

Customer Segmentation and Buyer Targeting Approach

Yogesh Jadhav, Deepa Parasar



Abstract: Nowadays, maintaining customer loyalty and attention span of the customers are major challenges faced by the retail industry. The availability of varied options in the market for similar purposes increases the competition between organizations, to market their product, tremendously. This leads to the need for reinforcement of marketing strategies from time to time. With the advancement of technology, this can be made possible. This paper proposes a systematic approach for targeting customers and providing maximum profit to the organizations. An important initial step is to analyze the data of sales acquired from the purchase history and determine the parameters that have the maximum correlation. We focus on the parameters recency and frequency of the purchases made by the customers to perform clustering. Based on respective clusters, proper resources can be channeled towards profitable customers using machine learning algorithms. This paper also deals with the draw-backs of the recommender system like cold start problem, sparsity, etc and how they can be overcome. K-Means clustering is used for customer segmentation and Singular Value Decomposition is used for providing appropriate recommendations to the customers.

Keywords: K-Means, customer segmentation, Singular Value Decomposition, cold start problem.

I. INTRODUCTION

Customer relationship management (CRM) is a market strategy that enables businesses to learn about customers' behaviors and needs that would help in developing stronger relationships with them. Advancements in technology have facilitated successful customer engagement in recent years as they help to solve business questions that in the past were too time-consuming to pursue because of manual computation. Particularly through data mining and the extraction of hidden patterns of customer purchases from large databases, organizations can identify valuable customers, predict their future behaviors and this enables firms to make proactive, knowledge-driven decisions. Customer clustering and buyer targeting are the two intelligent components of Customer Relationship Management [1]. In segmentation, deciding upon the optimum number of clusters and the variables used for clustering is an important step. As there could be a large number of variables that can be used to differentiate customers, optimum number of variables have to be

determined that can form the most distinct clusters [2]. The clustering parameters can broadly be classified as geographic, demographic, psychographic and behavioral. Geographic clustering will group customers belonging to the similar area together with an assumption that the needs of the same area people will be similar. However, it is majorly seen that rural and urban needs can be distinguished; but otherwise, the area does not provide much information for targeting customer. Psychographic clustering includes grouping on the basis of personality, lifestyle or the social class. However, this kind of personalized recommendations might affect the privacy of the customer. Demographic segmentation groups the market, based on gender, age, education, income, occupation, religion, and nationality. This may result in ignoring the fact that customers may not act on the basis of these parameters. For example, a mother may buy products for an age group of 1-5 years as well for an age group of 50-70 years. Hence, the broader parameter should be taken into consideration like the behavioral parameters. Behavioral parameters include clustering on the basis of regency and frequency of purchases. Regency is how recent was the last purchase of customer and frequency is how often the purchase happens. Despite the simplicity of these parameters, the clustering on this basis gives classes of customers which can be then handled differently leading to boost sales. According to the Pareto's rule, only 20% of the customers contribute to 80% of the sales of the organization. So as per [3], the best as well as the weakest cluster obtained from behavioral clustering can be targeted differently to gain maximum profit to the businesses. One of the most successful ways to target customers with marketing campaigns is through automated merchandising. This concept involves providing the customers with relevant and customer specific recommendations and this can be achieved through the use of recommender systems. Content based, Collaborative and Hybrid are three major types of recommender systems. In content-based recommender systems the products recommended to a user are similar products that the user has purchased in the past or is examining currently. In collaborative recommender systems, similarities between the users are recognized on the basis of their purchases and new recommendations are generated on the basis of inter-user comparisons. Hybrid recommender systems are but not necessarily a combination of collaborative and content based recommender systems. According to [4] they are highly preferred by organizations as they combine the strengths of two recommender systems whilst eliminating any weakness (es) that exist when only one recommender system is used. Rectification is not possible.

Revised Manuscript Received on November 20, 2020.

* Correspondence Author

Prof. Yogesh Jadhav,* Assistant Professor at Amity School of Engineering and Technology, Amity University Mumbai.

Dr. Deepa Parasar Associate Professor at Amity School of Engineering and Technology, Amity University Mumbai.

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

II. LITERATURE SURVEY

Large amount of data is collected in the retail industry on sales and customer shopping history [5]. Due to increasing ease, popularity and availability of the business conducted on web, the quantity of collected data continues to expand at a rapid pace.

Retail industry is considered a very rich source for the field of data mining. Data mining in retail helps to discover shopping patterns and trends of customers identify customer behavior, improve customer service quality, achieve better customer satisfaction and retention, enhance product consumption ratios, distribution policies and reduce the cost of business. Some of the applications of data mining in retail are:

A. Customer Relationship Management

Customer segmentation: Customer segmentation is one of the crucial factors in an organization's marketing strategies. It offers insights into how different segments of customers respond to changes in fashions, demographics and trends. For example, it can help classify customers into the following segments:

1. Customers responding to new campaigns
2. Customers responding to new product launched
3. Customers responding to discounts
4. Customers showing propensity in purchasing specific products.

Customer Lifetime Value (CLV):All customers are not equally profitable. This factor attempts to compute some relative measure of value by calculating the Risk Adjusted Revenue (which is the probability of a customer owning categories or products in his/her portfolio that he/she currently does not have), as well as the Risk Adjusted Loss (which is the probability of the customer dropping categories/products in his/her portfolio that he/she currently owns), adding to any Net Present Value, and deducting value of providing service to the customer.

Customer Potential: There are some customers who are not quite profitable today but have the potential of becoming profitable in the future. Hence it is very much essential to identify such customers with a high potential before deciding market strategies. One of the best ways to realize the potential of customers is by introducing the right marketing strategies.

Customer Loyalty Analysis: Retaining an existing customer is more economical than to acquiring a new one. For developing effective customer retention strategies, it is crucial to analyze the reasons for attrition of customers. Business Intelligence provides help in understanding the customer attrition in accordance with various factors influencing customers and sometimes drilling down to individual transactions can be done, which can tell us what might have lead to the change in loyalty.

Cross Selling: Retailer organizations use large amount of information of customers available to them for cross selling other products at purchase time. This is done using product portfolio analysis after which the products missing from their portfolios are sold. Market basket analysis is another method for performing effective cross selling. Look-a-like modeling is another strategy where a model is produced

where some quantitative measure of the affinity of customers to specific products is analyzed.

Product Pricing: One of the most vital marketing decisions taken by the retail organizations is pricing. Often, increase in the price of products may result in lowered sales and adoption of replacement products by customers. Using data mining, retailers may develop some sophisticated pricing models for different products that can establish price - sales relationships to check how change in prices of products affect the product sales.

Target Marketing: Retail industries can optimize overall promotion and marketing efforts by targeting campaigns to appropriate customers. Target marketing can be according to a very simple customer buying habit analysis; but data mining tools are being increasingly used to determine specific customer segments that are most likely to respond to particular kinds of promotions.

Store Segmentation: In this analysis the data that is common for a variety of different stores is taken, and which stores are similar in terms of customer or product dimensions is found out. In other words, which stores are similar based on the products that are sold very quickly or more slowly in comparison to the other stores. The next step is building profiles of customers who buy from a specific store.

Market Basket Analysis: It is used for studying naturally existing affinities between products. Beer-diaper affinity, a common example of market basket analysis, in which it can be found that men who buy diapers are very likely to buy beer. It is an example of two-product affinity. But market basket analysis can get extremely complex in real life resulting in unknown affinities between various products. This analysis has numerous uses in retail industry. One very classic use is for product placement in-store. One more popular use is of product bundling i.e. grouping products to sell in combo deals.

Out-Of-Stock Analysis: It probes into the reasons that lead to an 'out of stock' scenario. Basically, various variables are involved and the process can get complicated. A crucial part of the analysis is to calculate the lost revenue because of product running out of stock.

B. Data Mining Techniques for CRM

The retail businesses need to continuously reinforce their marketing strategies to maintain the attention span and loyalty of the customers [6]. Data mining approach provides a guide to solve this issue. Basic steps considered for customer relationship management using data mining and machine learning are as follows:

1. Defining the problem in business
2. Building the marketing database
3. Exploring the data
4. Preparing data for modeling purpose
5. Building model

Each business has its own objective. The model needed to be built for each goal depends on the requirement of the business. It also depends on factors like whether demographic conditions, psychographic conditions, behavioral conditions should be taken into consideration.



Increase in the revenue of the business by increasing the customer engagement is the underlying aim. Data Preparation is the most crucial step in any data mining application. The data may consist of various parameters and may be having many tables. Performing exploratory data analysis to find out prominent parameters that need to be considered and cleaning data of the data need to be done in this step. This will result in integration and consolidation of the data into single dataset.

For performing data analysis, various statistical operations like standard deviation, average, mean, variance can be calculated. Whereas various visualization tools can also be used to plot pie chart, violin graphs, bar graphs, etc and gain insight into correlation in various parameters as a result. Broadly four steps of data preparation can be considered as shown below:

1. First step is selection of the variable for the purpose of building a model. Ideally, all the variables must be fed to data mining tool and best predictors should be found out.
2. The proceeding step is to derive variables that might prove useful for predictive operations ahead.
3. Next, a subset of the dataset should be decided as sample for training purposes. This will reduce loss of information as well.
4. Finally, the process of building model is iterative, hence should be carried out in that way.

Alternative models also should be tried, and the model that proves most useful in solving the problem should be taken into consideration. The overall goal of customer satisfaction should be solved by offering profitable services to customers.

C. A Comparative Analysis of Clustering Algorithms

In the field of data mining, there are two approaches to solve a problem- supervised learning and unsupervised learning. Among that clustering is an unsupervised algorithm. Clustering does not have the group descriptions predefined as in segmentation. Clustering divides the object in dataset in such a way that object having similar properties are grouped together. Hence, clustering has wide applications in various fields like pattern recognition, machine learning, bio-informatics, artificial intelligence etc.

Various algorithms can be used for clustering [7]. Some of them are as follows:

1. K-Means Clustering

K-means is an unsupervised method used for clustering. It is largely used because of its efficiency and simplicity. If compared with hierarchical clustering, k-means clustering has a time complexity of $O(n)$ and hierarchical clustering has a time complexity of $O(n^2)$. Hence, k-means clustering works better for large dataset. Also, hierarchical clustering is used when there exists complex shapes of clusters and no predefined number of clusters. Moreover, as complexity of computation hierarchical clustering is high, that hinders its use for clustering in real time applications. Hence the use of K Means algorithm is made for carrying out the task of clustering. K means will work as follows:

- As is the prerequisite of k-means clustering, determine the number of clusters.
- Choose initial centroids randomly.
- Calculate the Euclidean distance which is the square of the distance of each data point and the centroid.

- Form clusters by assigning each data point to the cluster with least Euclidean distance.
- Calculate the new centroids from clusters obtained.
- Perform number of iterations to give the desired number of optimum clusters based on regency and frequency.

2. Hierarchical Clustering

A dendrogram is a tree like structure. The hierarchical clustering uses dendrogram for construction of hierarchy of clusters. Hierarchical clustering is of two types- Divisive Hierarchical Clustering and Agglomerative Hierarchical clustering. The later is a bottom up approach. Hence, in agglomerative clustering every single object starts as a cluster and then clusters combine on the basis of the distance in between them. Divisive hierarchical clustering is top down approach. Hence, in divisive hierarchical clustering, all the objects start as one cluster and the break down to finite number of clusters.

3. Expectation maximization (EM) algorithm

For the purpose of finding the maximum likelihood, this is an iterative method. It consists of two steps- the expectation step and the maximization step. The expectation maximization initially assumes the initial parameters. Then the probability distribution is computed on the basis of the parameters assumed. This is the expectation step. In the maximization step, new parameters are determined on the basis of the current completions of computing distribution. After several such repetition of computation of expectation and maximization, the algorithm converges.

4. Density based algorithm

Density based algorithm is a common algorithm. In density based algorithm from given set of points, points that are closely packed are grouped together. Hence, density based algorithm does not require the number of clusters to be predetermined. It is also robust to outliers and notion of noise. Also, density based algorithm identifies the cluster which may have arbitrary shapes.

D. Comparison between Clustering Algorithms

A comparative analysis of these four clustering algorithms namely K-means algorithm, Hierarchical algorithm, Expectation and maximization algorithm and Density based algorithm was done. The performance of these clustering algorithms was compared in terms of accuracy and efficiency. From Figure 1 it can be seen that K-Means clustering algorithm takes the lowest time i.e. 0.03 seconds and K-means algorithm is more accurate.

Algorithm	Number of Clusters	Cluster instance	Number of Iteration	Time	Accuracy
K-means	2	256(43%) 344(57%)	4	0.03s	56.66%
Hierarchical Algorithm	2	599(100%) 1(%)	4	1.51s	54.16%
EM Algorithm	2	311(52%) 289(48%)	4	0.44s	57.83%
Density based Algorithm	2	256(43%) 344(57%)	4	0.12s	56.66%

Figure 1: Result of clustering algorithms using Bank’s un-normalized data



Customer Segmentation and Buyer Targeting Approach

Algorithm	Number of Clusters	Cluster instance	Number of Iteration	Time	Accuracy
K-means	2	210(35%) 390(65%)	6	0.02s	55.20%
Hierarchical Algorithm	2	599(100%) 1(%)	4	0.59s	54.16%
EM Algorithm	2	365(61%) 235(39%)	4	0.14s	53.83%
Density based Algorithm	2	213(35%) 387(65%)	6	0.05s	50.83%

Figure 2: Result of clustering algorithms using Bank's normalized data

The accuracy of K-means algorithm is 56.66 percent when the input data is unnormalized. Distribution of cluster instance is better in Density based algorithm. However, Density based algorithm takes more time. It takes 0.12 seconds. From Figure 2 it can be said that when data is normalized, K-Means clustering algorithm takes the lowest time 0.02 seconds and more accuracy i.e. 55.20 percent as compared to all other algorithms. So, K-Means clustering algorithm produces better result as compared to the other algorithms with normalized and unnormalized data in terms of efficiency and accuracy.

From a marketing perspective, it makes sense to put in the effort to understand the parameters used for clustering to identify preferences of best customers for at least two reasons: Continuation to provide the clusters with what they are looking for and retain them as customers. Target marketing efforts toward prospects who happen to be best customers. Recency and Frequency Analysis uses sales data to segment a pool of customers based on their purchasing behavior. The resulting customer segments are neatly ordered from most valuable to least valuable [8]. This makes it straightforward to identify best customers. The idea behind Recency- Frequency Analysis is as follows: The customers, who have purchased recently, are more likely to buy again than customers who haven't been active for a while. The customers, who buy more often, are more likely to buy again than customers who buy infrequently.

E. Recommender Systems

Now days, the amount of information that we are retrieving and using has increased rapidly. Data mining is the process of mining relevant data from the large amount of data. It is the procedure of discovering and finding the appropriate pattern from the huge amount of data sets. The main aim of data mining process is to bring out appropriate and related information from the huge amount of data sets and convert it into comprehensible structure. One of the sub parts of data mining is recommender system. On the internet, there are many options for items or anything, so there is a need to filter and efficiently gives the information. Recommender systems solve these by analyze through the huge amount of data to discover users with personalized content and services [9]. Recommender Systems are primarily categorized on the basis of personalized and non-personalized recommendations. Personalized recommendations are those that are offered as a ranked list of products. Personalized recommender systems are used by e-commerce web- sites to recommend products to their customers based on past purchase behavior. Non-personalized recommendations are those which are simple to generate and normally feature in magazines or newspapers. These systems recommend products to customers depending on average ratings by other customers. Here, the recommendations do not depend on the customers, so all customers get same recommendations [10].

Recommender systems can be broadly classified into the following categories:

- Content-based Filtering Systems: Information of active customers and data about the products is used.
- Collaborative Filtering Systems: Information about set of customers and their correlation with the products is used to provide recommendations.
- Demographic Filtering Systems: Demographic information such as age, gender, education, etc. of people is used to identify types of customers.
- Hybrid recommender Systems: It uses a combination of Content based and Collaborative filtering techniques.

F. Challenges faced by Recommender Systems

The recommender systems suffers from various issues[11]. They are mainly categorized as given below.

Cold start problem is classified into two types:

- Cold start for new products
- Cold start for new customers

Cold start problem for a product occurs when there aren't enough previous ratings related to that product, or it has been introduced for the first time. In cold start problem for new customers, it becomes difficult to recommend products to new customers as there is no information related to the past purchases by the customer.

1. Scalability

As the numbers of customers and products grows, more resources are needed by the system to provide accurate recommendations to the customers. Majority of the resources are used for determining customers having similar interests, and products having similar attributes. It is a common problem in collaborative filtering method.

2. Sparsity

Suppose, there is an online shop having a huge amount of customers and products. If a customer purchased few products from the shop and has not rated a few of them, then, it will lead to a problem of sparsity. Also, it can be said that sparsity is a problem of lack of information.

3. Privacy

Privacy is also a major issue, especially in case of demographic recommender systems. For providing the most appropriate recommendation to the customers, the system must obtain appropriate information of the customers, including demographic information (age, gender, occupation, hobbies etc.) and data about the geographic location of customers which may exploit the customers' privacy.

4. Over-Specialization

It is a common problem faced by the content-based technique of recommendation systems. An ideal recommender system should suggest diverse products which content-based system lacks. It hinders the customers from getting exposed to something other than the usual products often purchased by the customer. Customers are recommended products that they are familiar to.

5. Freshness (Predictability)

In the problem of predictability, even if the products recommended to the customer are diverse, the customer might be familiar to them. For example, a system recommends best sellers only. The recommendations here are indeed diverse but the customer might already be familiar with the recommended products.



G. A Comparative Study of Recommendation Algorithm in E-Commerce Applications

Recommender systems are extensively used in applications that suggest various services, products and information items to customers. The most important step in building a recommendation system is to select an algorithm. The type of algorithm to be chosen is highly based on the type of input that the system takes.

In recommendation systems, there are generally three types of inputs: product aspects, customer aspects and previous interactions between customer and products. An example for the inputs can be the customers buying, the rating given by the customers and the browsing through the catalog that the customer does.

Some of the recommendation algorithms [12] are given below:

1. User-based Algorithm

The user-based algorithm is used to predict customer’s future purchases by collecting the purchase history of similar customers. The algorithm first computes a similarity matrix of all the customers. The score of similarity is calculated based on the row vectors of the interaction matrix A using a vector similarity function. A high similarity index tells us that two customers a and b may have similar inclinations since they have previously purchased a large set of common products. In other words, as the similarity between the target customer and a set of customer’s increases, the probability that a customer would buy the target product also increases.

2. Item-based Algorithm

The Item based algorithm makes use of the product similarities. A product similarity matrix is first computed and the similarity index is calculated from the column vectors of the interaction matrix. When a high index is obtained, it can be said that two products a and b are similar based on the fact that many customers have purchased both these products together. In other words as the similarity between the target product and the products bought by the target customer increases, the probability that the target customer would be interested in the target product also increases. The item-based algorithm is more efficient and provides better quality recommendations than the user-based algorithm for some of the databases.

3. Dimensionality Reduction Algorithm

The dimensionality reduction based algorithm is mostly used to reduce the sparsity of the interaction matrix. One of the examples of the dimensionality reduction algorithms is the standard singular value decomposition algorithm. In this algorithm, the interaction matrix is decomposed into 3 matrices U, Z and V’. The matrices U and V are orthogonal matrices and Z is a diagonal matrix. The matrix U represents the correlation between the user and the concept and V represents the correlation between the concept and the products. The matrix Z tells the strength of each user with each product. After getting the three matrices, the matrix Z is first reduced by keeping the first k i.e. k largest singular values to obtain the new version of the matrix with only k dimensions. Then using this, the other two matrices are also reduced to k dimensions. Now, these three new versions of matrices U, Z and V’ provides the best lower rank approximation of the original interaction matrix. By doing this, the primary data patterns are not affected and thus can be used effectively for making recommendations. To derive the

similarities between customers, the new versions i.e. the reduced versions of U and Z are used. After this, recommendations can be generated the same way as that in user-based algorithm. Now a suppositional interaction matrix is taken as input to illustrate the three recommendation algorithms introduced above.

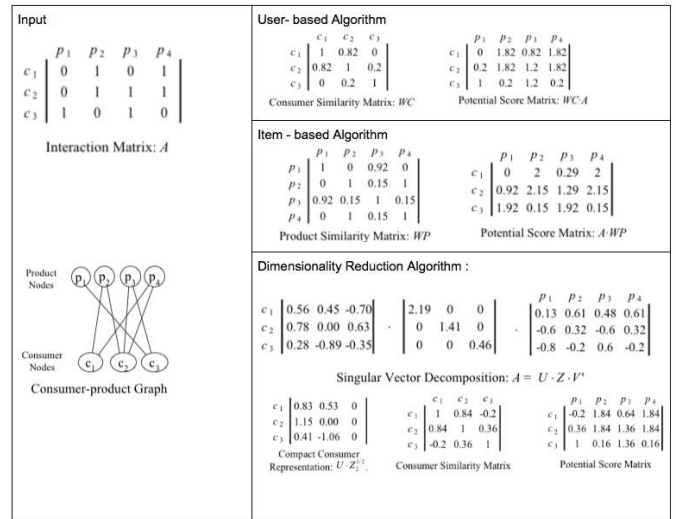


Figure 3: Three Recommendation algorithms illustrated with a simple example

In Figure 3, the interaction matrix and consumer-product graph representations of a simple transactional recommendation dataset is given. This dataset takes into consideration the information of 3 consumers, 4 products, and 7 observed transactions. The steps needed for computation of the three algorithms are present on the right side of Figure 3. As mentioned above, in user-based algorithms, the similarity between the customers is taken into consideration. Thus the customer similarity matrix is computed and then the similarity index is calculated as the dot product of this similarity matrix and the given interaction matrix. In item-based algorithm the product similarity. Matrix is shown as that is the first step in the algorithm. Next, the potential similarity index matrix is shown as the dot product of the interaction matrix and the product similarity matrix. It can be seen that there is little difference in the two score matrices and the two matrices have similar patterns in rows and columns. In the dimensionality reduction algorithm, it is known that the first step is the decomposition of the interaction matrix. So, the first step was computed and the results of the same are shown in the figure. Then the compact consumer representation based on a rank-2 approximation was calculated. Using this and cosine similarity, the consumer similarity matrix was found out. Then, as in the user based algorithm, the score matrix is calculated and shown in the figure 4. The evaluation measures that were taken into consideration are precision, recall, rank score and the area under Receiver Operating Characteristics (ROC) curve. Precision is the ratio of the number of hits to the number of products to be recommended. The recall is the ratio of the number of hits to the number of products that any customer c interacted with in the testing set. The measures precision and recall are sensitive to the number of products to be recommended.



Customer Segmentation and Buyer Targeting Approach

The ROC curve tries to measure the extent to which a learning system can successfully distinguish between relevant and non-relevant signals. Here, relevance is based on consumer- product pairs i.e. a recommendation that corresponds to a transaction in the testing set is deemed as relevant and otherwise as non-relevant.

The x-axis and y-axis of the ROC curve are the percent of non-relevant recommendations and the percent of relevant recommendations, respectively. From Figure 4, it can be deduced that when the unreduced data was used, all the algorithms had better performances. The difference in the performances increases when the total number of target customers is also considered. For the unreduced retail dataset, when the dimensionality reduction algorithm was used, it was found that the precision and recall were higher than when the other algorithms were used. When the computational efficiency analysis was done, it was seen that the dimensionality reduction algorithm takes the highest amount of time. All the three algorithms showed mixed performances under different datasets. The item-based algorithm performed exceptionally well for the movie datasets, but had relatively lower quality with the retail datasets and had the worst performance with the book dataset. The dimensionality reduction algorithm consistently achieved a mediocre performance across all datasets and performance measures and the user- based algorithm was almost always dominated by other algorithms.

Measure	Algorithm	Dataset						Avg Algorithm Rank
		Retail		Book		Movie		
		Reduced	Unreduced	Reduced	Unreduced	Reduced	Unreduced	
Precision	User-based	0.0041	0.0088	0.0103	0.0202	0.0460	0.0648	4.17
	Item-based	0.0044	0.0096	0.0045	0.0091	0.0460	0.0759	4.50
	Dimensionality Reduction	0.0050	0.0064	0.0097	0.0191	0.0384	0.0530	5.50
Recall	User-based	0.0359	0.0663	0.0534	0.1041	0.0503	0.0686	4.33
	Item-based	0.0359	0.0711	0.0251	0.0454	0.0538	0.0864	4.33
	Dimensionality Reduction	0.0411	0.0408	0.0564	0.1026	0.0428	0.0580	5.17
Rank Score	User-based	1.8750	4.6519	3.8270	7.8635	4.5250	7.4500	4.17
	Item-based	2.0313	4.8257	1.0261	3.1281	4.2167	8.7500	4.33
	Dimensionality Reduction	3.4896	3.0210	3.0227	6.9486	4.1000	6.8000	5.17
Area under ROC curve	User-based	0.4308	0.4628	0.3798	0.5287	0.7984	0.8544	5.67
	Item-based	0.4463	0.4810	0.3873	0.5558	0.8289	0.8930	4.33
	Dimensionality Reduction	0.4908	0.5663	0.5535	0.6894	0.8051	0.8602	3.00
# of target consumers		320	498	601	674	1000	1000	-

Figure 4: Experimental results: performance measures

III. CUSTOMER SEGMENTATION AND BUYER TARGETING APPROACH

A systematic approach to customer segmentation and buyer targeting for profit maximization is illustrated in Figure 5.

• Data Preparation

The data file containing all the orders is split into prior and train depending upon the eval set attribute. The prior orders file is merged with the products file. It is then merged with aisles to obtain a final data consisting of complete details of users and their orders. The dataset consist of number of variables which are correlated to each other heavily or lightly. For clustering, it is necessary to determine the most appropriate variable(s) for differentiating distinct data points. Principal Component Analysis can be used to solve the above purpose.

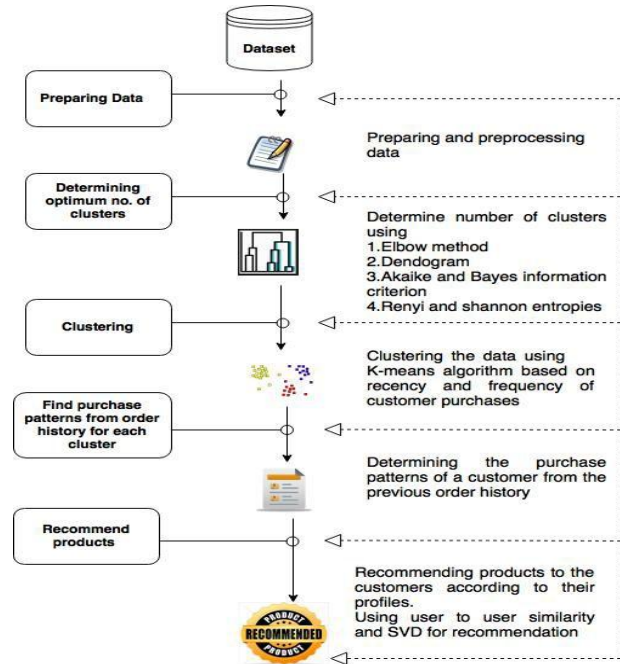


Figure 5. System Design

Determine optimum number of clusters

• Elbow method

The elbow method runs K-means algorithm for different values of K. The sum of the squared mean is calculated for each K. Using these values, a graph is plotted between the sum of squared error (SSE) and the K values [13]. 6.

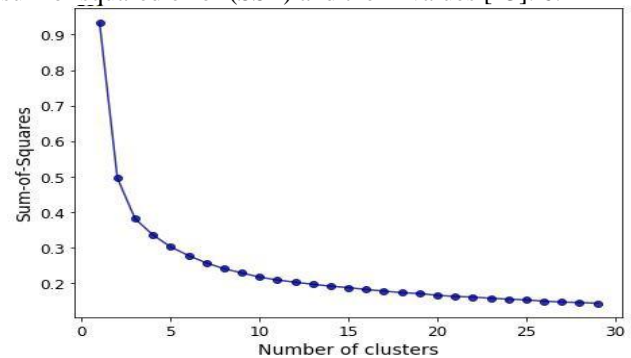


Figure 6. Number of clusters using Elbow method

Our aim is to find a value of K that is small and has a lower SSE. In the line graph, an elbow i.e. a bend shows such a value. This value of K is the optimum number of clusters. To verify the value of K, double differentiate the SSE and plot this against the K-values. A spike in this graph shows the value of K for optimum clustering, as shown in Figure.

• Dendrogram

A dendrogram is a tree diagram which is formed using hierarchical agglomerative clustering. The dendrogram shows the distance between the clusters and the various clusters formed at each level. In the graph, when there is a clear distance gap between the clusters then choose the number of clusters as the total clusters present at that level. The dendrogram is a good way for verifying the K value rather than interpreting it [13]. From Figure 7, it can be seen that for k value as 3, there is clear distance between all the three clusters. This tells us that the clusters are distinctive.



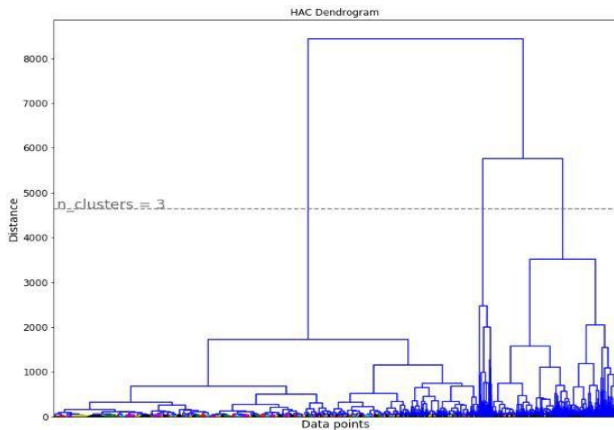


Figure 7. Number of clusters using Dendrogram

• Akaike and Bayes Information Criterion

Akaike and Bayes Information Criterion is used to select appropriate statistical model. It is similar to likelihood function. It is also used to eliminate the issue of overfitting which often occurs in machine learning. For a fit model, low value of Bayes Information Criteria is expected. Equations (1) and (2) give the formulas for the computation of AIC and BIC.

$AIC = 2k - 2 \log_e(L)$ equation 1

$BIC = \log_e(n) * k - 2 \log_e(L)$ equation 2

L : maximum value of likelihood function

k : estimated parameters of the model

n : number of sample points in data

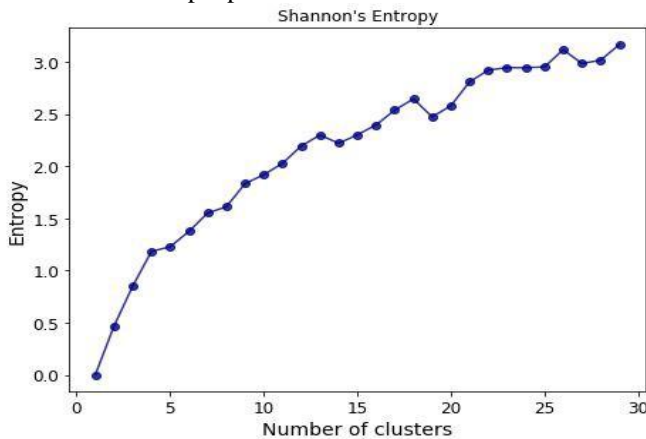


Figure 8: Number of clusters using Elbow method

• Renyi and Shannon Entropies

The Renyi and Shannons entropies are the measure of the randomness and uncertainty of the data. They are calculated based on the probabilities. A graph can be plot of the number of clusters against the entropies. Hence, decision of the number of clusters to be made can be made on the basis of the occurrence of knee in the graph. In Figure 8, the Shannon entropies are calculated and plotted for various values of k. The knee at 3 tells us that it is a good estimate for the number of clusters.

• K-means Algorithm for clustering based on recency and frequency

K-means is an unsupervised method used for clustering. It is largely used because of its efficiency and simplicity. If compared with hierarchical clustering, k-means clustering has a time complexity of O(n) and hierarchical clustering has a time complexity of O(n²). Hence, k-means clustering works better for large dataset. Also, hierarchical clustering is used

when there exists complex shapes of clusters and no predefined number of clusters [14]. Moreover, as complexity of computation hierarchical clustering is high, that hinders its use for clustering in real time applications [15]. Hence the use of K Means algorithm is made for carrying out the task of clustering. K means will work as follows:

- As is the prerequisite of k-means clustering, determine the number of clusters.
- Choose initial centroids randomly.
- Calculate the euclidean distance which is the square of the distance of each data point and the centroid.
- Form clusters by assigning each data point to the cluster with least euclidean distance.
- Calculate the new centroids from clusters obtained.
- Perform number of iterations to give the desired number of optimum clusters based on recency and frequency.

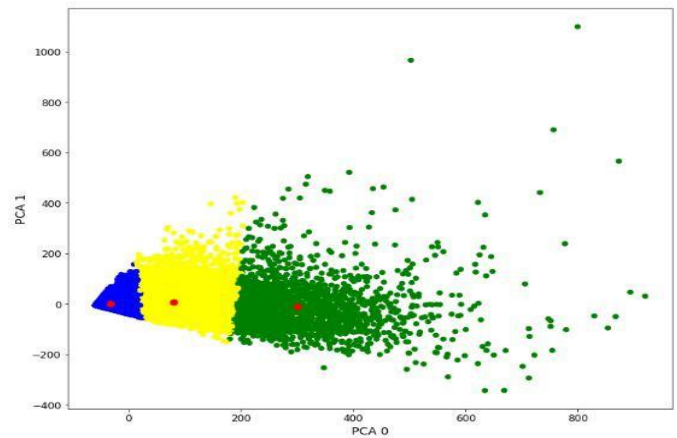


Figure 9: Cluster data points scatter plot

Hence, after application of k-means clustering algorithm to the dataset, the scatter plot obtained between principle components recency and frequency is shown in figure 9.

• Recommendation using Singular Value Decomposition

The model of a recommender system is represented using a utility matrix where every cell represents a users response to a product (whether it was purchased or not, if yes, the number of times it was purchased). In a utility matrix, not all features are of significant importance hence SVD is used to remove redundant features making it easy to store, process and visualize data [16]. It is an efficient technique for matrix factorization that factors an m * n matrix R into three matrices.

$R = U \cdot \Sigma \cdot V^T$Equation 3

In (3), R represents the m * n user-item utility matrix, U represents the m *r user-concept similarity matrix and V represents the r * n product-concept similarity matrix is a singular r * r matrix that holds the strength of each concept. For computing recommendations, calculate the dot product of the user-concept and product-concept matrices which gives us the predicted scores for every user to every product. Based on these scores in descending order of values, determine which products are most probable to be bought by the particular user and hence recommend the same.

Customer Segmentation and Buyer Targeting Approach

However, this might recommend some products that have already been bought by the customer, hence, to avoid this check the main utility matrix and consider scores only for those products that were absent in the utility matrix.

• Recommendation using User to User similarity

In this method, first compute an $m * n$ utility matrix consisting of cluster wise user to product responses. The top N similar users within the respective cluster for every user in the dataset are found using cosine similarity and obtain the top products bought by all the similar users combined. To check the similarity, the recall value by comparing purchased products for these sets of users is calculated. A high recall value corresponds to high similarity. For finding recommendations, the combined list of products bought by top N similar users is traversed masking the products already purchased by the targeted user and among these, the top K products are found appropriate for recommendation.

IV. RESULTS AND DISCUSSION

The Instacart Grocery Shopping Dataset 2017 is anonymized and contains purchase information of over 3 million orders of grocery from more than 200,000 users of Instacart [17]. Figure 10 describes the relationship model of the dataset.



Figure 10: Relationship model of dataset.

The number of orders per user between 4 and 100, the date and time of the orders placed and time between orders are also mentioned for each user.

• Exploratory Data Analysis

The Exploratory data analysis(EDA) was performed on the dataset before obtaining the appropriate recommendations. Through Principal Component Analysis, recency and frequency were found to be the most appropriate parameters for clustering. The number of clusters considered here is 3 which was found using elbow method and dendrogram. Based on the number of clusters obtained, recency and frequency analysis was carried out.

The reorder ratio graph as shown in Figure 11 gives an account of the frequency of orders of the customers. It helps to deduce that Cluster 2 consists of the most frequent customers, Cluster 1 with moderate frequency customers and Cluster 0 customers being rare visitors.

Figure 12 shows the recency of orders by taking into consideration the number of days since the last order was placed. According to the graph, it can be deduced that Cluster 0 customers tend to order in a span of 30 days, whereas Cluster 1 and 2 customers had last ordered in within a week's time.

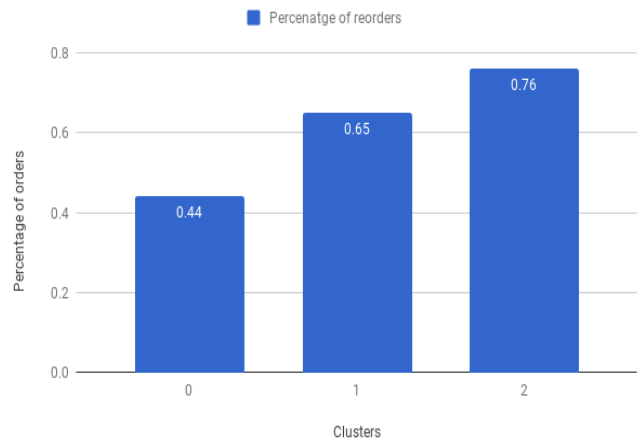


Figure 11.Reorder ratio per cluster

Based on the results obtained in Figure 11 and Figure 12, the following conclusions can be drawn:

1. Cluster 0 customers exhibit low recency and frequency, hence they fall into 'At Risk' category

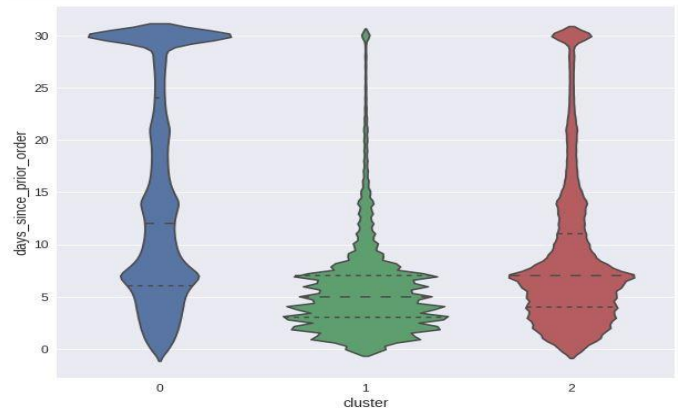


Figure 12. Days since prior order for every cluster

2. Cluster 1 customers exhibit high recency and high frequency, hence they fall into 'Champions' category
3. Cluster 2 customers exhibit moderate recency and moderate frequency, hence they fall into 'Potential Loyalists' category.

The most popular products were determined for every cluster as shown in Figure

13. It helps to solve the cold start problem by recommending common popular products of every cluster as the cluster of a new customer is initially unknown. The recency and frequency data analysis can be used for further recommendations.

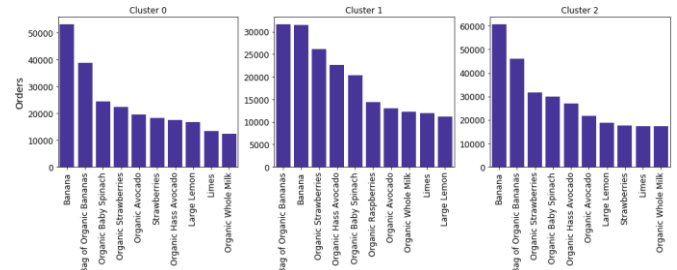


Figure 13. Most popular products per cluster

V. FUTURE SCOPE

The dataset provided by Instacart Grocery Shopping is been used for the purpose customer clustering and buyer targeting for profit maximization in this project. However, valuable information can also be acquired with the help of RFID i.e. radio frequency identification, instead of explicitly collecting it from the customer purchase details.

RFID or Radio frequency identification is an emerging tracking technology that makes use of small tags that emit distinct signals. Retail industry owners can make use of remote scanners to read RFID tags attached to individual products, thus enabling them to record information, including quantities of stock items and their precise locations in the store.

VI. CONCLUSION

Customer satisfaction can be achieved by employing a model that accurately predicts customer preferences and hence channels resources to gain maximum profit. Clustering was done to obtain groups of customers that exhibit similar behavior and hybrid recommender system was used to eliminate drawbacks that are experienced when only a single recommender system is used. Accurate recommendations were provided to customers by dealing with the common challenges of recommender systems such as overspecialization, sparsity, and cold start problem. Using Principal Component Analysis (PCA), it was found that Recency and Frequency were the components that resulted in maximum variance and then used K Means algorithm for clustering. Singular Value Decomposition was used as an algorithm for recommendation to deal with sparsity problem while User-User similarity recommendations were made using Cosine similarity to deal with cold start problem. The performance of the approach can be evaluated using the aforementioned performance metrics after online experimentation.

REFERENCES

1. R. Siva Subramanian, Dr. D. Prabha, 'A Survey on Customer Relationship Management', International Conference on Advanced Computing and Communication Systems, Coimbatore, India, January 2017.
2. Witschel H.F., Loo S., Riesen K. 'How to Support Customer Segmentation with Useful Cluster Descriptions' Advances in Data Mining: Applications and Theoretical Aspects, vol. 9165, Springer, 2015.
3. Ina Maryani, Dwiza Riana, 'Clustering and profiling of customers using RFM for customer relationship management recommendations' 5th International Conference on Cyber and IT Service Management, Denpasar, Indonesia, August 2017.
4. Anshu Sang, Santosh K. Vishwakarma, 'A Ranking based Recommender System for Cold Start & Data Sparsity Problem'
5. D. Prabha, R. Siva Subramanian, 'Data mining techniques for CRM', International Conference on Information Communication and Embedded Systems (ICES), Chennai, India, February 2014.
6. Mrs. Pradnya Muley, Dr. Anniruddha Joshi, 'Application of Data Mining Techniques for Customer Segmentation in Real Time Business Intelligence', International Journal of Innovative Research in Advanced Engineering (IJIRAE) Issue 4, Volume 2, April 2014.
7. Raj Bala, Sunil Sikka, Juhi Singh, 'A Comparative Analysis of Clustering Algorithms', International Journal of Computer Applications, pp. 35-39, August 2014.
8. Ilung Pranata, Geoff Skinner, 'Segmenting and targeting customers through clusters selection & analysis', under review for International Conference on Advanced Computer Science and Information Systems, Depok, Indonesia, October 2015.

9. Marius Kaminskas, Derek Bridge, Franclin Foping, Donogh Roche, 'Product Recommendation for Small-Scale Retailers', 16th International Conference on Electronic Commerce and Web Technologies, vol. 239, Springer, Spain, September 2015.
10. Pratik Ghanwat, Anu Chacko, 'Improved Personalized Recommendation System with Better User Experience', International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udipi, India, September 2017.
11. Dr. Sarika Jain, Anjali Grover, Praveen Singh Thakur, Sourabh Kumar Choudhary 'Trends, Problems And Solutions of Recommender System' International Conference on Computing, Communication and Automation, Noida, India, May 2015.
12. Zan Huang, Daniel Zeng, Hsinchun Chen, 'A Comparative Study of Recommendation Algorithms in E-Commerce Applications', Proceedings of the IEEE Region 10 Conference, The Pennsylvania State University, pp.1-23. Available: <http://citeseerx.ist.psu.edu>. [Accessed: March 08, 2018]
13. Trupti M. Kodinariya, Prashant R. Makwana, 'Review on determining number of Cluster in K-Means Clustering', International Journal of Advance Research in Computer Science and Management Studies, vol. 1, pp. 90-95, November 2013.
14. P. Praveen, B. Rama, 'An Empirical Comparison of Clustering using Hierarchical methods and K-means', 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bioinformatics, Chennai, India, February 2016.
15. Xiaojun Chen, Yixiang Fang, Min Yang, Feiping Nie, Zhou Zhao, Joshua Zhexue Huang, 'PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data', IEEE Transactions on Knowledge and Data Engineering, vol. 30, pp. 559-572, March 2018.
16. Naznin Akter, A. H. M. Sajedul Hoque, Rashed Mustafa, Mohammad Sanaullah Chowdhury, 'Accuracy Analysis of Recommendation System Using Singular Value Decomposition', 19th International Conference on Computer and Information Technology, North South University, Dhaka, Bangladesh, December 2016.
17. The Instacart Online Grocery Shopping Dataset 2017.[Online]. Available: <https://www.instacart.com/datasets/grocery-shopping-2017>.

AUTHORS PROFILE



Prof. Yogesh Jadhav Currently working as Assistant Professor at Amity School of Engineering and Technology, Amity University Mumbai. He completed M.E. in Computer Engineering from Mumbai University.



Dr. Deepa Parasar Currently working as Associate Professor at Amity School of Engineering and Technology, Amity University Mumbai. She completed PhD in Computer Science and Engineering.