Enhanced Feature Specific Collaborative Filtering Model For Aspect Opinion And Temporal Based Product Recommendation

J. Sangeetha, Dr.V. Sinthu Janita Prakash

Abstract: nowadays, the online purchasing and advertising becomes massively increased due to the increase in utilization of internet services by the users. For the product sale and its quality description, the product's customer review plays a significant role. Thus, the words and phrases with large size in a raw data is converted into numerical values based on the opinion prediction method. The fault prediction of the reviews and inappropriate recommendation of the best product to the users are the main challenging issues in recent days. To avoid these issues, an Enhanced Feature Specific Collaborative Filtering Model (EFCFM) is proposed for Aspect Opinion and Temporal Based Product Recommendation system. Initially, the raw data is preprocessed using stop word removal technique and the keywords from that preprocessed data is extracted using POS tagger which has both positive and negative polarity. The features of the keywords are extracted from the Senti-WordNet database, product property from the POS tagger and the reviews from the user ratings. Then the Enhanced Feature Specific Collaborative Filtering Model is used to calculate the product's strength and weakness. Also it helps to predict the corresponding characteristics and its opinions. After that, the user query is also analyzed and finally, the opinion score based product recommendation is obtained. The proposed EFCFM technique is analyzed comparatively with other existing techniques with the metrics like precision, recall, f-measure, RMSE, and the MAE. The evaluation results show that the proposed EFCFM technique offers best product recommendations accurately to the users.

Index Terms: User Reviews, Opinion Mining, Collaborative Filtering, Opinion Score, Product's Property, EFCFM, Recommendation Systems.

I. INTRODUCTION

In modern era, recommendation systems Lu, et al.[1] received much consideration from analysts in the web based e-commerce community because of their part in expanding deal profits and also mitigating the data overload issue, by giving users customized suggestions about products and services. So as to avoid uncertainties in circumstances where they need to select amongst the products they are tackled with, it generally depends upon on recommendations provided by other users. These recommendations are provided directly through words or via texts and videos. Examples for influencer are Movie and book reviewers, newspapers and online social networks. A recommendation system aids to improve the capability and efficiency of transferring and getting ideas from an eminent progression in the social interactions amongst the people. In a usual system, the user deliver assessments of products they have bought. These assessments are represented as ratings which are utilized by some users. With these ratings the best products are recommended to the other users. These systems need to obtain the best mixture of user prospects and suitable product to be recommended, i.e., determining the connection of interest and options which is a major problem. For instance, the review mining and determining the relation amongst the reviewers in social networks has been turned into a major drawback in machine learning, web mining, and natural language processing. They are largely concentrated on the prediction task of ratings. Though, star-level information of the user's rating is not always present on several review websites. On the other hand, reviews had sufficient comprehensive product statistics and user opinion data, which have great reference value for a user's choice.

- *J. Sangeetha, Asst. Professor, Cauvery College For Women, Trichy-18*
- *Dr.V. Sinthu Janita Prakash, Head, Department of Computer Science, Cauvery College For Women, Trichy-18 Email: sinthujanitaprakash@yahoo.com*

Subsequently, there are numerous unrated products in a grid. In such cases, it is advantageous and important to the user reviews for predicting the unrated things. Recommender systems is categorized into 3 most important classes about the method used to produce the recommendations: (i) content-based approach Thorat, et al [2], in which similar products that are showed according to the user's preferences are recommended; (ii) collaborative filtering Wang, et al [3], which recommends products that are selected by people with related preferences to the user and; (iii) hybrid approaches that combine the methods of both the previous methodologies which tries to resolve some complications. Content-based filtering is utilizing the method to examine the documents and depictions of products that are formerly rated by a user, and then form a profile of the users interests based on the rated products' features. By this profile, the recommender system can filter out the recommendations which are fitted for the users. Collaborative filtering (CF) Riya and Varghese [4] is one among the typically used recommendation methods. This method computes the similarity amongst the users and utilized this data to recommend the products which are not attempted by the target user. This similarity is depends upon the former reviews of shared products and it is utilized to give product recommendations that were previously assessed by these related users but were not assessed by the target user. The chief objective of this proposed approach are listed as follows:

- Review is recommended based on the Time and Location of the product.
- Based on the user query, the features are extracted concerning its strength and weakness for the better recommendation of product.

The remaining of this paper is structured as follows: The detailed descriptions of the related works on the recommendation system, filtering techniques, and the rating prediction technique are discussed in section II. The implementation process of an Enhanced Feature Specific collaborative filtering based clustering mechanism (EFCFM) for Aspect Opinion and Temporal Based Product

Recommendation in section III. The comparative analysis of proposed ESCFM technique with prevailing approaches is provided in section IV. Lastly, section V gets concluded the proficiency of the proposed ECFCM approach.

II. Related Work

This section gives the existing work on the data preprocessing techniques including Stop words removal, Keyword Extraction, finding the polarity from the datasets. Further, the usages of existing data mining techniques are to be discussed. Najafabadi, et al. [5] offered the Explicit Factor Model (EFM)to produce reasonable recommendations, in the meantime preserved a maximum prediction accuracy. They extracted the features of the specific product and opinions of the user by the sentiment analysis using phrase-level on user reviews, then produced both recommendations and non-recommendations with regards to the features of specific product to the interests of user and the unknown features were learned. Here, the drawback was only one measure per aspect were used. Zhao, et al. [6] represented a novel demographic based product recommender system that perceived the purchase plans of users from their microblogs in the real-time and created product recommendations with corresponding users' demographic data mined from their profiles according to the product demographics obtained from online reviews and microblogs. This system was not controlled by any precise e-commerce website and was able to create instant product recommendations to users who have stated their purchase plans in micoblogs. Ling, et al. [7] proposed a model that integrated the content-based and collaborative filtering. By developing the information in both ratings and reviews, it can be able to increase the prediction accuracy considerably through several classes of datasets than the existing methods, specifically under the cold-start scenarios where the data were tremendously low. They developed an effectual collapsed Gibbs sampler for determining the model constraints. This model also determined the topics that are interpretable, allowing us to accomplish prior knowledge to reduce the issue of cold start. Yang, et al. [8] suggested the hybrid collaborative technique with the content based filtering technique. They adopted a successful Statistical Relational Learning (SRL) approach for the weights and learning features are the first to implement such a system in a real-world big data context. This algorithm is capable of handling different costs for false positives and false negatives making it extremely attractive for deploying within many kinds of recommendation systems, including those within the domain upon which tested. There are three main works are employed such as 1) the first work which employs probabilistic logic models to build a real-world large-scale job recommendation system; 2) it is the first work which allows the recommender to incorporate special domain requirements of an imbalanced cost matrix into the model learning process; 3) it is the first to prove the effectiveness of statistical relational learning in hybrid recommendation system data. McAuley and Leskovec [9] suggested a HFT (Hidden Factors and Topics) model that integrated the scores with text review for product recommendations. HFT was employed by arranging the unknown aspects in product ratings with unknown subjects in the reviews of a product. Basically, these subjects operated as normalizes for concealed user and product constraints. This helps the users to fit accurately and product constraints with a limited reviews only. This was not done in existing methods. Skabar and Abdalgader [10]proposed a novel fuzzy clustering algorithm for handling the relational input data. The suggested algorithm exploited the graph representation of data and designed an Expectation-Maximization (EM) framework. Experimental results proved that the suggested algorithm has the ability to detect overlapping clusters with semantically related sentences. When the qualitative, quantitative and scientific methods were integrated with the suggested algorithm, the performance of the suggested algorithm would have been increased. Lei, et al. [11] suggested a sentiment-based rating prediction method (RPS) to avoid the wrong recommendations from the textual reviews. In this RPS, three major works were performed to tackle the recommended problems. First of all, mined the sentiment words and sentiment degree words from the user reviews by a social user sentimental measurement method, secondly, considered the interpersonal sentimental influence with own sentimental attributes that displayed the product reputation words. Finally, predicted a rating of a product with incorporated the 3 aspects as sentiment similarity of user, interpersonal sentimental influence, and the product's reputation similarity. The rating prediction accuracy was the major issues in this recommendation system. Sun, et al [12] proposed a novel privacy preserving proximal support vector machine technique for classifying the vertical partitioned data. The suggested algorithm provided a fast and simple generation of the linear and non-linear classifier. The performance of the proposed technique was validated with the existing privacy-preserving SVM classifier for the metrics such as classification accuracy and computation time. The validation results proved that the suggested approach produced optimal results for all the metrics. Singh et al [13] explored the new scheme called SentiWordNet based scheme in two levels such as document and aspect level. The utilization of linguistic features supported the effective document-level classification. The proposed SentiWordNet scheme located the opinionated text around the feature and computed the sentiment. Esparza, et al. [14] explored the fragmented noisy snippets that were directly used in recommendations. The validation of whether the Real-Time Web (RTW) services were used as the basis for the recommendation and the performance with the traditional systems. They developed the product recommender system based on the Twitter-like comments and showed the better recommendation performance compared to the CF approaches. The relationship between the web services and the providers described by the two-dimensional form called the user-item matrix. But, the matrix model cannot reveal that relationship accurately orientation. The identification of preference similarity among the reviewers was the major issue in the review-based recommendation. Verma, et al [15] analyzed six clustering algorithm such as K-means clustering, hierarchical clustering, DBScan clustering, density based clustering, optics and Expectation Maximization (EM) algorithms for clustering in the data mining. The comparison results proved that the K-means algorithm has increased Root Mean Square Error (RMSE) value, speed and optimal performance than the hierarchical algorithm. The EM and K-means algorithms were efficient in handling large datasets. Further, it analyzed that the hierarchical clustering algorithm was sensitive to noise. Kannan, et al [16] proposed an effective quadratic effective fuzzy c-means clustering algorithm with a quadratic term, mean distance function, kernel distance function and regularization function for reducing the proposed objective function. As the suggested clustering algorithm performed the clustering using cluster center the numbers of iterations were reduced. The number of clusters and the cluster validity was chosen using the silhouette method. The disadvantage of the proposed clustering algorithm was increased time complexity. Recamonde, et al [17] analyzed the distributed classification task with vertically partitioned data. The preferences of the agent were ordered using agent based classification system. The advantages of the suggested approach were simplicity, lack of demand for transfer of a large volume of data, parameter setup and preservation of data privacy. Wang et al [18] utilized the medium called Solid State Drive (SSD) that replaced the Hard Disk Drive (HDD) in their infrastructure point of view. The property of SSD was casual access potential alternate to the consecutive access potential. They conducted the series of experiments governing SSD effect on the search engine cache management. The grouping of similar services together in a single cluster increased the data size in the large factors. Besides, they compared the proposed SDD with the Static-Query Result Cache (S-QRC) and Dynamic Query Result Cache (S-QRC) in terms of hit ratio and average query latency. Salehan and Kim [19] offered the sentiment mining method for analyzing the big data of the online consumer reviews (OCR). In this method, classified the consumer reviews into positive sentiment and the neutral polarity of the text. The readership and helpfulness were majorly affected by the length and longevity of a review positively. Therefore, the offered method was provided successful sorting and classification results of the big OCR data. But, the classifier accuracy rate was the major requirement of the online reviews. X.Zheng,et al[20] elucidated a novel approach for searching the destination of tourist via online which has been difficult task due to its restrictive factors. To solve this issue, recommender system could be placed for helping the user to set information overload, still there might be occurrence of cold start problem and fewer recommendation accuracy. Here the situation calls for best recommender system that utilizes opinion-mining technology for filtering user sentiment, and also temporal dynamics were utilized for signifying the destination popularity drifting over time and user preference. Through linking user sentiment and temporal influence, the elements were merged with SVD++ method. A well-known recommendation approaches were taken to analyze the performance of this framework and it attained better recommendation accuracy and excellence. A sequential experimental evaluation on publicly available dataset was made to prove the superiority of this method than the other methods. S.I.Nikolenko et al [21] elucidated an innovative method using end-to-end Aspect-based Rating Prediction model where the user's review was estimated on the other hand for elucidating the profile users or predictions, a coherent aspects of reviews was also discovered. This model only utilizes max margin losses for combined item which was one of the advantage to enhance the performance including dual-headed architecture and user embedding learning. This framework outperformed well than other state-of-the-art models like HFT, NARRE, DeepCoNN, and TransRev with two real world datasets. Here the role of aspect embedding's employed in recommender system were also exposed qualitative inspection of the aspects and quantitative evaluation of rating prediction models. J.Shokeen et al[22] described the idea of social recommender system, where the recommender system was integrated with social networks to generate a new social recommender system. The different features were examined to analyze its importance towards generating the recommendations effectively, better recommendations were reliant on the role of features. Still there was a lack in social recommender system quality, in future author will have a deep investigation to find the way of improving the quality through resolving the issues in conventional work.

III. PROPOSED METHODOLOGY

EFCFM: ENHANCED FEATURE SPECIFIC COLLABORATIVE FILTERING MODEL

This section discusses methodology of the proposed an Enhanced Feature specific Collaborative Filtering Model for Aspect Opinion and Temporal Based Product Recommendation about the product reviews through the implication of collaborative filtering with the opinion score estimation technique. The workflow of the proposed EFCFM is shown in Fig. 2. Initially, the user's reviews about the products are taken as an input. Here, the stop word removal operation is performed as preprocessing step to eliminate all stop words in the user's reviews. But the stop words that contains negative polarity statements are excluded in this techniques which helps in the better analysis of the review. From these preprocessed review, the keywords are extracted which represents the polarity or features of the products. Additionally, the features of the product's property are also extracted. Based on their polarity using EFCFM is implemented the reviews are grouped as a feature specific clusters. After that the polarity detection is done and the products' overall opinion can be determined using time based Opinion Score Estimation Algorithm. Finally the user query is taken and the query is searched based on the location and time specified by the users. This makes the search more easier and quick to access. Then the query is pre-processed and the requirement of the user is analyzed by the extracted keywords. The overall opinion about the product is estimated and checks whether the query result is in the Quick Access Memory (QAM). If so, the feature based polarity score is estimated and recommends the query result. Else, the query result is obtained from the knowledge base and the feature based polarity score is estimated. Finally the best products are recommended accurately according to the user query.

Table 1. *NOTATIONS AND DESCRIPTION*

List of Notations	Description
$\mathrm{Wd_L}$	Word list
$\mathrm{P_{Sr}}$	Positive Score
$\mathrm{N_{Sr}}$	Negative Score
N	Number of User
R	Reviews with sentences or paragraphs
M	number of aspects represent A_M from $U_R(i)$
U_L	User index
Kw_x	List of keywords
U_R	User Reviews
C_{id}	Cluster Indices
C_W	Clustered Keywords
Sn_{wd}	SentiWordnet dictionary words
NU_R	Number of U_R
AO_p	Aspect-Opinion Pair
F_{pos}, F_{neg}	Positive and negative features
$Score_{StrengthFeat}$, $Score_{WeakFeat}$	Polarity Scores

AO_p	Aspect-Opinion Pair (A_M, O_p)	O_p	opinion

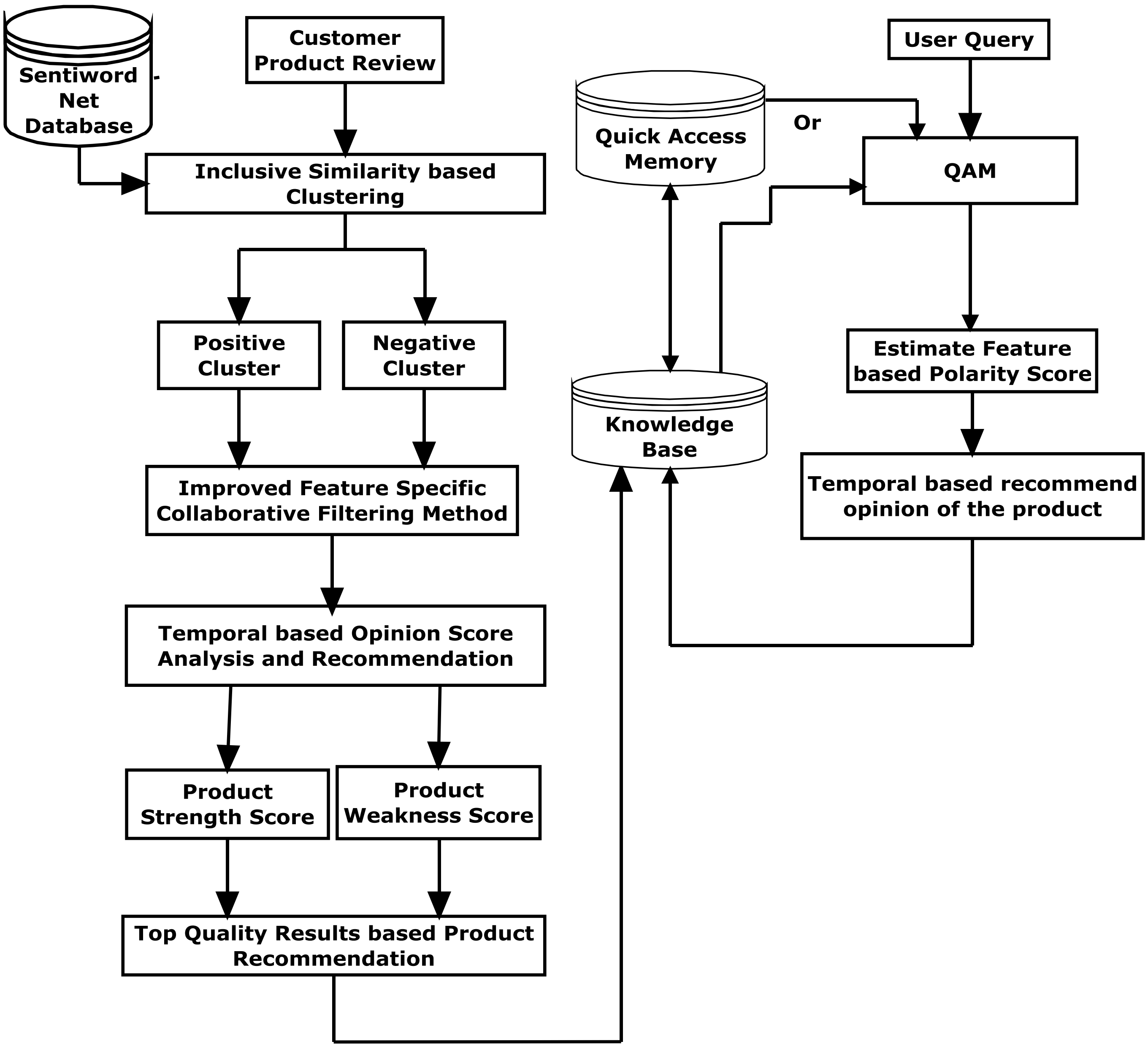


Fig.2 *Workflow of the proposed EFCFM*

A. INCLUSIVE SIMILARITY BASED CLUSTERING (ISC)

Generally, the user's reviews are obtained as the form of raw text data. This text data can be first preprocessed using stop word removal technique to recommend the best products accurately based on the reviews. The raw text data may contain negative sentences and comparative sentences which makes the analysis of the user review as a challenging task. For that some words such as "Wh" words and verbal words are removed using stop word removal technique. Further, the spaces and special characters in the raw text data can be removed in this preprocessing step as described in [23]. After that, the words are categorized based on the parts of the speech (POS) such as Noun, Verb, Adverb, Adjective, etc. After the words from the user's review are preprocessed and categorized, the keywords from those words are extracted. These keywords are extracted from each category and assigned a tag for each keywords using POS tagger. This POS tagger is the computer readable program which plays a vital role in opinion mining. Using these tags the keywords from the user's review data are extracted. The extracted keywords are validated with SentiWordNet database that displays two numerical scores ranging from 0 to 1, which is representing synset's positive and negative bias polarity. Thus, the positive and negative polarity based keywords Sn_{wd} are split. Then, the user reviews (U_R) the polarity and the product features are given to the Inclusive Similarity-based Clustering algorithm. At last, ISC algorithm is applied to cluster where the most similar keywords based on the adjacency matrix. Then combine the similar index words and illustrates of positive and negative based polarity clusters.

B. IMPROVED FEATURE SPECIFIC COLLABORATIVE FILTERING BASED CLUSTERING MODEL:(IFSCF)

The extraction of related features based opinions for the belonged product is calculated from this proposed IFSCF model. It is used to estimate the strength of the product and percentage of weakness efficiently. The user's aspects and the corresponding opinion of the product is retrieved by giving the cluster words and the user's reviews to the IFSCF model. Then the product strength and weakness score are estimated. Adjust the cluster indices, clustered keywords, and the SentiWordNet dictionary words with the set of words based on polarity score for every word. For every reviews i, the condition is applied in the clustered keywords which has to fulfill the subsequent cases in step 3. When the clustered keyword denotes the SentiWordNet's positive word, it has to be assigned as P_{cw}, or else it has to be named as N_{cw}. Then condition is applied in the j of the user reviews denotes in step 9. Then categorize the aspects and its corresponding opinion of the product based on the following cases. Initialize the M number of aspects represent A_M from $U_R(i)$ and the opinion O_p be the set of opinions that corresponds to the A_M. Then, the aspect–opinion pairs are termed as the AO_p. If the opinion is said to be positive then update the aspect-opinion term as strength category or it is called as weak category as,

$$AO_{StrengList} \leftarrow (A_M(k), O_p(k)) \quad (1)$$

$$AO_{WeakList} \leftarrow (A_M(k), O_p(k)) \quad (2)$$

Then, continue this process until reach the overall user reviews. After calculating the strength score value as,

$$Score_{Strength} = \frac{AO_{StrengList}\left(Count\ (A_M) + Count\ (O_p)\right)}{Size(AO_{StrengList}) + Size(AO_{WeakList})} \quad (3)$$

Convert the numeral strength score value of $AO_{StrengList}$ into the product percentage strength value as,

$$SS_p = Score_{Strength} * 100; \quad (4)$$

Afterwards, the same procedure is followed in the negative score estimation process. To calculate the weakness score of the product as,

$$Score_{weakness} = \frac{AO_{WeakList}\left(Count\ (A_M) + Count\ (O_p)\right)}{Size(AO_{StrengList}) + Size(AO_{WeakList})} \quad (5)$$

$$WS_p = Score_{weakness} * 100; \quad (6)$$

Finally, predict the overall percentage score of strength and weakness value of the product based on the user reviews [24]. The algorithm for an improved feature specific collaborative filtering based clustering model is illustrated as follows:

Improved Feature specific Collaborative filtering Model
Input: cluster Words, User Reviews (U_R) **Output**: Aspect-Opinion pairs, Strength and Weakness Score
Step 1: Initialize C_{id} and C_W **Step 2**: Initialize Sn_{wd} with polarity score value **Step 3**: For i=1 to Size (C_W) **Step 4**: If (C_W(i) in $Sn_{wd}(pos)$ **Step 5**: Add $C_W \rightarrow P_{cw}$ Else **Step 6**: Add $C_W \rightarrow N_{cw}$ **Step 7**: End if **Step 8**: End for i **Step 9**: For j=1 to N **Step 10**: Initialize A_M , O_p **Step 11**: Assign $AO_p \leftarrow$Aspect-Opinion Pair (A_M, O_p) **Step 12**:For k= 1 to size (A_M) **Step 13**: If (O_p $is\ positive$) **Step 14**: Update AO_p as strength by using the equation (1) Else **Step 15**: Update AO_p as weak by using the equation (2) **Step 16**: End if **Step 17**: End for K; **Step 18**: End for j **Step 19**: Compute Strength Score by using the equation (3) **Step 20**: Estimate the product strength percentage value by using equation (4) **Step 21**: Compute Weakness Score by using the equation (5) **Step 22**: Estimate the product weakness percentage value by using equation (6)

C. QUICK ACCESS MEMORY(QAM)

The Quick Access Memory (QAM) [25] is utilized to check the matching the user query along with the memory of cache. Retrieval of the results from cache memory is happened when it's matched otherwise result is retrieved from the knowledge database. The process time and execution time are minimized according to the optimal memory access. The reviews of multiple user as well as the clustered words are specified to the model of preprocessing in order to extract the appropriate key features. It decreases the repetitive words and thus leads to the reduction in time complexity. The product query and the user reviews are collected together in parallel in order to achieve the cache memory efficiently. Then, the huge size phrases are divided into the plentiful segments. The stop words removal and the relevant keywords extraction are the succeeding steps in the processing of user query. The outcomes from the query processing are matched with the cache memory contents. Then, the available relevant results are verified in the cache memory. If they are existing, then the results are extracted from the cache memory or else from the database, results are extracted efficiently. At last, feature based opinion prediction as well as the score approximation process are recommended.

D. OPINION SCORE ANALYSIS BASED RECOMMENDATION

Finally, the opinion score based product recommendation [24] is carried out after calculating the score values for both the user review and product property. Here a product is recommended based on the opinion score based analysis. For that, the user query, temporal selection of the product property by the user is taken as an input. Let 'Q_U' be the user query. The keywords are extracted from the query of the user and those extracted keywords are represented as Q_K. Let Q_{Pr} be the specific property for which the user wanted to search and Q_T be the Temporal Feature by which the user's response has to be filtered. If a user wanted to search a product, first it checks whether the property of that product is in the extracted keywords. If so, it computes the overall opinion score using the following equations:

Let T_C be the total Number of user Reviews

$$\text{Positive Score } (P_s) = \frac{T_C}{T_P} * 100 \quad (9)$$

$$\text{Negative Score } (N_S) = \frac{T_C}{T_N} * 100 \quad (10)$$

The overall opinion score is calculated and the positive and negative scores are obtained by the equations (9) & (10). With this overall opinion score a decision rule is made to determine the recommendation of the product. Similarly, the opinion score for the features specific polarity based products. Let 'r' be the set of User list and it checks temporal features, text features and product's property. If so, the strength and weakness of both the features of the user review as well the product property for all the polarity can be calculated using the following equations.

$$Str_F = Str_F + Str_{Score}(h) \quad (11)$$

$$wk_F = wk_F + weak_{Score}(h) \quad (12)$$

$$Pr.Str_F = Pr.Str_F + Prop_{Strength}(h) \quad (13)$$

$$Pr.Wk_F = Pr.Wk_F + Prop_{Weakness}(h) \quad (14)$$

If the polarity of the user query is positive, then update the positive score and if the polarity of the user query is negative then update the negative score. After that the positive and negative scores are calculated as follows:

$$\text{Positive Score } (P_s) = \frac{P_{Score}}{N(U_r)} * 100 \quad (15)$$

$$\text{Negative Score } (N_s) = \frac{N_{Score}}{N(U_r)} * 100 \quad (16)$$

Algorithm IV - Opinion Score analysis based Recommendation	
Input:	User Query, Temporal Selection of Product Property by the user
Output:	Collaborative Score Based recommendation status of the product
Step 1:	Initialize U_Q
Step 2:	U_k←pre-process (U_Q) // where U_k is the keywords in the user Query
Step 3	Asp_Q←Extract (Aspects,U_k)
Step 4:	if ($Size(Asp_Q) > 0$)
Step 5:	Initiate aspect specific Recommendation
Step 6:	Let AO_p be the Aspect –opinion pair
Step 7:	Sub_AO_p←Similar (AO_p, Asp_Q)
Step 8:	F_{pos}←positive features (Sub_AO_p)
Step 9:	F_{neg}←Negative features (Sub_AO_p)
Step 10: Step 11: Step 12: Step 13: Step 14: Step 15: Step 16: Step 17:	$Score_{StrengthFeat}$ = Polarity Score (F_{pos}) $Score_{WeakFeat}$ = Polarity Score (F_{neg}) apply Score Based ranking to $Score_{StrengthFeat}$ and $Score_{WeakFeat}$ $Strength_{Score} = \frac{\sum_{i=1}^{N} Score_{StrengthFeat}}{Size(Score_{StrengthFeat})}$ $Weakness_{Score} = \frac{\sum_{i=1}^{N} Score_{StrengthFeat}}{Size(Score_{StrengthFeat})}$ $Results = \begin{cases} Recommended & if\ (Strength_{Score} > Weakness_{Score}) \\ Not\ Recommended & Other\ Wise \end{cases}$ else Initialize Overall Product Recommendation
Step 18:	$U_r > Q_T$(reviews given after the temporal selection)
Step 19:	For h=1 to r
Step 20:	Compute the overall opinion score using equations (11) to (14)
Step 21:	If (polarity (h) is positive)
Step 22::	Increment P_{Score}
Step 23:	Else (polarity (h) is negative)
Step 24:	Increment N_{Score}
Step 25:	End if
Step 26:	End for h
Step 27:	F_{pos}←positive features (AO_p)
Step 28:	F_{neg}←Negative features (AO_p)
Step 29:	Follow Step 10 to 15
Step 30:	End if

Initializing AO_p and U_Q to find the overall opinion score and recommendation. Then the U_Q requested query is pre-processed to extract the relevant query and remove the unwanted informations. Then the extracted features is considered into aspects and its opinion to categorize the positive features and the negative features. Then estimate the polarity score of positive and negative features to retrieve the rank of the scores. Finally, establish the strength and weakness score of the product and then show the suggested product is recommended or not. After calculating the overall opinion score and the positive and negative scores of the user query, the decision is made based on the opinion score to determine the recommendation of the product.

IV. PERFORMANCE ANALYSIS

This section demonstrates the performance analysis of the proposed Enhanced Specific Feature Collaborative Filtering for Aspect opinion and temporal based recommendation. The experiment is tested with Java and NetBeans tool. The JAVA Net beans offers a framework for management, storage and retrieval of large amount of data. The dataset that has the customer reviews of five products are collected. The five products are Canon G3 (digital camera), Nikon coolpix 4300 (digital camera), Nokia 6610 (cellular phone), creative labs Nomad Jukebox Zen Xtra 40 GB (Mp3 player) and the Apex AD2600 progressive–scan (DVD player). The effectiveness of this proposed technique is analyzed using this dataset and are evaluated with parameters like Precision, Recall, F-Measure, RMSE and MAE. Also the performance of the proposed techniques is compared with several existing techniques.

A. Evaluation Metrics

Precision:

The precision which is referred to as the positive predictive rate is the proportion of retrieved reviews that are appropriate to the user query. It can be deliberated as,

$$P = \left\{\frac{\{relevant\ reviews\ \cap retrieved\ opinions\}}{retreived\ opinions}\right\} \quad (17)$$

Recall:

A recall in data recovery is the proportion of the user reviews that are appropriate to the successfully retrieved user query.

$$R = \left\{\frac{\{relevant\ reviews\ \cap retrieved\ user\ query\}}{relevant\ user\ query}\right\} \quad (18)$$

F-Measures:

The f-Measure is well-defined as the biased mean of the precision and recall of the opinion calculation based on the user query.

$$F - M = \frac{2 \times P \times R}{P+R} \quad (19)$$

RMSE:

The RMSE which is also termed as the root mean square deviation (RMSD) is a repeatedly utilized amount of the dissimilarity amongst the values attained by a system and the values that are actually perceived from the background that is being exhibited. These distinct dissimilarities are also termed as residuals, and the RMSE functions to combine them into a distinct amount of predictive power.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{exp,i} - X_{est,i})^2}{n}} \quad (20)$$

MAE:

The MAE measures the normal size of the mistakes in an arrangement of expectations, without considering their behavior. It is normal for the test of the total contrasts amongst expectation and perception where every individual dissimilarity have the level of weight.

$$MAE = \frac{\sum_{i=1}^{n}|X_{exp,i} - X_{est,i}|}{n} \quad (21)$$

Where X_{exp} represents the experimental results and X_{est} represents the estimated variable at time i.

B. Results and Discussion

The performance metrics such as Precision, Recall, F-Measures, RMSE and MAE are evaluated using Customer Review Dataset (CRD) for the proposed EFCFM model. Table 2.shows the RMSE and MAE values for the proposed model.

Table 2. *RMSE and MAE values*

Customer Review Dataset	RMSE	MAE
Digital camera1	0.6	0.74
Digital camera2	1.1	0.61
Phone	0.99	0.75
Mp3 Player	0.71	0.54
DVD Player	0.54	0.63

This table shows the error rate for various products that are obtained from the CRD dataset. From the table, it is observed that the proposed EFCFM provides reduced average error rate.

1) RMSE, MAE Measures

The performance of all collaborative filtering methods in terms of MAE and the RMSE values are validated with the customer review datasets.

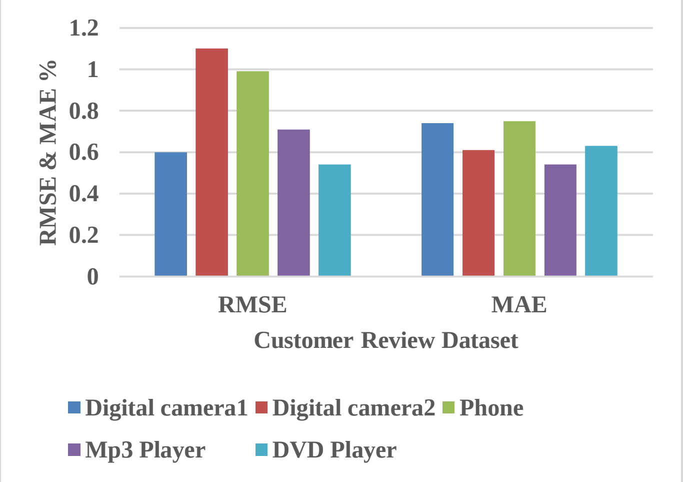


Fig. 2 *Performance analysis of RMSE and MAE for the proposed EFCFM approach.*

Fig. 2 shows the graphical representation of RMSE and MAE for the proposed EFCFM model. For various datasets, our proposed EFCFM technique provides reduced error rate which resulted that the products based on the user review are recommended more accurately

2) Precision, Recall and F measure:

The precision, recall and F measures are calculated for proposed EFCFM technique with different range of sparsity. At sparsity level of 0.2 to 1.0, the precision recall and F measure are evaluated for various techniques to analyze the performance of the proposed EFCFM and how much efficient it is when compared to other methods. The different experimental analysis is performed in the proposed EFCFM model to tackle the data sparsity. The proposed EFCFM model precision, recall, and the f-measure values are compared with the different sparsity levels between “0.2 to 0.4”, “0.4 to 0.6”, “0.6 to 0.8” and “0.8 to 1.0”.

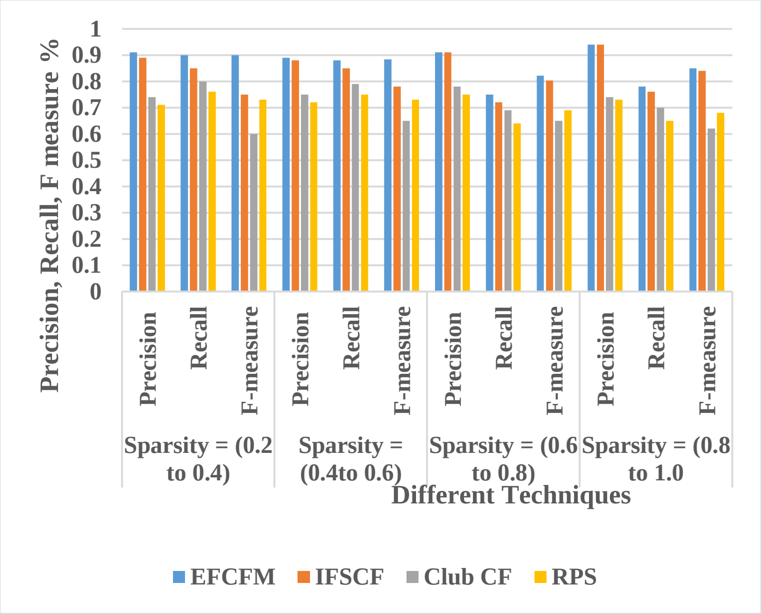


Fig. 3 *Precision, Recall, F-Measures of the proposed EFCFM technique*

Fig. 3 shows the comparative results of the precision, recall and F-Measures for the basic CF and EFCFM technique [5]. From the figure it is observed that at the sparsity level of 0.8 to 1.0, the proposed technique provides increased precision value which indicates that the positive predictive rate is high. This results that the highly accurate recommendations are provided for the user

3) Average Precision:

The average precision that shows the rating prediction for both the existing and proposed technique is given in the table 3. The proposed EFCFM method is analyzed and compared with the existing RPS [11] ClubCF techniques and IFSCF. This uses various datasets such as Movie, SFU, Yelp [11] and CRD for analyzing the average precision rate.

Table 3. *Average Precision*

Dataset	EFCFM	IFSCF	Club CF	RPS
Movie	76.5	75.1	71.04	72.7
SFU	77.7	76.6	72.1	73.5
Yelp	90.2	89.2	79.2	87.1
CRD	86.5	85.3	80.1	82.5

In the preceding figure Fig. 4 shows the comparative analysis of the average precision for both the existing and proposed technique. For various datasets, the average precision for the proposed technique provides better results than the existing technique. For CRD dataset proposed EFCFM technique provides 86.5 average precision value but the existing technique provides only 82.5 average precision value. Form this analysis, it is observed that the proposed techniques offers high positive prediction rate than the existing technique.

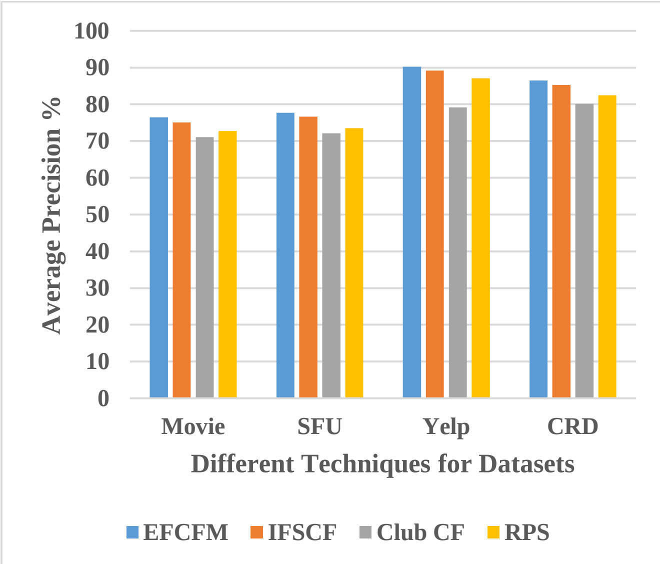
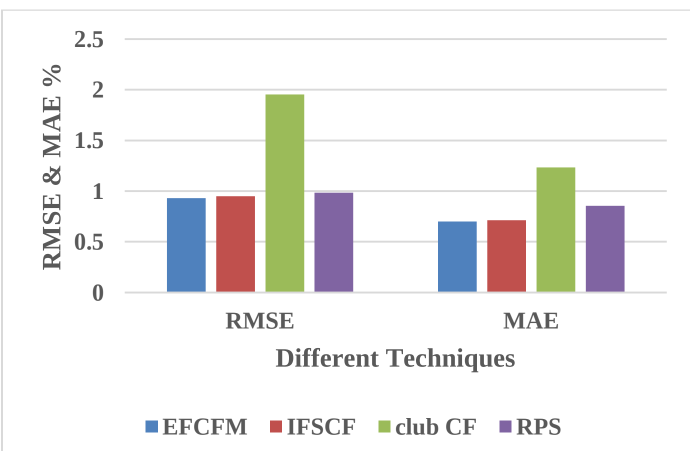
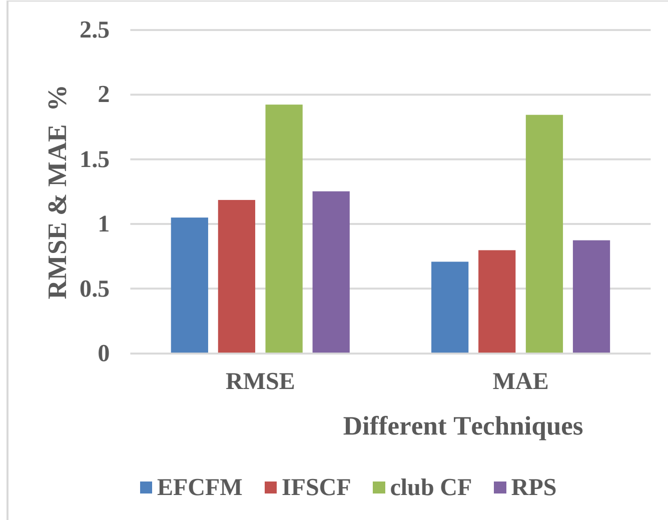


***Fig. 4** Average precision of the proposed EFCFM technique*

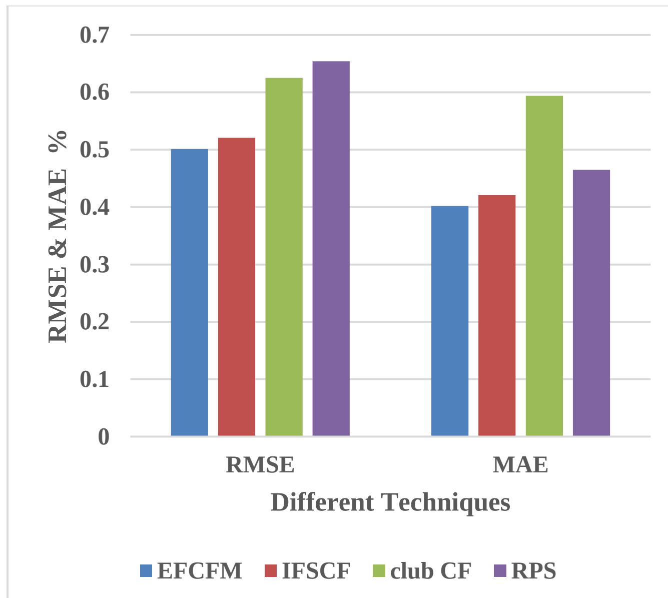
4) Error Value Prediction for Different Techniques:
The RMSE and MAE values are evaluated for EFCFM and existing methods and it also shows the comparative results for various existing techniques and proposed technique. Several existing techniques [26] such as IFSCF, Club CF and RPS are compared with the proposed EFCFM.



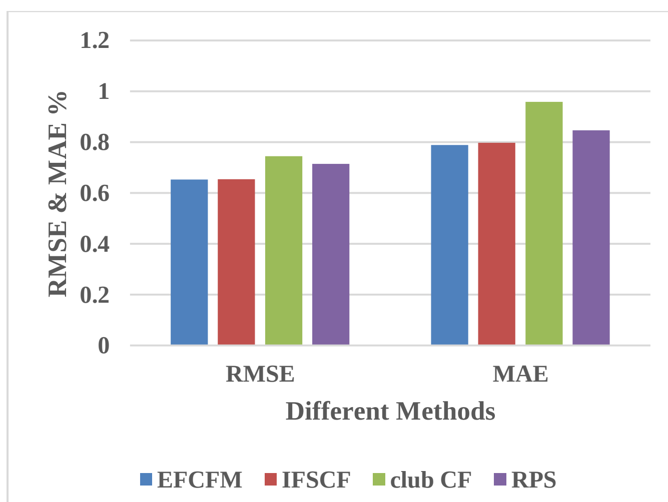
(a)



(b)



(c)



(d)

Fig. 5 (a) the MAE and RMSE value of Yelp 2013, (b) the MAE and RMSE value of Yelp 2014, (c) the MAE and RMSE value of Opinions dataset, (d) the MAE and RMSE value of the proposed CR dataset. Fig. 5 (a) shows the MAE and RMSE value of Yelp 2013, (b) represents the MAE and RMSE value of Yelp 2014, (c) represents the MAE and RMSE value of Epinions dataset, and (d) represent the MAE and RMSE value of the proposed CR dataset. The proposed EFCFM model is compared with the various existing collaborative filtering models and the recommendation models. Hence, the proposed EFCFM model increases the efficiency by reducing the error rate successfully.

V. Conclusion and Future Work

In opinion mining and recommendation system, there exist several issues such as false prediction of user review and inappropriate recommendation of products. To avoid these drawbacks, a novel technique called an Extended Collaborative Filtering based Clustering Mechanism is introduced in web service recommendation systems. The main purpose of this proposed technique is to recommend the best product based on the customer reviews accurately to the users. In this proposed work, a raw text data can be initially preprocessed to remove the redundant words and spaces using stop word removal approach and the keywords are extracted using POS tagger. After that the features of the keywords are extracted which is the major challenging task. Here the features of the user review along with the property of the products are extracted. Then the polarity estimation is carried out. Using these scores, the strength and weakness of the user review and the product's property are also evaluated using EFCFM technique. Finally, the overall opinion score based recommendation of the best product is obtained. The performance of the proposed EFCFM technique is evaluated with the metrics like RMSE, MAE, Precision, Recall and F-Measures. These parameters are also compared with various existing techniques. The precision value for the CRD dataset provides 4.62% improvement compared to that of the existing technique. This results that the proposed techniques may afford more accurate product recommendation. Also the RMSE and MAE values of proposed technique provides 7.88% and 1.87% reduction than in existing LDA technique. From the results it is observed that the proposed EFCFM technique provides accurate recommendation of the best products with respect to the user queries. This proposed method helps to obtain the services based on the product reviews. In future, this work will try to obtain the multi services based on the multi domains such as google reviews for schools, colleges, hospitals, etc.

VI. References

[1] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: a survey," Decision Support Systems, vol. 74, pp. 12-32, 2015.

[2] P. B. Thorat, R. Goudar, and S. Barve, "Survey on collaborative filtering, content-based filtering and hybrid recommendation system," International Journal of Computer Applications, vol. 110, pp. 31-36, 2015.

[3] H. Wang, N. Wang, and D.-Y. Yeung, "Collaborative deep learning for recommender systems," in Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, 2015, pp. 1235-1244.

[4] P. Riyaz and S. M. Varghese, "A scalable product recommendations using collaborative filtering in hadoop for bigdata," Procedia Technology, vol. 24, pp. 1393-1399, 2016.

[5] M. K. Najafabadi, M. N. r. Mahrin, S. Chuprat, and H. M. Sarkan, "Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data," Computers in Human Behavior, vol. 67, pp. 113-128, 2017.

[6] X. W. Zhao, Y. Guo, Y. He, H. Jiang, Y. Wu, and X. Li, "We know what you want to buy: a demographic-based system for product recommendation on microblogs," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014, pp. 1935-1944.

[7] G. Ling, M. R. Lyu, and I. King, "Ratings meet reviews, a combined approach to recommend," in Proceedings of the 8th ACM Conference on Recommender systems, 2014, pp. 105-112.

[8] S. Yang, M. Korayem, K. AlJadda, T. Grainger, and S. Natarajan, "Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach," Knowledge-Based Systems, vol. 136, pp. 37-45, 2017.

[9] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: understanding rating dimensions with review text," in Proceedings of the 7th ACM conference on Recommender systems, 2013, pp. 165-172.

[10] A. Skabar and K. Abdalgader, "Clustering sentence-level text using a novel fuzzy relational clustering algorithm," IEEE transactions on knowledge and data engineering, vol. 25, pp. 62-75, 2013.

[11] X. Lei, X. Qian, and G. Zhao, "Rating prediction based on social sentiment from textual reviews," IEEE Transactions on Multimedia, vol. 18, pp. 1910-1921, 2016.

[12] L. Sun, W.-S. Mu, B. Qi, and Z.-J. Zhou, "A new privacy-preserving proximal support vector machine for classification of vertically partitioned data," International Journal of Machine Learning and Cybernetics, vol. 6, pp. 109-118, 2015.

[13] V. K. Singh, R. Piryani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification," in 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013, pp. 712-717.

[14] S. G. Esparza, M. P. O'Mahony, and B. Smyth, "Mining the real-time web: a novel approach to product recommendation," Knowledge-Based Systems, vol. 29, pp. 3-11, 2012.

[15] M. Verma, M. Srivastava, N. Chack, A. K. Diswar, and N. Gupta, "A comparative study of various clustering algorithms in data mining," International Journal of Engineering Research and Applications (IJERA), vol. 2, pp. 1379-1384, 2012.

[16] S. Kannan, S. Ramathilagam, and P. Chung, "Effective fuzzy c-means clustering algorithms for data

clustering problems," Expert Systems with Applications, vol. 39, pp. 6292-6300, 2012.
[17] M. Recamonde-Mendoza and A. L. C. Bazzan, "Social choice in distributed classification tasks: Dealing with vertically partitioned data," Information Sciences, vol. 332, pp. 56-71, 3/1/ 2016.
[18] J. Wang, E. Lo, M. L. Yiu, J. Tong, G. Wang, and X. Liu, "Cache design of ssd-based search engine architectures: an experimental study," ACM Transactions on Information Systems (TOIS), vol. 32, p. 21, 2014.
[19] M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics," Decision Support Systems, vol. 81, pp. 30-40, 2016.
[20] X. Zheng, Y. Luo, L. Sun, J. Zhang, and F. Chen, "A tourism destination recommender system using users' sentiment and temporal dynamics," Journal of Intelligent Information Systems, vol. 51, pp. 557-578, 2018.
[21] S. I. Nikolenko, E. Tutubalina, V. Malykh, I. Shenbin, and A. Alekseev, "AspeRa: Aspect-based Rating Prediction Model," arXiv preprint arXiv:1901.07829, 2019.
[22] J. Shokeen and C. Rana, "A study on features of social recommender systems," Artificial Intelligence Review, pp. 1-24, 2019.
[23] J. Sangeetha and V. S. J. Prakash, "An efficient inclusive similarity based clustering (ISC) algorithm for big data," in 2017 World Congress on Computing and Communication Technologies (WCCCT), 2017, pp. 84-88.
[24] J. Sangeetha and V. S. J. Prakash, "Improved Feature-Specific Collaborative Filtering Model for the Aspect-Opinion Based Product Recommendation," in Advances in Big Data and Cloud Computing, ed: Springer, 2019, pp. 275-289.
[25] J. Sangeetha, V. S. J. Prakash, and A. Bhuvaneswari, "Dual Access Cache Memory Management Recommendation Model Based on User Reviews," 2017.
[26] Z. Ren, S. Liang, P. Li, S. Wang, and M. de Rijke, "Social collaborative viewpoint regression with explainable recommendations," in Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017, pp. 485-494.