A Recommender for Retail Business using Consumer Segmentation with Clustering Techniques

Dr. Sabah Mohammed [1], Sarthak Rawat[2]

[1] *PEng (Full Professor), PhD, ISP, IEEE Senior Member, Department of Computer,*
Lakehead University, mohammed@lakeheadu.ca

[2] *Student, Msc Computer science, Department of computer science,*
Lakehead University, srawat@lakeheadu.ca

I. Abstract

Retail marketers are constantly looking for ways to improve the effectiveness of their campaigns. One way to do this is to target customers with offers that are more likely to attract them back to the store and to spend more time and money on their next visit. Demographic market segmentation is an approach to segmenting markets. A company divides the larger market into groups based on several defined criteria. Age, gender, marital status, occupation, education and income are among the commonly considered demographics segmentation criteria.

We propose a study wherein we would try to help online retail business by better understand its customers and therefore conducting customer-centric marketing more effectively. Based on Recency, Frequency, and Monetary model, customers of the business have been segmented into various meaningful groups using the different clustering algorithm. Furthermore, using this knowledge we would build a better recommender system personalized to different segment of customers.

Keywords: Customer segmentation, Consumer Segmentation, Pattern Clustering, Clustering algorithms, Recommender Systems, RFM-model, K-means/Fuzzy K-means, collaborative filtering.

II. Introduction

For internet-based business, the importance of appropriate recommendations is growing fast and people are increasing expecting suitable recommendations from those businesses to identify products and services. That is why many companies and websites have initiated the implementation of recommendation systems in recent years to identify customer interests. The recommender system is designed to assist users to identify necessary items. To get a better recommendation, an insight of likes and dislike of customers can help a lot. To gain such insights Customer segmentation is an excellent way. The process of grouping customers of individuals who share common characteristics is called Customer Segmentation [1]. This enables Companies to create targeted marketing for a specific group of customers which increases the chances of the person buying a product. These product recommendations are personalized and specific to every customer. It allows a firm to create and use specific communication channels to communicate with different segments to attract them. Customer segmentation personalizes the messages of individuals to better communicate with the intended groups. The most common attributes used in customer segmentation are location, age, sex, income, lifestyle and previous purchase behaviour. Here, segmentation is done using behavioural data since it is commonly available and continuously evolving with time and purchase history. RFM (Recency, Frequency, and Monetary) analysis is a renowned technique used for evaluating the customers based on their buying behaviour. A scoring method is developed to evaluate scores of Recency, Frequency, and Monetary. Finally, the scores of all three variables are consolidated as RFM score. In this context, it has been observed that the scores of three factors Recency, Frequency and Monetary directly proportional to customer's lifetime and retention. Once the values of recency, frequency and monetary are calculated, the K-Means algorithm is applied to the variables to clusters of the customer base. The behaviour of each cluster is analysed to find the group of customers who give more profits to the company. Similarly, clustering is performed using two other algorithms namely, Fuzzy C – Means clustering

III. RELATED WORK ON CONSUMER SEGMENTATION

March Liao [2] identified that both customer segmentation and buyer targeting are necessary to improve the marketing performances. These two tasks are integrated into a step-by-step approach, but the problem faced is unified optimization. To solve the problem, the author proposed the K-Classifiers Segmentation algorithm. This approach focuses on distributing more resources to those customers who give more returns to the company. A sizable number of authors had written about different methods for segmenting the customers.

Cho [3] proposed a recommendation system using weighted frequent pattern mining. Where Customer profiling is performed to find the potential customers using the RFM model. Using the RFM model with varied weights for each transaction provided a more accurate recommendation to the customer which in turn increases the profit of the firm.

Sheshasaayee [4] designed a new integrated approach by segmentation with the RFM and LTV (Lifetime Value) methods. They used a two-phase approach with the first phase being the statistical phase and the second phase is to perform clustering. They aim to perform K-means clustering after the two-phase model and then use a neural network to enhance their segmentation.

Christy[5] proposed a novel idea for choosing the initial centroids in K- Means is proposed. The results obtained from the methodologies are compared with one another by their iterations, cluster compactness and execution time. This clustering algorithm is then used to find segments of customer more efficiently and accurately.

F.Darvishi-mirshekarlou [6] gave the review about the Recommender systems using collaborative filtering. It is the most popular and successful method that recommends the item to the target user. These users have the same preferences and are interested in it in the past. Scalability is the major challenge of collaborative filtering. With regard to increasing customers and products gradually, the time consumed for finding nearest neighbour of target user or item increases, and consequently more response time is required.

Zan Wang [7] proposed a hybrid model-based movie recommendation system which utilizes the improved K-means clustering coupled with genetic algorithms (GA). Thus, user space becomes much denser and reliable.

The research discusses about customer segmentation as a case study and implements different ways to segment customers based on RFM models. Whereas research on recommender system, mostly focused on improving recommender systems and not on incorporating customer centric knowledge. We would try and build our model on top of a case study of customer segmentation, by using the information gained by consumer segmentation to incorporate in our recommender system to personalize recommendation.

IV. DESIGN ARCHITECTURE

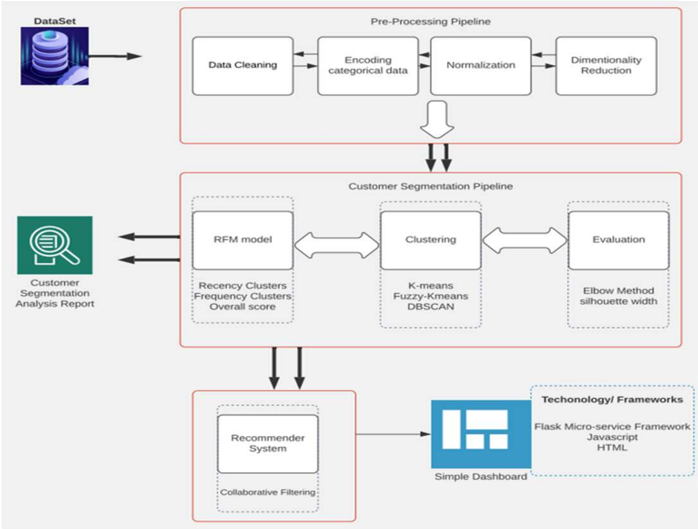

Fig 1. Design of Application

To recommend a product to a particular customer segment we divided our project into three different pipelines:- Pre-processing pipeline, Segmentation pipeline, Recommender/Dashboard. Every pipeline is dependent on the previous pipeline to get data for processing. Fig 1 shows a brief architecture of our project.

The pre-processing pipeline is responsible for data cleaning, data encoding, normalization, and dimensionality reduction. Data cleaning includes removing duplicates, irrelevant observations and errors, unnecessary columns, handling inconsistent data, missing value imputation. Data encoding and normalization would encode categorical variables and normalize the data so that we don't have different magnitude of columns. This pipeline also would be responsible for dimensionality reduction so that we are able to remove unwanted feature's and visualize our data at the same time.

Segmentation pipeline utilizes data given by the pre-processing pipeline and then performs RFM model segmentation using different clustering algorithms using the SciKit Learn framework.

Recommender system pipeline would receive data from segmentation pipeline and would provide recommendation to customers on a web dashboard. Fig 1.1 shows a detailed sequence diagram of our project

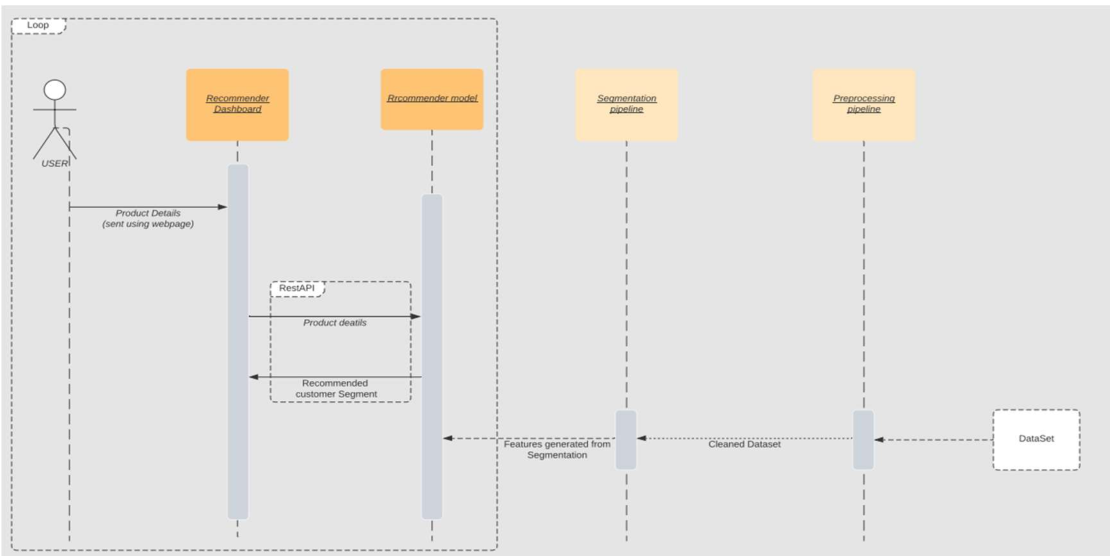

Fig 1.1 Sequence Diagram

A. Recommender Architecture

The recommender architecture is a hybrid collaborative filtering architecture which takes into input the cluster information along with CustomerId in the user input section. Whereas the product input takes information about product. The information is then converted using embedding layer and then concatenates using concatenate API of Kera's the concatenated layer is then fed into two Dense layers to get the predicted rating of a product. Fig 2 shows us the neural network architecture of recommender system.

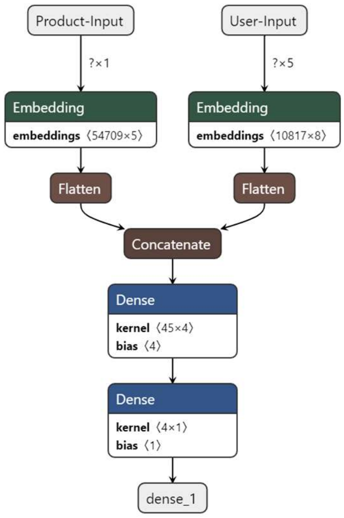

Fig 2. Architecture of Recommender

V. Clustering Algorithms

The transactional dataset of the customers of a company is used to perform the segmentation process. In this research, three different algorithms have been used to cluster the customers based on RFM analysis. The data is initially pre-processed to remove outliers and to filter meaningful instances. The pre-processed information is then fed into the RFM model to calculate the recency, frequency, and monetary values. The three attributes are then passed to three clustering algorithms namely K-Means and Fuzzy C-Means clustering algorithm. These algorithms cluster the customers into segments. The workability of the clustering algorithms is then analysed regarding the number of iterations, cluster compactness and the time taken for execution.

A. K-means Clustering

K-Means [8] is a standard algorithm which takes the parameters and the number of clusters as inputs and partitions the data into the defined number of clusters such that the intra-cluster similarity is high. K-Means is an iterative approach which computes the value of centroids before each iteration. The data points are moved among different clusters depending on the centroids calculated at each iteration. The process is repeated until the sum cannot be decreased any more. K- Means algorithm is shown in Algorithm 1.

Algorithm 1 k-means algorithm

1: Specify the number k of clusters to assign.
2: Randomly initialize k centroids.
3: **repeat**
4:     **expectation:** Assign each point to its closest centroid.
5:     **maximization:** Compute the new centroid (mean) of each cluster.
6: **until** The centroid positions do not change.

B. Fuzzy C-means

Fuzzy C-Means is a clustering approach [9] which permits a specific data to be present in more than one cluster. It does not decide the membership history of a data point to a given cluster. Instead, the likelihood that a specific data point will belong to that cluster is calculated. The advantage that Fuzzy C-Means has over K-Means is that the result obtained for the large and similar dataset is better than K-means algorithm because in K-means a data point must entirely be present in only one cluster. In this study, a customer may belong to more than one cluster which increases the chance of retaining the customers by treating them with different offers for each segment. The time complexity of Fuzzy C-Means is O (n+k+d2+i), where d is the number of iterations.

Algorithm 2 Fuzzy C means Algorithm

Specify iteration threshold and random init fuzzy partition matrix U^s
Specify ϵ
while $U^{(s+1)} - U^{(s)} < \epsilon$ **do**
    Calculate cluster center matrix
    Compute Euclidean distance
    Update fuzzy partition matrix $U^{(s+1)}$
send cluster results

VI. PROPOSED METHODOLOGY

In this section the proposed model to evaluate the accuracy of the predictions is described. The main purpose of this model is to performing customer segmentation using k-means algorithms and then use this information to give better recommendations. To evaluate our models, we use two approaches first using Shillouet width and Calinski Harabasz Score for Cluster evaluation, Second Using RMSE to evaluate our recommender algorithm.

Brazil Dataset Description

The data set contains actual sales data for a chain of Brazilian stores. The names of products, customers, and employees were changed to preserve their identity. This dataset contains 13 fields some of the important fields are: Order Number, ProductName, ClientID, Sale Date Time, Product Cost, Item Quantity, Total item value. The data includes 26,951,165 sales records for 10817 customers for about 51,671 different products.

Data Collection and Clean-up

Out of the 26,951,165 fields many fields do not have customer_Id value which is imperative for our analysis. So we decided to drop the values which do not have these values. Consequently, the number of transactions is reduced to 1,628,925 which is more than enough to perform our analysis. These transactions are in fact related to the products and customers during the under-study period.

The RFM Variables Extraction

Bose and Chen[10] believe that data on customers behaviour include customers transaction records, feedbacks from customers which can be easily indicated usi

ng the RFM models. To create a RFM model we must take more of a statistical approach. For each customer we generate three features (Recency, Frequency, Monetary) based on their transactions. These features have a range of between 1 to 2000 which was a arbitrary value chosen to help us visualize better clusters. Recency is generated by calculating the time elapsed from last purchase. For example, a customer who purchased product recently will have higher recency than a customer who made his last purchase few months ago. Similarly, Frequency is determined the number of times a person purchases a product from the retail website. And monetary aligns with his total spending on the website to date. The datapoints of each customer is shown in Fig 3.

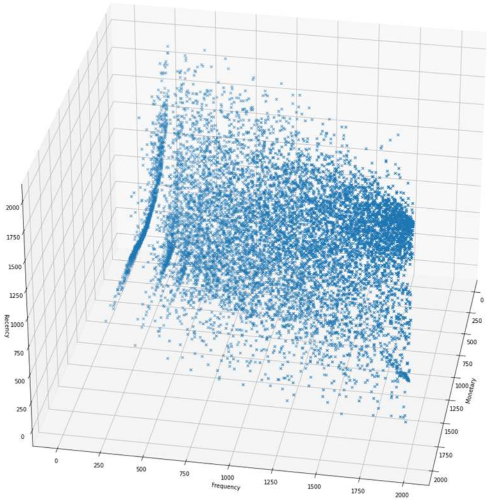

Fig 3. Plotting 3-D graph of the Data Points

Recommender System

One of the methods widely prevalent in the recommender systems is collaborative filtering [11] (CF) method which is the best method for this research considering existing data. The proposed way creates a collaborative filtering neural network as baseline model. Then we compare this baseline model with the two proposed models with additional RFM information. The recommender architecture is a hybrid collaborative filtering architecture which takes into input the cluster information along with CustomerId in the user input section. Whereas the product input takes information about product. The information is then converted using embedding layer and then concatenated and fed into two Dense layers to get the predicted rating of a product.

VII. Clustering and Recommender Evaluation/Result

The results were evaluated in two phases. First evaluation was done for Clusters and different clustering algorithm during the customer segmentation phase. And Second evaluation was done for our Collaborative filtering recommender system.

Clustering Evaluation

After generating RFM features for every customer different clustering algorithm were performed on our dataset in order to extract different segments of customers. Algorithm's applier were K-means, Fuzzy C-means and Agglomerative Clustering [12]. To select the optimal algorithm two different Evaluation methods were used namely Silhouette Score [13] and Calinski Harabasz Score [14]. For finding the ideal number of segments Elbow method[15] was applied. The elbow method is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters and picking the elbow of the curve as the number of clusters to use. The graph of inertia vs number of clusters is shown in Fig 4.

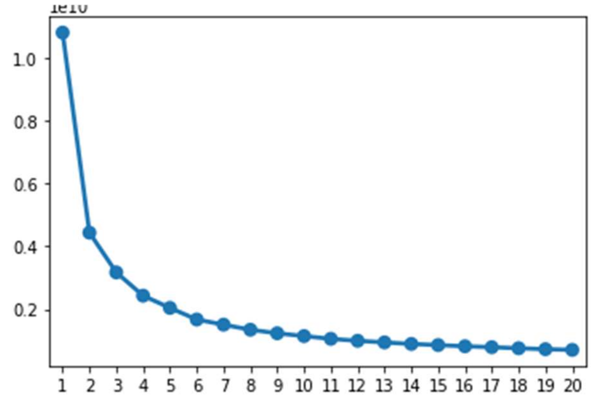

Fig 4. Elbow point evaluation for clustering.

The best number of clusters from elbow method was determined to be 5. To verify our results, Graphs of Silhouette Score and Calinski Harabasz Score Vs number of clusters were plotted. Tt verified that indeed 5 was the best number of clusters in the vicinity. Fig 5 & 6 show the evaluation result of Silhouette width and Calinski Harabasz Score respectively.

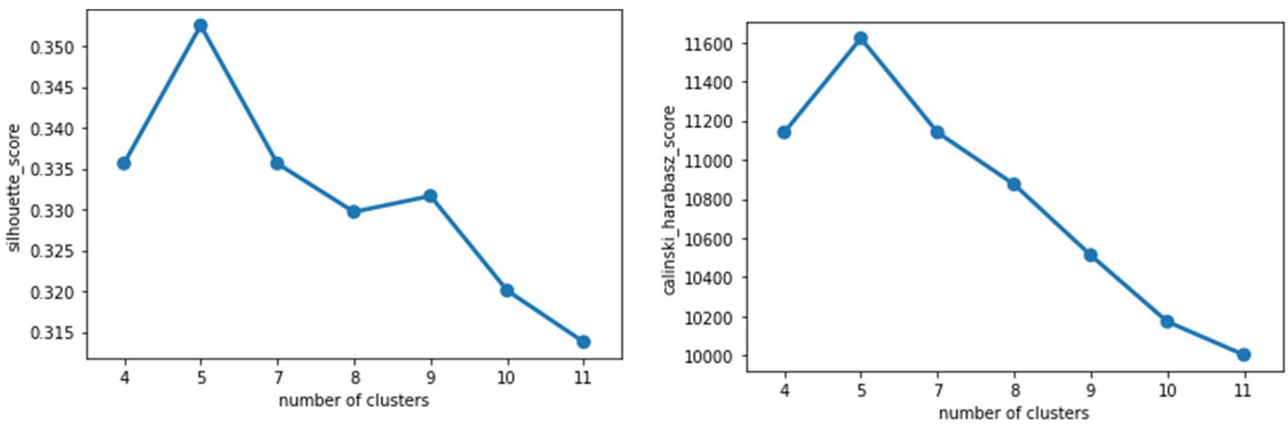

Fig 5. Silhouette width and Calinski Harabasz Score Vs Number of Clusters (K-means)

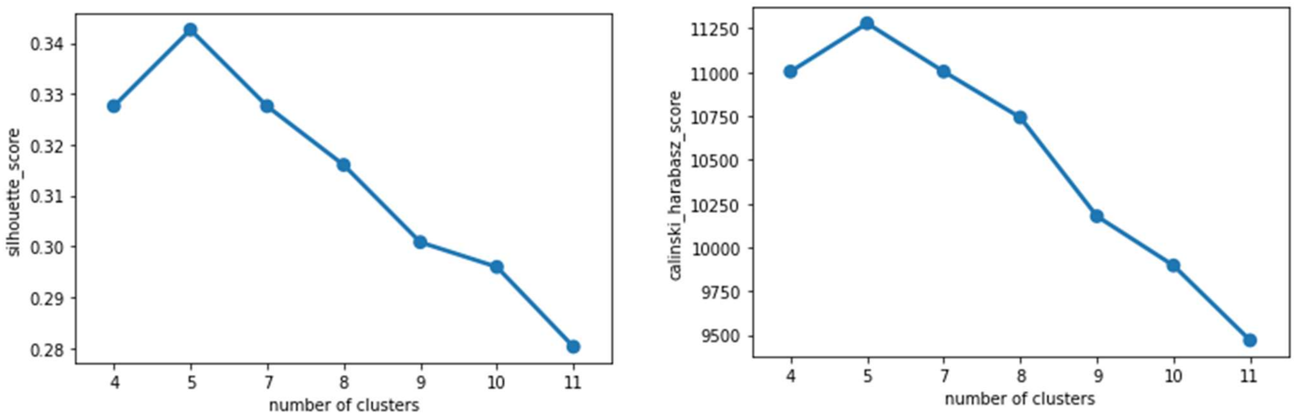

Fig 6. Silhouette width and Calinski Harabasz Score Vs Number of Clusters (Fuzzy C-means)

Evaluation for different algorithms were performed and the best performing algorithm: K-means was selected for our final segmentation model. Table 1. shows comparison of different algorithms along with evaluation scores.

Table 1. Evaluation of Different Algorithms

ALGORITHM	Silhouette Score	Calinski Harabasz Score
K-MEANS	0.3525	11620.96
FUZZY C-MEANS	0.3426	11279.83
AGGLOMERATIVE	0.3381	10432.13

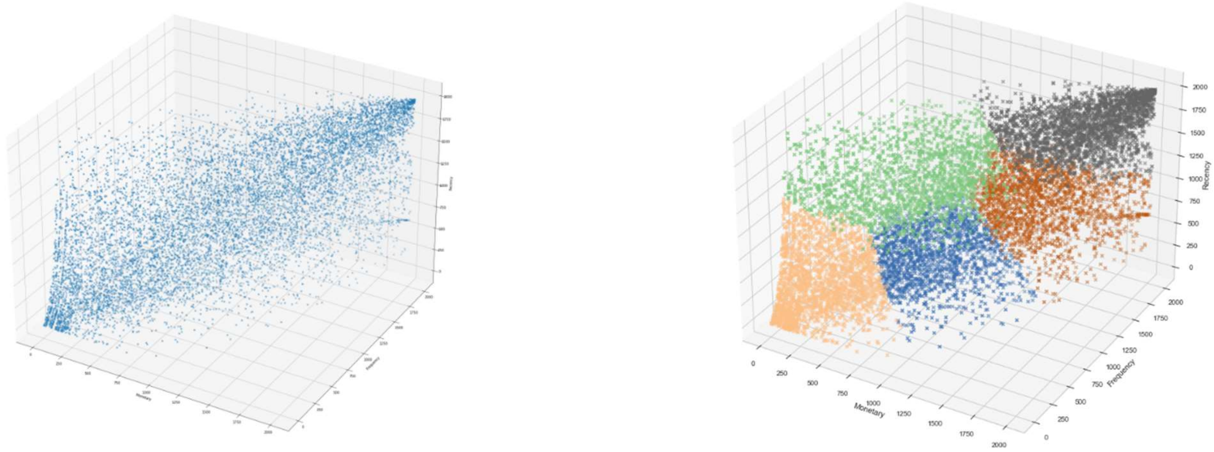

Fig 7. Initial and final Clustering results

The final results of clustering is show in Fig 7. After creating clusters with the help of k-means algorithm these segments were further analysed to decide the best and worst segments. The segments were divided into Champions, Loyal Customers, Potential Loyalist, Needs Attention, At Risk/Lost. Fig 8 shows an average graph of Recency, Frequency and Monetary for all the segments. Using this information, we can create different strategies/actions to appeal each segment. Some of these actions are shown in Table 2 below.

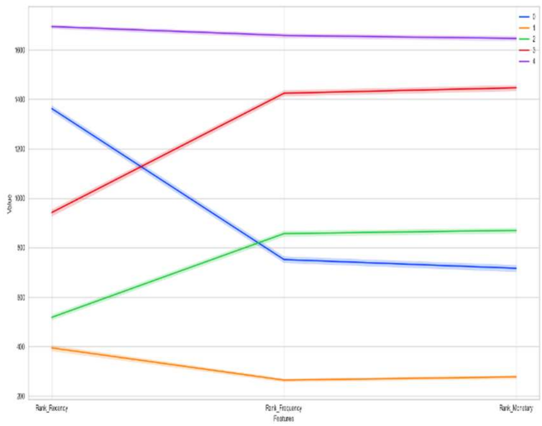

Fig 8. Average RFM for every cluster

Table 2. Identification of each segment

Cluster Identified	Cluster label	Description	Activity	Actions to be done
Champions	4 (Purple)	High RFM	Bought recently, buys often and spends the most.	Reward them. Can be early adopters of new product
Loyal Customers	3 (Red)	Medium R, High F& M	Spends good money often. Responsive to promotions.	Ask for reviews. Engage them.
Potential Loyalist	0 (Blue)	Good recency but medium freq and monetary	Recent customers, buys and spends often.	Offer membership/loyalty program.
Needs Attention	2 (Green)	Low recency, high freq and monetary	Used to spend decent amount. But have stopped recently.	Make limited time offers. Recommend products. Reactivate them
At Risk/Lost	1 (yellow)	Low RFM	Bought only once and lost interest.	Revive interest with reach out campaign.

Recommender Evaluation

This useful information regarding each segment is then used in our recommender. The cluster label along with RFM scores are fed into the recommender neural network alongside with product and customer IDs. Three models are evaluated alongside each other. Where Baseline model is simple collaborative filtering model using neural network and other two are collaborative filtering models with segment information. These models were run for 50 epochs each with 20-80% test train split. Each model had exponential decaying learning rate with initial rate of 0.01, every model was trained using Adam optimizer[16] and a dropout of 0.25, The results for all the models are shown in the Fig 9 below.

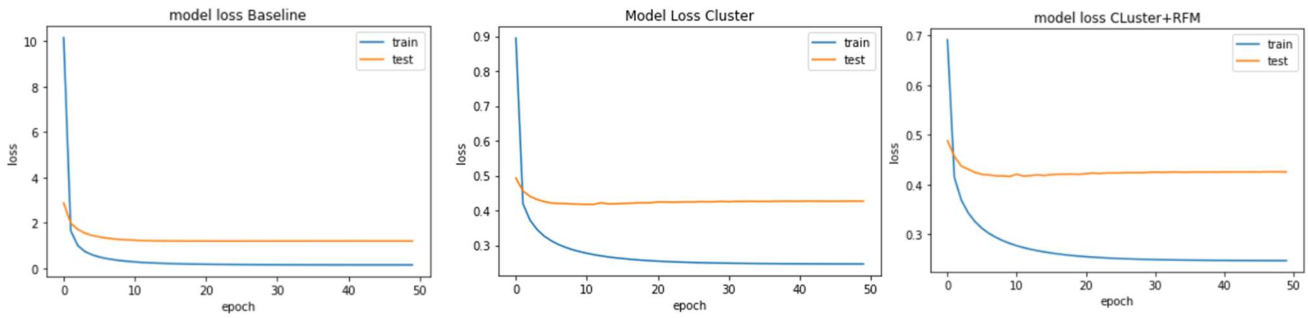

Fig 9. Test and Train loss for three recommender models

Baseline model when compare with other models with clustering information and RFM scores, showed that the other models were better. The baseline model had a MSE of 1.182 whereas the other two models had MSE of about 0.426 indicating that these models performed much better.

Table 3. Loss comparison of each model

Model	Training Loss (MSE)	Test Loss (MSE)
Baseline (Collaborative filtering)	0.139	1.182
Baseline + clustering info	0.245	0.426
Baseline + clustering and RFM info	0.246	0.425

VIII. Conclusion

Customer segmentation improves a company's relationships with its customers. Finding new customers for the enterprise is vital, meanwhile retaining the existing clients [17] is even more important. In this project, segmentation is done using RFM analysis and then is extended to other algorithms like K –Means clustering, Fuzzy C – Means. These clustering algorithms are then evaluated using Silhouette width and Calinski Harabasz, and it is observed that K-means performed better. Since segmentation is done based on the values of recency, frequency, and monetary values, the company can customize their marketing strategies to the customers based on their buying behaviour. The information gained by segmentation is then further used to develop an improved recommender system to predict ratings of customers for each product. The model is then evaluated against a baseline collaborative filtering model by plotting test and train loss. The proposed model is more effective because it considers customer similarity as a feature from beginning rather than trying to find similarity (Like in baseline model). Future work includes studying the performance of the customers in each segment and improving the recommender to tackle cold start problem. This would help better in providing better promotional offers to specific products.

GITHUB link: **https://github.com/sarthak7295/recommender_segmentation.git**

IX. References

[1] Kim, Su-Yeon, Tae-Soo Jung, Eui-Ho Suh, and Hyun-Seok Hwang. "Customer segmentation and strategy development based on customer lifetime value: A case study." Expert systems with applications 31, no. 1 (2006):101-107.

[2] Jiang, Tianyi, and Alexander Tuzhilin. "Improving personalization solutions through optimal segmentation of customer bases." *IEEE transactions on knowledge and data engineering* 21, no. 3 (2008): 305-320.

[3] Cho, Young Sung, and Song Chul Moon. "Weighted mining frequent pattern based customer's RFM score for personalized u-commerce recommendation system." *JoC* 4, no. 4 (2013): 36-40.

[4] Sheshasaayee, Ananthi, and L. Logeshwari. "An efficiency analysis on the TPA clustering methods for intelligent customer segmentation." In *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pp. 784-788. IEEE, 2017..

[5] Christy, A. Joy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa. "RFM ranking–An effective approach to customer segmentation." *Journal of King Saud University-Computer and Information Sciences* (2018).

[6] Darvishi-Mirshekarlou, F., S. H. Akbarpour, and M. Feizi-Derakhshi. "Reviewing cluster based collaborative filtering approaches." *International Journal of Computer Applications Technology and Research* 2, no. 6 (2013): 650-659.

[7] Wang, Zan, Xue Yu, Nan Feng, and Zhenhua Wang. "An improved collaborative movie recommendation system using computational intelligence." *Journal of Visual Languages & Computing* 25, no. 6 (2014): 667-675.

[8] Krishna, K., and M. Narasimha Murty. "Genetic K-means algorithm." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 29, no. 3 (1999): 433-439.

[9] Bezdek, James C., Robert Ehrlich, and William Full. "FCM: The fuzzy c-means clustering algorithm." *Computers & geosciences* 10, no. 2-3 (1984): 191-203.

[10] Bose, Indranil, and Xi Chen. "Quantitative models for direct marketing: A review from systems perspective." *European Journal of Operational Research* 195, no. 1 (2009): 1-16.

[11] He, Xiangnan, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. "Neural collaborative filtering." In *Proceedings of the 26th international conference on world wide web*, pp. 173-182. 2017.

[12] Murtagh, Fionn, and Pierre Legendre. "Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?." *Journal of classification* 31, no. 3 (2014): 274-295.

[13] Shahapure, Ketan Rajshekhar, and Charles Nicholas. "Cluster Quality Analysis Using Silhouette Score." In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 747-748. IEEE, 2020.

[14] Baarsch, Jonathan, and M. Emre Celebi. "Investigation of internal validity measures for K-means clustering." In *Proceedings of the international multiconference of engineers and computer scientists*, vol. 1, pp. 14-16. sn, 2012.

[15] Liu, Fan, and Yong Deng. "Determine the number of unknown targets in Open World based on Elbow method." *IEEE Transactions on Fuzzy Systems* (2020).

[16] Bock, Sebastian, and Martin Weiß. "A proof of local convergence for the Adam optimizer." In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8. IEEE, 2019.

[17] Tong, Lili, Yiting Wang, Fan Wen, and Xiaowen Li. "The research of customer loyalty improvement in telecom industry based on NPS data mining." *China Communications* 14, no. 11 (2017): 260-268.

