

**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH  
TECHNOLOGY****BOOK RECOMMENDATION USING K-MEAN CLUSTERING AND  
COLLABORATIVE FILTERING****Ritu Rani<sup>\*1</sup> and Rahul Sahu<sup>2</sup>**<sup>\*1</sup>Research scholar, Department of Computer Science & Engineering, LNCT, Bhopal ,  
India<sup>2</sup>Assistant Professor and, Department of Computer Science & Engineering, LNCT, Bhopal, India

DOI: 10.5281/zenodo.1042100

**ABSTRACT**

With the increase in demand of items amongst customer enhances the growth in information technology and e-commerce websites. This demand is increased due to the availability of web services **Personalized preferences** and **decision making** are generated in an application called Recommendation system using an **information filtering technique**. Web related services and applications through which searching and selecting becomes easy, the related use which are in demand are selecting **books and newspaper, best hotel and its location, ticket bookings like movies, flights, buses, trains etc.** Relevant features and related items are the characteristic on which this technique works. Suggestion of items, according to user preferences are most important, so suggestion according to similarities provides suitable recommendation. The working of recommendation system for administration have been researched in recent years. Network resources, clients, and administration are all connected and quickly developed. This method of recommendation system **works as suggestion, customization, learning, administration and this all provides user for the items suggestion and decision making.** In this paper , variety of **algorithms like k-mean clustering, collaborative filtering** are used for the information suggestion

**KEYWORDS:** Recommendation system, k-mean clustering, collaborative filtering.**I. INTRODUCTION**

We have lot of choices in the real world and selecting one out of many is the biggest problem. We always take help from some or the other person in choosing the one amongst many. Our family members, friends who have same preference like our suggest us. **But if the same thing happens in virtual world then there selection of interested item is based on recommender system**, which is very helpful in selecting the one amongst the many. **Through recommender system people can share preferences and most preferred items are offered among them to user from which he can select the interested one.** This technique deals with the **information overload problem.** Useful suggestions are created for customers to get the interesting items.

Book recommendation is created and deployed in this approach of work, which helps in recommending books. Recommendation achieved by the users feedbacks and rating, this is the online which analyze the ratings, comments and reviews of user, negative positive nature of comments using opinion mining. **User searching for the interested book will be displayed in top list and also can read feedback given by people about the book or any searched items.** Whenever user search for any book from the large data available, he gets confused from the number of displayed item, which one to choose. In that case recommendation helps and displays on the **interested items.** This is the trustworthy approach, which is used in this paper where selection is based on the **services of cloud using collaborative filtering.** Set of user requirements are searched in cloud services. From the targeted user it computes the similar preferences to provide the service of recommendation to each user. Sorting of services are done based on the degree of recommendation for each user. In recent years use of recommendation is increased because of the use of some techniques. This techniques displays the selected and preferred items on the top list, it helps in suggesting user, the relevant items which is the need of requirement to user.

Recommendation system can be used for items, for friends also on social sites, for selecting of hotels, music, book, application etc. here selection is difficult because number of data is available. The methodologies like

content-based and collaborative filtering are used here to predict from the available choices. On the basis of attribute of items content-based filtering works and on the basis of past behavior of user collaborative filtering works.

## II. METHODOLOGY

Methodology of our work depends upon the technologies used in this paper. These technologies are collaborative filtering and K-mean clustering algorithm. The proposed algorithm mainly involves K-mean clustering and CF, we first explain about these involved algorithm and then describes the specific purpose of the used algorithm.

### Collaborative Filtering

The most successful algorithm used in the recommender system is the Collaborative Filtering. A recommender system is the system of intelligence which is helpful for user in getting interested items. Collaborative filtering is based on the technique of information filtering and data mining. It suggest users on the basis of neighbor's preferences. And suffers with the parameter like deficient scalability and low accuracy. This method matches the people of similar taste and then on the basis of personalized recommendation, recommend the user.

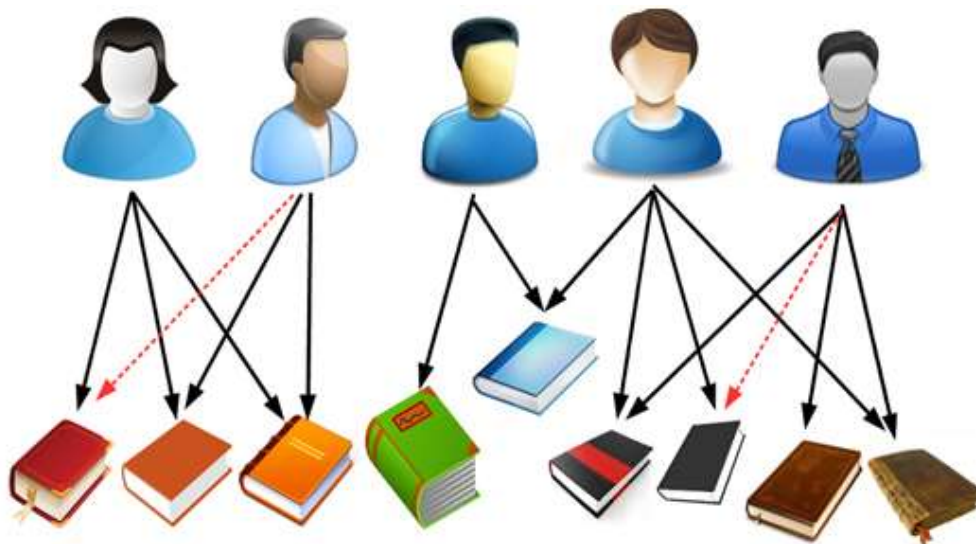


Figure 2.1. Collaborative Filtering for Book Recommendation

This algorithm is classified into two entities, the user entity and the item entity. The user entity works on the basis of rating, they rate the item according to their opinion about that item. Recommender system mainly uses collaborative filtering or the combination of it with other algorithm. It mainly focuses on user with same preference and taste and suggest items to them on the basis of selection of items by those users.

Collaborative Filtering is an algorithm which works on the basis of similarities in users choice. This algorithm can be explained as the choice of a person is relevant with the taste of his friend or belonging peoples. Recommendation is produced on the basis of neighbor's choices and taste. The neighbors are similar to the users which are the active user. This algorithm has two major step. In the first step computation is done between the similarities of active user and the most similar user and also the users in database. In second step recommendation is developed for those active users of first step.

Memory based and model-based are the two main types of this algorithm. Where, techniques of machine learning like Neural network and Bayesian network etc are used in Model-based algorithm.

Memory based algorithm generates recommendation using database of user-items. Each user has neighbor, and this neighbors are group of people with same taste. After it, prediction is generated for the items which are not rated by user. This algorithm mainly use two users or two items for the computation of similarities between them, after it recommendation for active user id generated.



### K-Mean Clustering

A clustering approach is used in our work, clustering is based on similarity where similar elements are kept in a single group. Likewise similar element, the irrelevant elements are also reside in a group, which is another group, based on similarity value or maximum size of cluster. The clustering approach which is used in our work is K-mean clustering for grouping of similar users. It is the unsupervised and simplest learning algorithm, which simplifies mining work by grouping similar elements forming cluster. This is done using a parameter called K-centroids. Distance between each element is calculated for checking the similarity and forming a single cluster to reside the similar elements, after comparing with K-centroid parameter.

K-mean is used for the analysis of cluster in data mining, it is a vector quantization method which is based on signal processing. K-mean clustering works as partitioning of observations into cluster form k. Every observation with the nearest mean belongs to cluster, it act as a prototype of cluster. K-mean clustering is used for the categorization of user on the basis of interest.

K-mean algorithm is proposed by Macqueen, in the year 1967. This algorithm is a simple and learning algorithm. The clustering algorithm used in data mining is k-mean clustering algorithm, it is popularly used in data mining for the clustering of large data. This algorithm works as selection of k objects which is the initial cluster center. Distance between each cluster center and object are calculated and is assign to nearest cluster. Average of all cluster is updated and the process repeats until starts functioning.

K-mean clustering is described as :-

Input data :

database,  $N = \{x_1, x_2, x_3, \dots, x_n\}$

n data objects and k number of cluster,

Output :-

1. From N datasets, k objects are randomly selected and cluster center ( $m_1, m_2, m_3, \dots, m_k$ ),
2. Distance between each object and cluster center is calculated and then each object is assigned to the nearest cluster.

### Proposed Work

Our proposed work involves working steps, which is as Data collected and preparation, clustering approach and recommendation approach.

#### Data Collection And Preparation

The proposed work implies that to maintain quality and authenticity of any system, information and data source are required. Collection of this data source is as important, data collection can be directly done from user or from any existing system. Primary data collection is done by survey and secondary data collection is from existing system or datasets. Primary data collection is a purpose of primary source. By using existing schema collection of data is possible, which consist of different categories and attributes. Then data cleaning is done on incomplete data from a data source. After it lemmatization module is developed and by using stop word removal quality of data source is improved. To break the entire work into a single value some schemes such as Tokenization and abbreviation removal is used, extraction is also done to get more accurate value. The complete approach works as getting most accurate answer for the book recommendation.

#### Clustering Approach

Similar elements are grouped on the basis of similarities, this approach of grouping a similar elements are called Clustering Approach. The similarity is based on the maximum size of cluster and relevant value. Different element who are having different values exist in another group. Here, K-mean clustering algorithm is used for grouping similarities of books based on similar user. In it work become easy for mining and classifying because it is a simple learning algorithm, which checks for similar elements using K-mean. Distance calculation is done by calculating distance between single cluster center and user. Quantification technique is also used in proposed architecture, which is used to convert relevant values into a matching score.

After it, the output which is generated is forwarded for making cluster of similar users. The complete work signifies that the similar users are based on the relevant distance. Below figure shows the outcome and working.

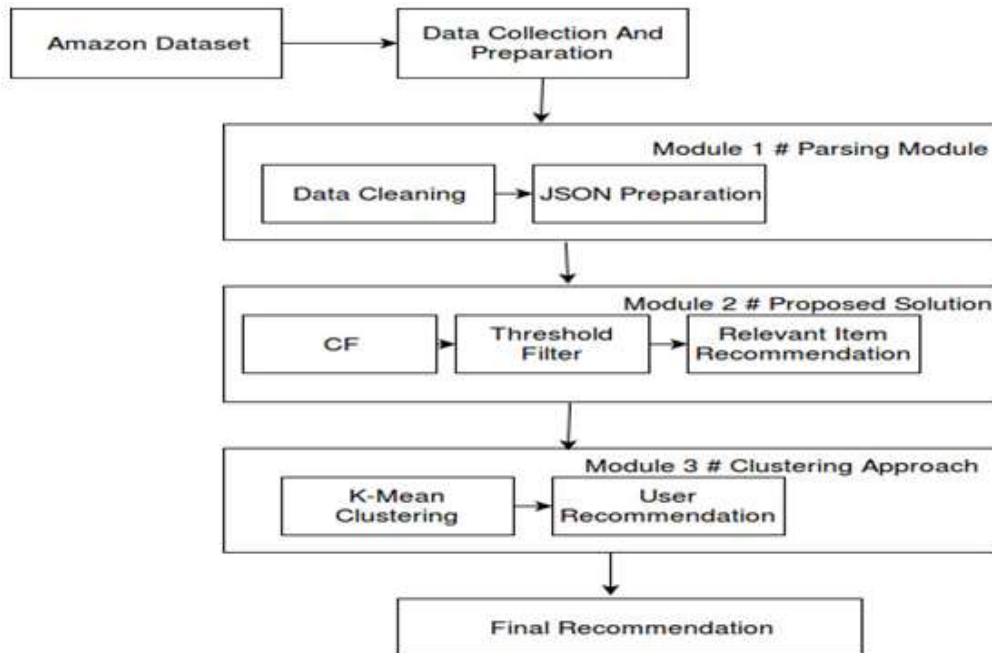


Figure 2.2 Cluster making

### Recommendation Approach

An approach defining filtering which works on the basis of rating and scoring. Mainly scores are used for predicting the frequency, frequency of items appearing which depends on similarity and preferences. The system of recommendation is now-a-days becoming more popular because of the use of searching items, many users are there who prefer it for books, articles, newspapers, different products and items etc. A recommendation for customized algorithm with K-mean clustering approach is used in order to convey the best recommendation and similar solution. Moreover, cluster recommended as input of data source through which similar score is calculated.

In simple words it can be indicated as maximum similarity scores more closeness. Threshold filter retrieves user id and most close user recommend. High threshold represent most close value. Whereas, low threshold represents high number of user.

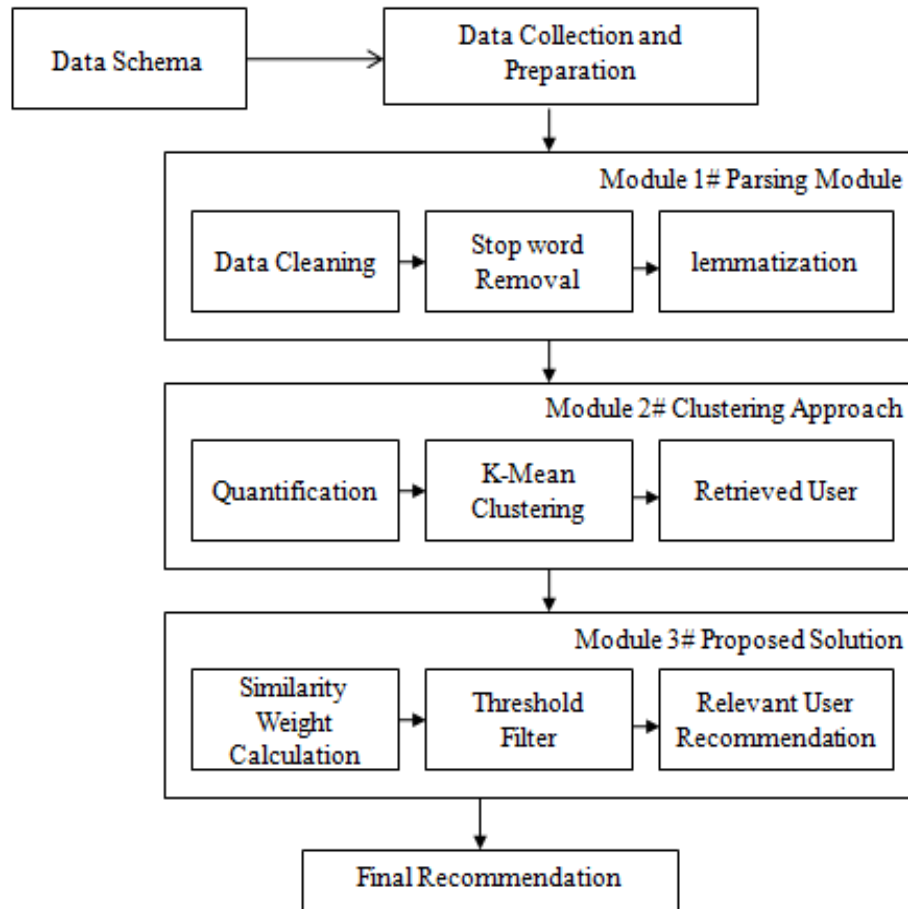


Figure 2.3 Proposed architecture

### III. CONCLUSION

This paper focuses on collaborative filtering and K-mean for book recommendation system, which on the mechanism of mining and classifying works easy as a simple learning algorithm.

We proposed an algorithm which works as personalized and customized recommendation system, this approach increases accuracy and quality of clustering for recommendation.

The proposed work implies that to maintain quality and authenticity of any system, information and data source are required. Similarity for calculating distance between user and cluster center is adjusted and calculated. While calculating mean, the user who gave scores are only considered. K-mean clustering improves the accuracy of clustering algorithm and is suitable with collaborative filtering when compared. Rating scales for items are different in collaborative filtering for different user. Most of the people gives low scores many gives high. Similarity calculation for the factors are not considered but can be adjustable to overcome the defects. The approach of using it can provide with the solution balancing the average score and adjust the similarity using method of K-mean clustering. User is assigned with the most similar cluster of his search, while computation of similarity in user and cluster.

This system of recommendation is simple and convenient for user to use and reduce searching efforts, while using method by avoiding answers of complicated questions and other method depending on rating of data. The complete work signifies that the similar users are based on the relevant distance. The similarity is based on the maximum size of cluster and relevant value



#### IV. REFERENCES

- [1] T. Zuva, S.O. Ojo, S. M. Ngwira, K. Zuva, "A Survey of Recommender System Techniques, Challenges," International Journal of Emerging Technology and Advanced Engineering, vol. 2, no.11, pp. 382-386, November 2012.
- [2] A.S. Tewari, A. Kumar, and A.G. Barman, "Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining." Advance Computing Conference (IACC), 2014 IEEE International. IEEE, pp. 500 - 503, 2014.
- [3] A. Chadha, and S. Kumar, "An improved K-Means clustering algorithm: A step forward for removal of dependency on K." Optimization, Reliability, and Information Technology (ICROIT), 2014 International Conference on. IEEE, 2014: 136-140.
- [4] C. Kaleli, "An entropy-based neighbor selection approach for collaborative filtering." Knowledge-Based Systems, vol. 56, pp. 273-280, 2014.
- [5] Jim Z. C. Laia, Tsung-Jen Huang, Yi-Ching Liaw, "A fast k-means clustering algorithm using cluster center displacement", Pattern Recognition 42 (2009) 2551-2556.
- [6] Al Mamunur Rashid, Gerge Karypis and John Riedl, "Learning Preferences of new users in Recommender System: An Information Approach.", SIGKDD Workshop on Web Mining and Web Usage Analysis (WEBKDD), 2008.
- [7] A. K. Jain and R. C. Dubes, Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [8] Kanungo, T.; Mount, D. M.; Netanyahu, N. S.; Piatko, C. D.; Silverman, R.; Wu, A. Y. (2002). "An efficient k-means clustering algorithm: Analysis and implementation" (PDF). IEEE Trans. Pattern Analysis and Machine Intelligence. 24 (7): 881–892. doi:10.1109/TPAMI.2002.1017616. Retrieved 2009-04-24.
- [9] Y. Cai, H.F. Leung, Q. Li, H. Min, J. Tang and J. Li, "Typicality based collaborative filtering recommendation," IEEE transactions knowledge and data engineering, vol. 26, no. 3, pp. 776-779, March 2014.
- [10] Shun-Hong Sie and Jian-Hua Yeh, "Library Book Recommendations Based on Latent Topic Aggregation", International Publishing Switzerland, pp. 411–416, 2014.
- [11] Pranav Bhure, Navinkumar Adhe, "Book Recommendation System Using Opinion Mining Technique", International Journal of Research in Engineering and Technology(IJRET), eISSN: 2319-1163, ISSN: 2321- 7308, Volume: 04 ,Issue: 1 ,pp333, Jan-2015

#### CITE AN ARTICLE

**Rani, R., & Sahu, R. (2017). BOOK RECOMMENDATION USING K-MEAN CLUSTERING AND COLLABORATIVE FILTERING. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, 6(11), 145-150.**