have been proposed by many researchers that resolve this issue. Online companies generally collect customer reviews and use them to recommend products to other companies at a lower cost. Such software systems that make customized responses are known as recommendation systems. The recommendation systems are beneficial for both the user and the vendor. They save the time and money of the users by helping them find out the best product and company within no time. However, the effectiveness and reliability of such recommendation systems are to be studied properly.

In this paper, a fully automated recommendation system is proposed. Here, association rule mining based on the FP growth algorithm is used. The review data set thus collected is analyzed using a review analyzer system. The review analyzer system is a two-stage system in which the collected review dataset is analyzed using a multilevel K-means algorithm and further classified into positive and negative reviews using an ensemble classifier. The best product is recommended using a product recommendation system based on the rating of a company based on the reviews. Methodologies and principles are explained further.

### A. Contributions

In this paper, the defects in traditional collaborative filtering recommender systems are rectified and a more personalized recommendation strategy based on a trusted community in user social networks has been proposed. In the first stage, most of the transacted product is detected through FP growth association rule mining method. The trusted reviews have been detected through user social network then an automated review analyzer is used for review prediction. Further, the personalized recommendation has been achieved using the trusted data reviews from ensemble review analyzer.

## II. LITERATURE SURVEY

The users in the social networking and online web application is growing enormously because of the information produced from the different sources in a rapid manner. Those sources like tweeter, Facebook, and another form of application that generates the data collects customer reviews as well.

Revised Manuscript Received on April 25, 2020.

\* Correspondence Author

Mr. Senthil S.Sekhar\*, P.hD Scholar, Department of Computer Science, VELS Institute of Science, Technology and Advanced Studies, Chennai.

Dr. K. Satyanarayana, M.Phil Computer Science, Manonmanium Sundaranar University, Tirunelveli.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

These data sources are integrated such that the different reviews are generalized based on the user and their profile. The raw data are collected which is pre-processed into specific data frames for better accuracy. The backbone of the review is an application user who only provides a review of the particular product or any other element in the online shopping arena [3].

The users are identified based on their profiles using a user profile analyzer. Various attacks affect the collaboration recommendation technique. Therefore, maximum reliability is attained only if the online review system is devoid of such attacks. Various studies have been done by researchers to reduce the attacks and to implement an efficient recommendation system. The recommendation system is referred to as a service that analyses customer reviews to determine the most customer-preferred product or service. It involves various filtering processes and algorithms depending upon the user. Collaborative filtering [4] is a method to predict the interests of customers automatically. The advantage of collaborative filtering is that it maintains the current customer behavior in the future also [5, 6]. It includes both a user-based and item-based approach in which user based is easier to implement and has fast processing. Moreover, data mining is a method that helps to find hidden relations, patterns and rules in huge amounts of data for extracting useful information from the same in a readily understandable form. Association rule is a data mining method that is mostly researched. Association rules describe the rules based on how one event is related to the other. It classifies data according to its relevance. Formally, it can be defined as a set of transactions,  $D = \{T_1, T_2 \dots T_n\}$ , where transaction  $T$  is defined as the set of items that build the transaction [7]. Association rule is expressed in form  $R: A \dot{\in} B$  which is referred to as when event  $A$  occurs, event  $B$  may occur. In association rule-based K-means algorithm, the value observed from association rules is inputted. The K-means algorithm repeats  $n$  dimension data items and further group them into clusters. It is highly efficient for the processing of a huge amount of data [8]. The FP Growth algorithm by Han focused an efficient method for mining the entire set of patterns and pattern growth. This is done with the help of an extended prefix tree structure for the storage of compressed and important information about frequent patterns named FP tree [9]. Through his study, Han proved that his method outperforms other methods like Apriori Algorithm for mining frequent patterns. When readers browse information about a book, the book recommendation system automatically recommends the books of their interest via mining. This is the data association relationship discussed in Apriori Algorithm which was proposed [7]. In [10] system implemented by acquiring system- customer interactions which can prevent inappropriate recommendation lead by interest alteration. In [11] introduced data weighting methods such as time-based data weight and item similarity-based data weight to adaptively track the transfer of user interest. However these systems have better accuracy of recommendation, but they are all from the perspective of the system in identifying the migration of user interest.

Malicious assault is another issue in a collaborative filter recommender system. Researchers have created numerous

techniques. As is appeared in reference, an attack recognition model has been introduced, given investigation of the rating information from five alternate points of view.

### III. PROPOSED SYSTEM

A fully automated recommendation system is proposed in this paper. This section consists of three modules; (1) Product analyzer (2) Review analyzer based ensemble classifier (3) Product recommender. The product analyzer is based on association rule mining using the FP growth algorithm. This helps in extracting useful information from the previous transacted data by mining the entire set of frequent patterns. Figure 1 shows the schematic diagram of the product recommendation system. The conventional approach for a recommendation system is the review analysis method. Manual collection of reviews may be time consuming and inefficient. Therefore, in this paper, an automatic review analysis based on the polarity shift method is used in the proposed system. Social networking platforms help in gathering input for the creation of a dataset in review analysis. However, genuine reviews cannot be sorted out as it has shilling attacks and grey sheep users. Therefore, the Multilevel clustering process with the generalization of data review is implemented to achieve an accurate review in any domain. The proposed method considers the user profile with history related to the particular review and also considers their relative interaction to other products with their reviews. The dataset is further classified into positive and negative reviews using polarity shift detection and ensemble classifiers [12]. This Polarity shift detection classification easily classify the reviews to predict the most suitable company. The detailed explanation is as follows.

#### A. Product Analyzer using Association Rule Mining

The product analyzer is based on association rule mining to find hidden relations, patterns, and rules in a huge amount of data. This is done to extract useful information from the previous transacted data. The collaborative method is a commonly used method to gather customer data. The advantage of collaborative filtering is that it maintains the current customer behavior in the future also. Collaborative filtering alone doesn't make sense to compare massive data to find the most customer-preferred product. Association rule is a data mining method that is mostly researched.

Association rules describe the rules based on how one event is related to the other. It classifies data according to its relevance.



Figure1: Schematic diagram of the product recommendation system

Formally, it can be defined as a set of transactions,  $D = \{T_1, T_2 \dots T_n\}$ , where transaction  $T$  is defined as the set of items that build the transaction. Association rule is expressed in form  $R: A \dot{\in} B$  which is referred to as when event  $A$  occurs, event  $B$  may occur.

There are two types of association rule mining, namely the FP Growth Algorithm and the Apriori Algorithm. Studies prove that the FP Growth algorithm is efficient as compared to the Apriori algorithm. Therefore, FP Growth Algorithm is used in the proposed system. A comparison between both association mining rules can be observed in the experimental results.

#### a. FP Growth Algorithm

The FP Growth Algorithm is an efficient method for mining the entire set of frequent patterns using pattern fragment growth. It is an alternate path to determine the frequent set of items without the involvement of candidate generations. The basic strategy of this method is the usage of frequent pattern tree called FP tree which helps to retain the item set association information. Initially, the algorithm compresses the input database thereby making an FP tree for representing frequent items. Further, it divides the database into conditional databases. Each conditional database is associated with a frequent pattern. In the end, each database that was obtained is mined separately. This method reduces the search costs that look for short patterns offering better selectivity. It requires only two scans as compared to Apriori Algorithm which requires  $k-1$  scans. FP Growth algorithm uses leverage, conviction and lift techniques with a minimum threshold for data mining.

### B. Review Analyzer based Multilevel K means Algorithm

In a review analyzer system, trusted review dataset availability is a challenging issue. Normally the user reviews are gathered from social networking platforms. These networking sites may have various forms of review for a particular domain in the information handling process. The application users can be of two types, mainly registered users and unregistered users. Registered users usually provide genuine reviews. Those reviews are not reliable as it may have false reviews. The unregistered users may provide genuine reviews but not trustworthy. The proposed system has two stages of review analysis, a multilevel K means algorithm-based analysis and review classification based on ensemble classifier. This helps in dataset creation devoid of shilling attacks and grey sheep users. The multilevel clustering process with the generalization of data review is implemented to achieve an accurate review of the product. During a review analyzing process, the data arrive from numerous sources which cause difficulty in analyzing. Therefore, the existing methods usually depend on a specific source for review analysis. However, review analysis from more than one source is essential for accurate and genuine results. This would improve the reliability of the user as well as the vendor. The proposed method considers the user profile with history related to the particular review and also considers their relative interaction to other products with their review. Two levels of clustering process have been carried out namely the data specific cluster and generalized cluster as outer

cluster and integrated cluster respectively. This is due to the insufficiency of one level clustering. The outlier of each cluster is also taken into account to achieve maximum accuracy. The K-Means clustering process is then considered for performance evaluation. But the collected reviews have both negative comments and positive comments.

#### a. Multilevel K means Algorithm

Social networking and online web application users grow tremendously because of the rapid production of information from various sources. Moreover, these sources like Facebook, Twitter, and other applications collect customer reviews along with the generation of data. The integrated data set used in the reason for different reviews are generalized based on the user and their profile. The raw data are collected and are pre-processed further into specific data frames for better accuracy. The backbone of the review is an application user who only provides a review of the particular product or any other element in the online shopping arena. The users are identified based on their profiles using a user profile analyzer. Various attacks affect the collaboration recommendation technique. So it has to be handled properly to reach maximum reliability level. Shilling attack [13] is one type of attack in which the user gives positive reviews or feedback to their product or element and false review to their opponent i.e. competitors. This problem is rectified by using the user categorization to achieve better trust over the online application. In the proposed method used two varieties of major phases namely user profile analysis phase and integrated clustering phase.

The second phase of the proposed method is an integrated clustering process in which different data frames are clustered with a common data format with generalization property. The data collection process is carried out from a reliable user which is further identified with their profile and early history.

Various data are clustered to form various clusters with data set. In phase1 the K-Means Clustering method is applied based on the number of initial cluster sizes. The centroid is identified and further, it groups the relevant data as one cluster. So many clusters are formed based on the number of data sources in that identified type of data with the profile. They are clustered data and outlier's data. The clustered data holds relevant data whereas the outliers are data that do not belong to any cluster. The clustered data which are identified during phase1 have various forms of review with a different profile. This is generalized into a common form of review based on the reliable user and source..

The outliers identified in two phases are clustered together to refine and maximize the reliability in the review from the user. The clustered are integrated and it performs the second level of K-Means Clustering for generalized feedback and review over the product or relevant element. Finally, the analytical report of the reviewed data is analyzed.

#### b. Review Analyzer based Ensemble Classifier

Review analyzer based ensemble classifier is the second stage of review analyzer system. Here,

the created dataset using Multilevel K means algorithm is sorted into positive and negative reviews. This helps in extracting genuine information regarding the user reviews of a product. In the initial stage, a fusion model to detect different types of polarity shifts such as explicit negation, explicit contrast specifically is introduced. A Classifier model is used to detect explicit negations and contrasts. Figure.2 shows the system overview of the proposed review analyzer system. Multi-level K-means algorithm ensures the system to gather the review dataset from a trusted community. Here, there are two stages namely, a training stage and a testing stage.

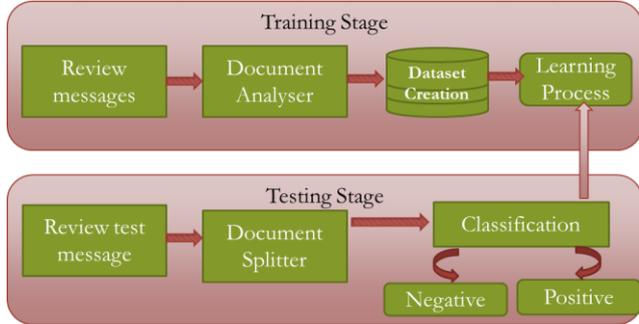


Figure 2: Overview of a Review analyzer

In the training stage, the classifier is trained with two sets of the list such as explicit negation and explicit contrast list. In the testing phase, k means clustered reviews are taken as the input of the review classifier. Initially, the reviews are called a document. Document splitter module splits the document into sentences which are further split into words by sentence splitter. The words are classified as a negative word or positive words using the classifier. In the training stage, both classes are trained using ensemble AdaBoost classifier [14]. So the rating of reviews can be easily predicted. But the Polarity shift is a complex linguistic structure [15,16] that may include explicit negations, contrasts, and intensifiers diminish, etc. Negation is the most common type of polarity shift.

The pseudo code for review analyzer based ensemble classifier is illustrated in figure 3.

C. Product Recommendation System

The previous transacted review data is gathered with the help of association rule mining using the FP growth algorithm. Multilevel K means algorithm creates a review dataset from a trusted community. The dataset thus created would be devoid of shilling attacks and grey sheep users.

Genuine user reviews determine the accurate prediction of the best product. The review dataset is further sorted into positive and negative reviews using polarity shift detection based on the ensemble classifier. The review analyzer thus helps in determining the rating of a company based on the reviews. The companies are sorted based on the rating. Therefore, the product recommendation system helps the user to easily identify the best product and company within no time. The existing methods fail to satisfy customer needs as it involves complex processes and inaccurate analysis systems. The proposed system experimentally proves to be the most accurate and user satisfying recommendation system among the existing methods.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The Experimental study has been done on the Amazon transaction dataset and also reviews data collected from social media. For achieving better results, the dataset used in the FP algorithm and Apriori algorithm should be free from redundant attributes and should be clear. For this purpose, data preprocessing is done.

```

• Input: document D, set of negators N = {n1,n2,n3 ... nt}, and set of contrast and C = {c1,c2 ... ct}
1. Document split into sentence D = {S1, S2 ... Sm} using Split function
2. Sentence split into words Sm = {w1, w2 ... wmm} using Split function
3. Trained the lists for classification class 1 is negator class 2 is contrast N = {n1,n2,n3 ... nt}, C = {c1,c2 ... ct} as result a model is created Model M using ensemble classifier
4. For i = 1: mn
5. if wi ∈ M; wi is class of N Sm put into Snegation
6. Continue
7. if wi ∈ M; wi is class of C Sm put into Scontrast
8. Continue
9. Sno-shift = D - Snegation - Scontrast
• Output: Snegation, Scontrast, Sno-shift
    
```

Figure 3: The Pseudo Code for Review Analyser based Ensemble Classifier

Figure 4 illustrates the performance evaluation of FP growth and the Apriori algorithm with execution time. From figure 4(a), it is clear that the FP growth execution time curve rate is less as compared to the Apriori algorithm concerning the evaluation support values of 0.2, 0.5 and 6500 transactions respectively. Similar results are obtained in Figure 4 (b) and (c) also. This clearly shows that the execution time for the FP growth algorithm is less than the Apriori algorithm. That is why the FP growth algorithm is used in the proposed most transacted product finder.

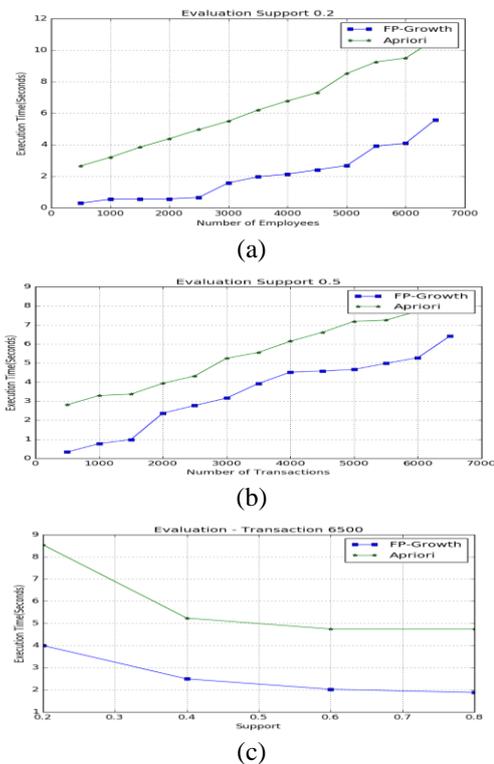


Figure 4: Graph for execution time to increasing number of employees for an evaluation support of (a) 0.2, (b) 0.5, (c) Graph of Execution time to Support level for 6500 Transactions.

In the system, the most transacted product is obtained using the parameters such as support, confidence and lift value. The most transacted product will have the maximum lift value. The concepts of three parameters are explained below with the help of an example.

Assume we have a list of 6500 customer transactions, and we have to find the Support, Confidence and Lift for two items e.g. mobile phone and headphone. Out of 6500 transactions, 650 contain mobile phone while 1625 contain headphones. Out of 1625 transactions where headphones are purchased, 780 transactions contain mobile phones as well. Using this data, we have to find the support, confidence and lift.

Support refers to the default most transacted item and can be calculated by finding several transactions containing a particular item divided by a total number of transactions. Suppose we want to find support for item headphones, it can be calculated as:

$$\text{Support (headphones)} = (\text{Transactions containing headphones}) / (\text{Total Transactions})$$

Confidence refers to the likelihood that a headphone is bought along with the mobile phone. It can be calculated by finding the number of transactions where Mobile phones and headphones are bought together, divided by a total number of transactions where the Mobile phone is bought. Mathematically, it can be represented as:

$$\begin{aligned} \text{Confidence (Mobile phone} \rightarrow \text{headphone)} \\ = (\text{Transactions containing both (Mobile phone and headphones)}) \\ / (\text{Transactions containing Mobile phone}) \end{aligned}$$

Lift (Mobile phone  $\rightarrow$  headphone) refers to the increase in the ratio of sale of headphones when the Mobile phone is sold. Lift (Mobile phone  $\rightarrow$  headphones) can be calculated by dividing Confidence (Mobile phone  $\rightarrow$  headphones) divided by Support (headphones). Mathematically it can be represented as:

$$\text{Lift (Mobile phone} \rightarrow \text{headphone)} = (\text{Confidence (Mobile phone} \rightarrow \text{headphone)}) / (\text{Support (headphone)})$$

For instance, from the first item, it is observed that mobile phones and wireless headphones are commonly bought together. This makes sense since people who purchase mobile phones are careful about flexible movements hence they are more likely to buy wireless headphones. The support value for the first rule is 0.0045. This number is calculated by dividing the number of transactions containing Mobile Phones by a total number of transactions. The confidence level for the rule is 0.2905 which shows that out of all the transactions that contain Mobile Phones, 29.05% of the transactions also contain Wireless headphones. Finally, the lift of 4.84 shows that the wireless headphone is 4.84 times more likely to be bought by the customers that buy the mobile phone as compared to its default sale.

The first rule has already been discussed. The second rule states that Television and Home Theatre are bought frequently. The support for Television actions containing Home Theatre is that 37.28% of the transactions are likely to contain Home Theatre as well. Finally, the lift of 4.47 shows that the Home Theatre is 4.47 more likely to be bought by the customers that buy Television, compared to its default sale.

Results are then compared with the results obtained when the Apriori algorithm was used. Here the FP growth algorithm

is chosen on the sorted transaction data over the Apriori algorithm as the performance of the FP growth algorithm is more efficient. FP growth algorithm, unlike the Apriori algorithm, produces frequent itemsets without the involvement of candidate itemsets thereby reducing the execution time and memory required. It uses a search technique of divide and conquers and for the entire process. It requires only two scans as compared to Apriori Algorithm which requires k-1 scans. From the experimental analysis, it is clearly understood that the FP- growth algorithm is better as compared to the Apriori algorithm.

Association rule mining includes two types namely, FP growth algorithm and apriori algorithm. In this analysis, both the FP growth algorithm and the apriori algorithm are compared to find the algorithm having less execution time.

In association, rule mining lift is a measure of association rule in predicting as having an enhanced response which is measured against a random choice.

Lift is simply the ratio of target response to the average response. For example, if a mobile phone has an average response rate of 5% but a certain model has a response rate of 20%. Then that model will have a lift value of 4. Support of an association rule is simply the measure of how frequently the items involved in it occur together. It is the number of transactions containing a pattern.

#### A. Performance Evaluation of Product Analyser

The system performance is evaluated with True Positive Rate (TPR), True Negative Rate(TNR), Positive Predictive Value (PPV), Negative Predictive Value (NPV), False Negative Rate(FNR), False Positive Rate(FPR), False Discovery Rate(FDR)[11]. The proposed product analyzer system has a better performance than existing systems.

Table 1 shows the statistical parameters of the performance analysis of the proposed system accuracy is attained at 96.75%.

**Table 1: Performance evaluation of product**

Parameters	Value
TPR	0.99162
TNR	0.90669
PPV	0.964019
NPV	0.97723
FNR	0.00838
FPR	0.09331
FDR	0.035981
FOR	0.02277
<b>Accuracy</b>	<b>0.9675</b>

#### B. Performance Evaluation of Review Analyser

The second evaluation is based on the review analyzer in which the most transacted product reviews are collected from the dataset for finding a best or suitable company for the most transacted product.

The classifier method automatically classifies the reviews. The company gaining the most number of positive reviews is tagged as the best

Figure 4 shows the clustered result of Multilevel K- means algorithm. These graphs (a), (b), (c) and (d) show the reviews of mobile buyers, headphone buyers, television buyers, and home theatre buyers respectively. In each of the graphs, five companies are represented here 1- comapny1, 2-comapny2, 3-comapany3, 4-comapny4, 5-comapny5 in the x-axis v/s number of users in the y-axis (like Samsung, Panasonic, Sony, etc.). The blue bar graph represents genuine reviews whereas green represents the forged reviews. In the case of mobile buyers, company 5 has 400 to 350 genuine reviews and zero forged reviews. Headphone buyers provided 60 to 52 genuine reviews and zero forged for company 5 and 35 genuine reviews for company 4. In graph (c), television buyers, company 5 has received 200 to 180 genuine reviews and negligible forged reviews. Similarly, in the case of home theatre buyers, company 5 have 48 to 45 genuine reviews and zero forged reviews. From these graphs, it is clearly understood that company 5 has received the most genuine reviews. Therefore, the most transacted product is also from company 5.

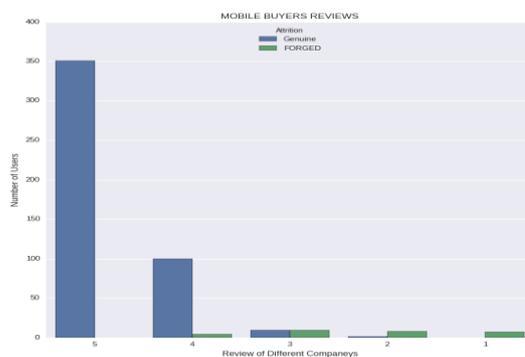
Performance evaluation of the Multilevel K means the Algorithm is shown in Table 2. Accuracy 92.72% for mobile review set, 91.81% for Headphone, 85.455 for Television review set is obtained in the proposed system. Table 3 shows the performance evaluation of the review analyzer with different classifiers, 94.04% accuracy is obtained in the review classification based on ensemble AdaBoost. Using this system, the product quality and robustness of trusted community detection and review classification can be increased and human labor costs, expenses and mental stress can be reduced.

**Table 2: Performance Evaluation of Multilevel K means Algorithm**

Dataset	Mobile	Headphone	Television
TPR	0.96	0.9789	0.9375
TNR	0.6	0.533	0.2857
PPV	0.96	0.93	0.9
NPV	0.6	0.8	0.4
FNR	0.04	0.021	0.0625
FPR	0.4	0.46	0.7142
FDR	0.04	0.07	0.1
FOR	0.4	0.2	0.6
<b>Accuracy</b>	<b>0.9272</b>	<b>0.9181</b>	<b>0.8545</b>



(a)



(b)

**Figure 4: Clustered Result of Multilevel k means Algorithm**

**Table 3: Performance Evaluation of Review analyzer with different classifiers**

Parameters	Value (%)
TPR	97.95
TNR	65.21
PPV	92.3
NPV	88.23
FNR	2.04
FPR	34.78
FDR	7.69
FOR	11.76
Accuracy	91.73

Table 4 is a comparison of the existing methods of review analysis. It is clearly understood that the method of having bigram features give more result as compared to unigram features.

**Table 4: Performance Comparison of Review analyzer with existing methods**

Methods	Accuracy %
Na.J C et.al[17](unigram features)	78.33
Na.J C et.al[17](bigram features)	79.33
Rui Xia et.al[15](unigram SVM)	82.5
Rui Xia et.al[15](bigram SVM)	83.8
Rui Xia et.al[15](unigram LR)	83.2
Rui Xia et.al[15](bigram LR)	84.6
Rui Xia et.al[15](unigram Ensemble)	84.6
Rui Xia et.al[15](bigram ensemble)	86.23
<b>Proposed</b>	<b>94.043</b>

### C. Performance Evaluation of Recommendation System

Table 5 illustrate the performance evaluation of the proposed recommendation system the proposed system has a greater performance based on PPV and accuracy.

Table 6 shows the performance comparison of a recommendation system with existing methods that the proposed system, it clearly understood the proposed system has greater accuracy than the existing methods. Therefore, it can be suggested as the best recommendation system for upcoming entrepreneurs.

Table 5: Performance Evaluation of Proposed Recommendation System

Dataset	Navies Bayes	Ensemble ada boost	SVM
TPR	0.9582	0.9892	0.9835
TNR	0.7188	0.8093	0.7466
PPV	0.9001	0.9329	0.9051
NPV	0.8669	0.9656	0.9484
FNR	0.0417	0.0107	0.0164
FPR	0.2811	0.1906	0.2533
FDR	0.0998	0.067	0.0948
FOR	0.133	0.0343	0.0515
Accuracy	0.892	<b>0.9404</b>	0.915

Table 6: Performance Comparison of Recommendation system with existing methods

Methods	Accuracy %
Pham et.al[16]	82.45
Raj et.al [18]	81.06
Nawrocka et .al [19]	84.55
Wenjuan et.al[14] (CF)	83.57
Wenjuan et.al[15] (Trusted community)	79.31
<b>Proposed</b>	<b>91.73</b>

### V. CONCLUSION

In this work, the classical defects in traditional collaborative filtering recommender systems are rectified and a more personalized recommendation strategy based on a trusted community in user social networks has been proposed. In the first stage, the most transacted product is detected through the FP growth association rule mining method. The trusted reviews have been detected through user social network then an automated review analyzer is used for review prediction. Further, the personalized recommendation has been achieved using trusted data. There are different types of review results that are available in the online application especially in social networking and online shopping. The existing method of analysis depends only on a particular data source and suffers from a huge number of attacks. These types of attacks are confused to the user by posting misinformation which leads the unreliable problem. The reviews are handled

properly using the cluster method in the Existing algorithms are restricted to the specific type of source in there, the standardization problems occur. The proposed method focused in customer reviews by considering new approach. The main phases of clustering techniques carried out with generalized review data which provides a high level of satisfaction to the customer. Various clustering algorithms are analyzed and compared with the proposed method. The multilevel clustering method achieves maximum performance with real-time data from the online application. Using the Polarity shift detection classification we can easily classify the reviews to predict the most suitable company.

### REFERENCES

- Chirita, P. A., Nejdl, W., & Zamfir, C. (2005, November). Preventing shilling attacks in online recommender systems. In Proceedings of the 7th annual ACM international workshop on Web information and data management (pp. 67-74). ACM
- Wenjuan, W., & Zhubing, L. (2015, July). A personalized recommendation strategy based on trusted social community. In 2015 10th International Conference on Computer Science & Education (ICCSE) (pp. 496-499). IEEE.
- ZENG, Z., & ZHANG, L. (2008). An intelligent shopping system based on multi-attribute decision and collaboration filtering [J]. Engineering Journal of Wuhan University, 2.
- Lim, S. L., & Finkelstein, A. (2011). StakeRare: using social networks and collaborative filtering for large-scale requirements elicitation. IEEE transactions on software engineering, 38(3), 707-735.
- Korea Communications Agency. Elements according to the step by utilizing big data technology Propulsion Trends and Implications. Information Communication Technology Issues& Prospects No.10, 2013.
- Mehta B, Hofmann T, Fankhauser. Lies P et al.: Detecting spam users in collaborative filtering.[C].In Proceedings of the 2007 International Conference on Intelligent User Interfaces ACM press January 2007. 14-21.
- R. Agrawal, T. Imielinski and A. Swami. Mining Association Rules between Sets of Items in Large Database. Proceedings of ACM SIGMOD on Management of Data, PP. 207-216, 1993.
- JinHyun Jooa, SangWon Bangb, GeunDuk Parka. Implementation of a Recommendation System using Association Rules and Collaborative Filtering, Procedia Computer Science 91 (2016) 944 – 952
- Han, J., Pei, J., Yin, Y. et al. Data Mining and Knowledge Discovery (2004) 8:53
- Min-Jeong Kang. Recommendation System for Make a Purchase to Application in Smart-Phone User. Hongik University Master's Thesis;2011
- Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. Item-Based Collaborative Filtering Recommendation Algorithms. In Proceedings of the 10th International World Wide Web Conference, Pages 285-295, 2011.
- Li,S.,Lee,S.,Chen,Y.,Huang,C.,&Zhou,G.(2010).Sentimentclassification andpolarityshifting.InProceedingsoftheinternationalconferenceoncomputationallinguistics(COLING).
- Zhang, Y., Tan, Y., Zhang, M., Liu, Y., Chua, T. S., & Ma, S. (2015, June). Catch the black sheep: unified framework for shilling attack detection based on fraudulent action propagation. In Twenty-Fourth International Joint Conference on Artificial Intelligence.
- Wenjuan, W., & Zhubing, L. (2015, July). A personalized recommendation strategy based on trusted social community. In 2015 10th International Conference on Computer Science & Education (ICCSE) (pp. 496-499). IEEE.
- Xia, R., Xu, F., Yu, J., Qi, Y., & Cambria, E. (2016). Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. Information Processing & Management, 52(1), 36-45.
- Pham, M. C., Cao, Y., Klamma, R., & Jarke, M. (2011). A clustering approach for collaborative filtering recommendation using social network analysis. J. UCS, 17(4), 583-604.

17. Na, J. C., Sui, H., Khoo, C. S., Chan, S., & Zhou, Y. (2004). Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews.
18. Raj, S. D. (2017, October). Automated service recommendation with preference awareness: An application of collaborative filtering approach in big data analytics. In 2017 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 766-769). IEEE.
19. Nawrocka, A., Kot, A., & Nawrocki, M. (2018, May). Application of machine learning in recommendation systems. In 2018 19th International Carpathian Control Conference (ICCC) (pp. 328-331). IEEE.

### AUTHORS PROFILE



**Mr. Senthil S. Sekhar** Received M.Sc. Computer Science in 1998 from Bharathidasan University, Tiruchirappalli and M.Phil Computer Science in 2007 from Manonmanium Sundaranar University, Tirunelveli. He is currently doing his research in Vels University, Chennai. His research interest Data mining, Machine Learning and Deep Learning, Recommender Systems.



**Dr. K. Satyanarayana** received M.Sc (Mathematics), M.Phil (Mathematics) from Bharathidasan University, Tiruchirappalli. His M.Phil (Computer Science) is from MS University, Tirunelveli, His Ph.D in Computer Applications is from Dr.M.G.R University, Chennai. Presently he is working as Principal, Sindhi College, Chennai. His research interests are in the areas like Data mining, Machine learning, ANN and Recommender Systems.