Journal of Engineering Sciences
Vol 13, Issue 04, APRIL/2022
ISSN NO:0377-9254

CATEGORIZATION OF CUSTOMERS USING RFM

Dr. BHAVANI M[1], ALAPAKA HADARSH[2], KATADY SAI SHIRISH[3], GANDRAJU RISHITA[4], MEKALA SAI MURALI MANOHAR[5],

#1 Assistant Professor, Department of CSE, GITAM (deemed to be University), Gandhi nagar Rushikonda Visakhapatnam 530045 Andhra Pradesh, INDIA

#2,#3,#4,#5 Student, Department of CSE, GITAM (deemed to be University), Gandhi nagar Rushikonda Visakhapatnam 530045 Andhra Pradesh, INDIA.

ABSTRACT_ Customer relationship administration is a advertising and marketing concept, which consists purchaser identification, attraction, retention and development. We can say that CRM is a method of acquiring, holding clients to get excessive cost for the enterprise and the customer. CRM approach affords integration of technology, approaches and all commercial enterprise works associated to the customers. This venture investigates how to combine algorithms and Analysis of Recency, Frequency, and Monetary value.This evaluation can be carried out in on-line buying and selling to grant techniques based totally on consumer buying behavior. To preserve and promote the business, many components are necessary to be viewed to make certain that advantages increase. Customer segmentation helps agencies to goal their offerings and prioritizing merchandise on the groundwork of its gain. Success of an company relies upon on attracting and preserving loyal customers. In this project, we utilized a number of Clustering Algorithms on Customers information to phase the clients into clusters in which every cluster has precise characteristics. By evaluating the above algorithms, we can discover key patron clusters based totally on Recency, Frequency, Monetary values.

1.INTRODUCTION

The RFM is used in the analysis of customer values. It consists of three values. The first is “Recency” which tells how recent the customer has made a transaction or purchase. Second is “Frequency” which tells how many times did the customers make a transaction or purchase. The Third is “Monetary” which is the sum of all amounts that the customer has spent buying. Using these three values, clusters are to be made by considering the scaled values of RFM. Each cluster has its definite

characteristics. By understanding the characteristics of each clusters, customer groups can be identified.

2.LITERATURE SURVEY

"This research built clustering models on customer profile data based on their usage of Internet Banking for customer segmentation in XYZ bank using K-Means method and K-Medoids method. The performance of clustering result was measured and compared. K-Means method outperformed K-Medoids method based on intra cluster (AWC) distance. While based on DBI, K-Means performs slightly better than K-Medoids."[2]

"The proposed procedure has shown that for the purpose of clustering, when we combine the expanded RFM model into K-means algorithm, we can see a tremendous improvement in classifying accuracy in order to reach to an excellent CRM."[1]

"In the past, retailers, banks, insurance companies and car dealers had close ties with their customers and knew what customers want; so they tried to grant their needs and wishes by offering special services to them. Later, with the arrival of mass production and marketing and increased number of consumer customers, the importance of building relationships with customers was reduced and the variety of products and their prices also declined. There are a lot of evidences indicating that the customer relationships were taken into consideration since the late 19th century. Today, through the effective use of information technology and communication, organizations can provide their customers with diverse products with lower prices and special services simultaneously. One of the most effective tools to study customers' behaviour is using clustering techniques. In this chapter, the use of segmentation to segment customers and its importance are described, then questions and related objectives are developed and finally the main research variables are defined."[3]

"Three machine learning methods, namely NEM, LIRM and LORM, are applied to segmented customer data of a company. The data has two features which are number of payments and total amount of payments for each customer. This study proposes to

solve a customer segmentation problem of a company by using customers' payment information. The methods are inherited from machine learning literature, applied to real data and it is shown that logistic regression method is a good choice for this segmentation problem. The outputs of this study can be implemented in modern customer relationship management (CRM) software."[4]

3.PROPOSED SYSTEM

Machine learning makes the system automatically learn and also to improve through experience without the need for being separately programmed.

Machine learning emphasizes on generation of computer programs that can use data to learn. Machine Learning is related to how to construct programs which automatically gets improved through experience.

3.1 SUPERVISED MACHINE LEARNING

In Supervised Machine Learning, the data is represented as labelled and the algorithm gets learned from the training data that is labelled.

Classification and Regression are the examples. Supervised Machine learning uses training data to teach the models to give the required output.

The training dataset does includes input values and correct output values of the attributes, that allows the model to get learned over the time.

3.2 UNSUPERVISED MACHINE LEARNING

Contrary to Supervised Machine Learning, Unsupervised Machine Learning uses the data that is not labelled.

From that unlabelled data, it finds patterns which help to solve clustering problems.

Examples: Hierarchical Clustering, K-Means Clsutering.

3.3 CLUSTERING

Clustering is an unsupervised machine learning algorithm. It's used for finding natural clusters in the dataset. Clustering algorithms analyse the input data and discover the natural clusters in the data.

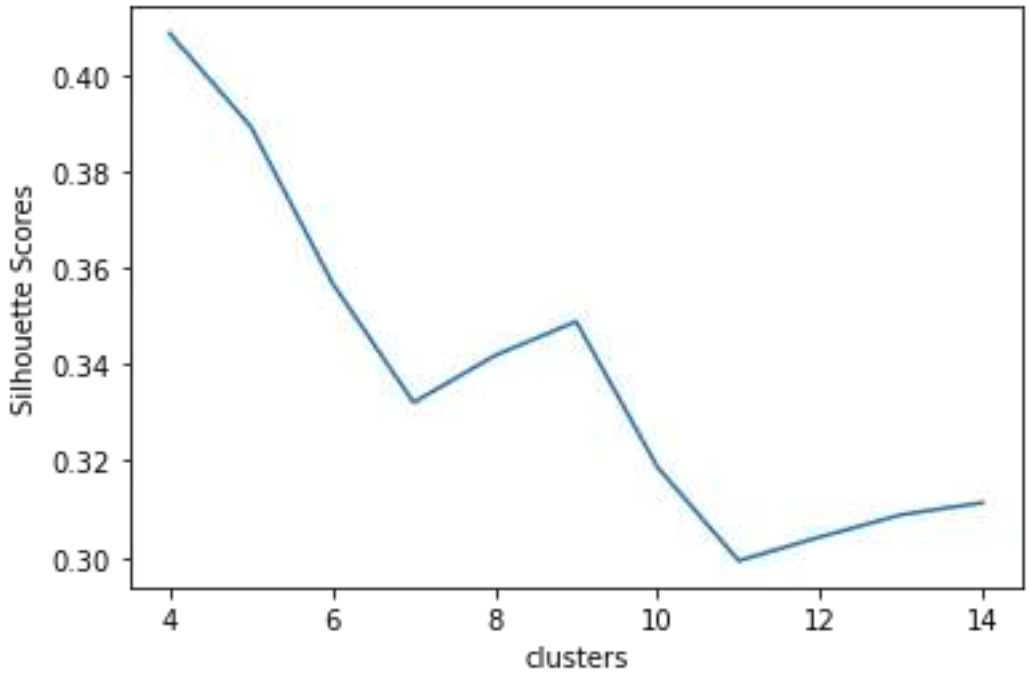
3.4 K-MEANS CLUSTERING

Unlike supervised learning algorithms, K-means clustering is an unregulated machine learning algorithm. K-Means is used when we have non-labeled data (Data with no categories or groups). Our customer segregation data is like this problem.

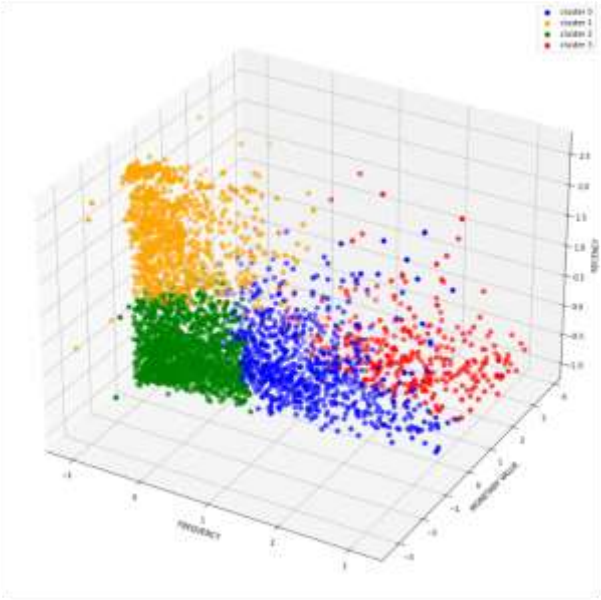
The algorithm finds Clusters in the data, in which the number of clusters is represented by the value of K. The algorithm acts repetitively to give each data point to one of K clusters, according to the features. This makes K-Means suitable for the Customer Categorization problems.

Given a set of data points are grouped as per feature similarity.

At the end, we are going to get various clusters along with cluster ids to which the customer belongs to.



The optimal value for k is 4.



3.5 AGGLOMERATIVE CLUSTERING

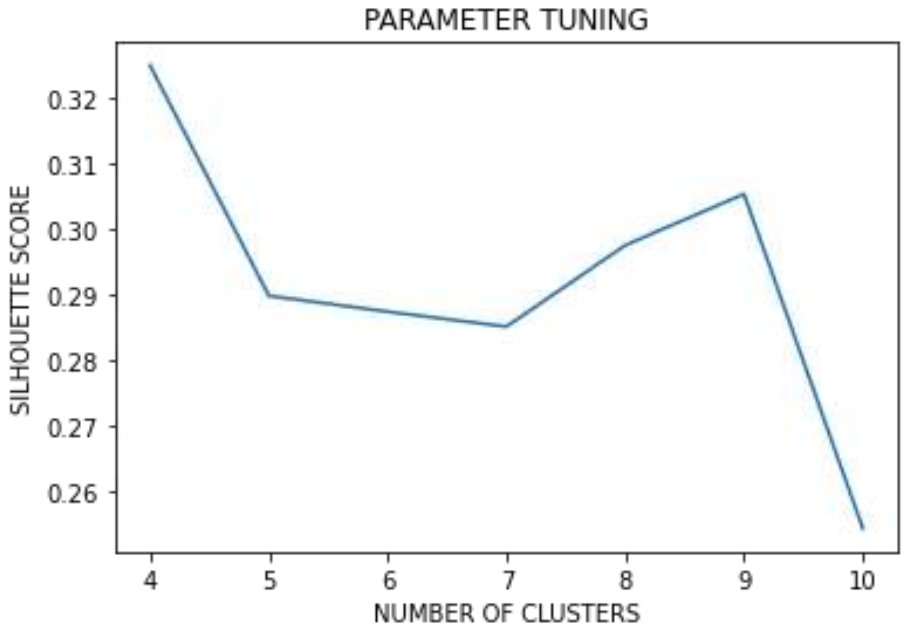
Agglomerative clustering is the most popular clustering algorithm that is used to classify points according to their similarity. The algorithm begins by considering each item as a singleton group. Next, the pairs of groups are grouped into sequences until all the clusters have been grouped together into a large collection which consists of all the items. The output is a tree representation of elements, known as a dendrogram.

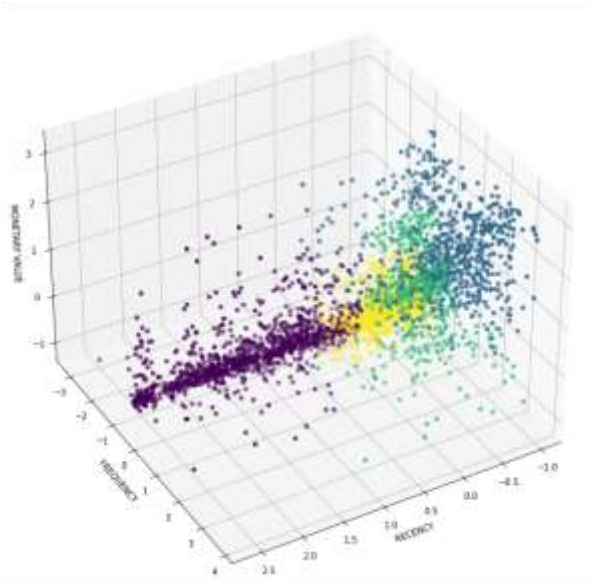
Agglomerative clustering is a “bottom-up” approach. Each object is initially considered as a single element (leaf). In each step of the algorithm, two very similar groups are grouped together into a large group. This process is repeated until all the points become part of one large group (root).

To determine which items / collections should be merged or separated, we need the methods to measure similarities between elements.

There are ways to calculate similarity, like Euclidean distance and Manhattan distance.

The linkage function considers distances and then combines pairs of elements into clusters based on the similarity. These newly formed groups are linked to one another to form larger collections. This steps are repeated until all the elements in the dataset are connected together in the tree.





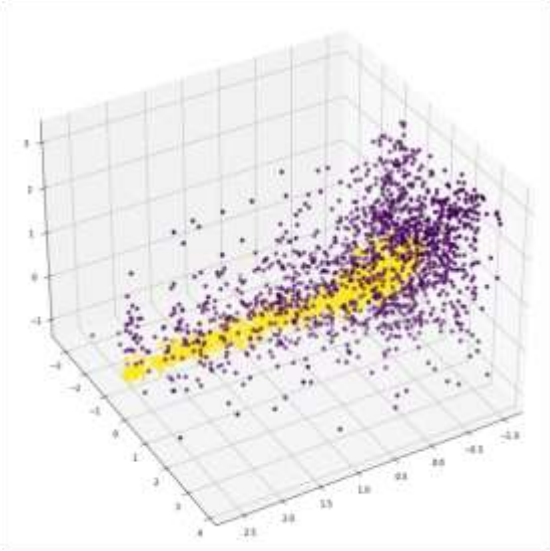
2.5 DBSCAN

DBSCAN is a widely used clustering algorithm which is used for data mining and machine learning. According to the set of points, DBSCAN collects adjacent points according to the distance measurement like Euclidean distance and a minimum points paarmeter. It marks the exterior of points in densely populated areas.

The DBSCAN Clustering algorithm requires two parameters:

Epsilon: Represents how close the points should be in order to be considered as part of a cluster. This means that if the Euclidean distance between two points is less than or equal to Epsilon, these points are considered to be neighbors.

minPoints: The minimum number of data points in order to form a cluster.

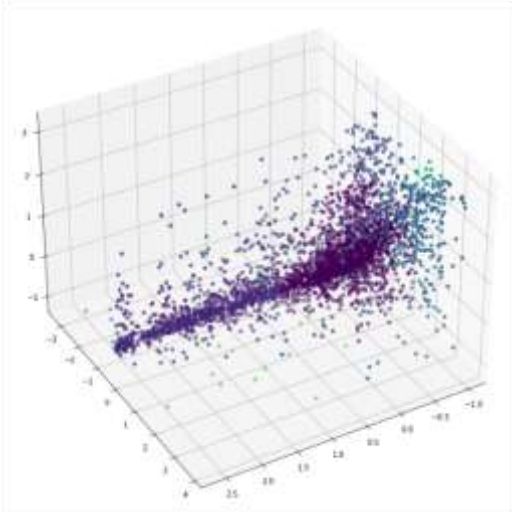


3.6 MEANSHIFT CLUSTERING

Mean shift is an unsupervised Machine Learning clustering algorithm which is known to be used for clustering. It is mostly used in image segmentation because it does not have any parameters and does not require any shape of the clusters to be pre-defined in the feature space.

"Mean Shift" means "Shifting to the Mean" in an iterative manner. In the algorithm, every data point shifts to the "regional mean" interatively and the location of the final position of each point represents the cluster to which it belongs to.

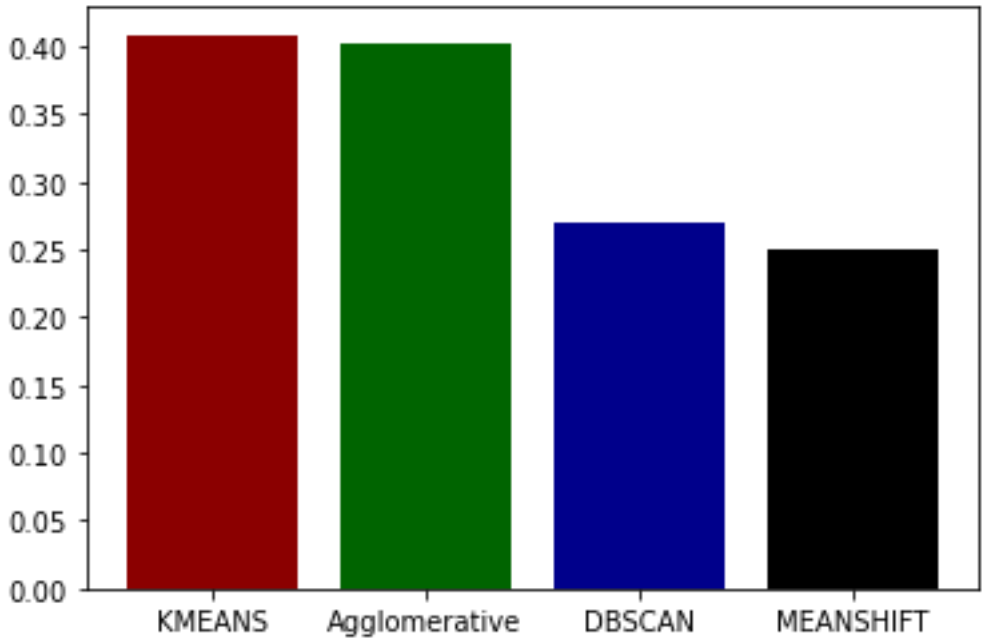
In mean shift algorithm, each point tries to find its cluster by moving towards the weighted mean of its local area in each iteration. The final position of each point will be the centroid of the cluster that the point belongs to. Then, all the data points with the final position point can be labeled with the same cluster.



4.RESULTS AND DISCUSSION

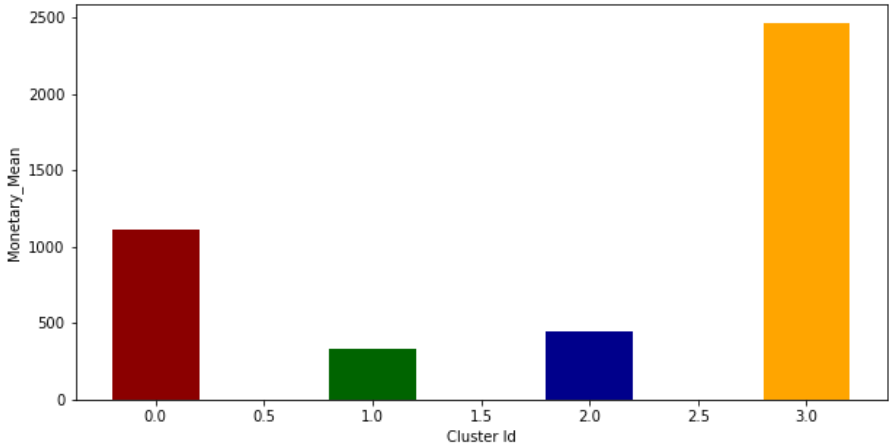
After implementing the clustering algorithms KMeans, Agglomerative, DBSCAN and Meanshift, we found that the KMeans algorithm shows good quality clusters over others. So, segments were done based on the results given by KMeans clustering.

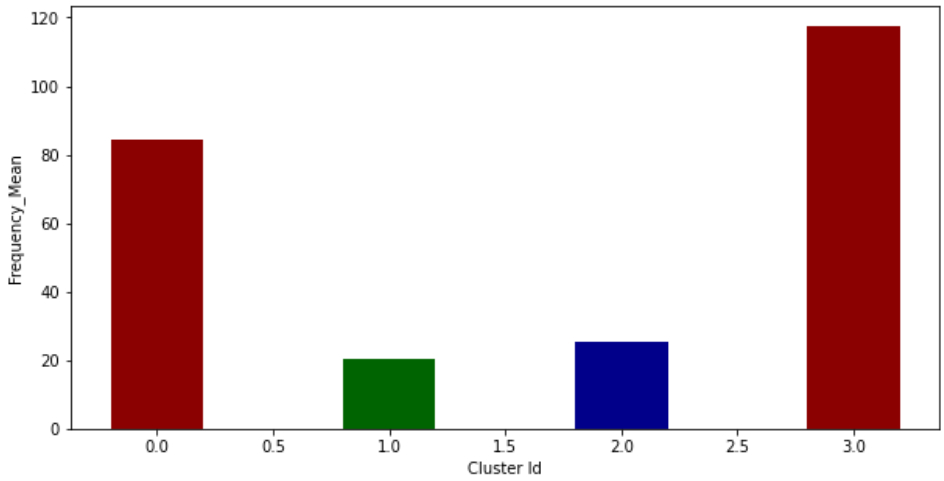
METHOD	SILHOUETTE SCORE
K-MEANS	0.40851721359807264
AGGLOMERATIVE CLUSTERING	0.3249143321884565
DBSCAN	0.2692400279336455
MEANSHIFT	0.2505909231704478

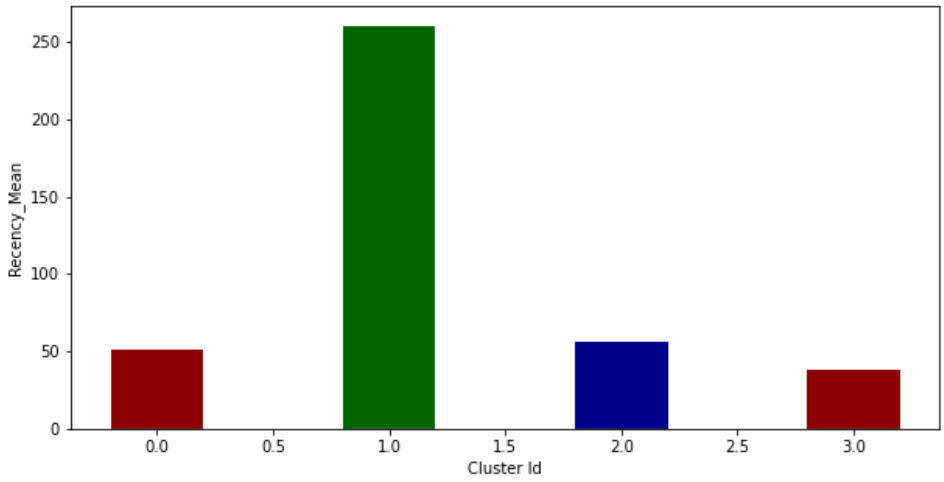


Four segments were formed each of which has definite characteristics.

Four clusters were formed. The analysis is shown below.







CLUSTER 0 : Moderate Monetary value, High Frequency value, Low Recency value - GOOD VALUE LOYAL CUSTOMERS

CLUSTER 1 : Low Monetary value, Low Frequency value, High Recency value - LOST UNSATISFIED CUSTOMERS

CLUSTER 2 : Low Monetary value,Low Frequency value, Low Recency value - UNSATISFIED CUSTOMERS

CLUSTER 3 : High Monetary value, High Frequency Value, Low Recency Value - BEST VALUE CUSTOMERS

5.CONCLUSION

This Project built various clustering models on Retail Customer dataset based on Recency and spending behaviours of the customers for customer segmentation. The performances of all the clustering models was measured and are compared. K-Means model outperformed Agglomerative Clustering, DBSCAN and Meanshift Clustering based on Silhouette Scores. The optimal number of clusters for KMeans is 4, according to the highest value of Silhouette Score.

REFERENCES

[1] I. Maryani, D. Riana, R. D. Astuti, A. Ishaq, Sutrisno and E. A. Pratama, "Customer Segmentation based on RFM model and Clustering Techniques With K-Means Algorithm," 2018 Third International Conference on Informatics and Computing (ICIC), 2018, pp. 1-6, doi: 10.1109/IAC.2018.8780570.

[2] M. Aryuni, E. Didik Madyatmadja and E. Miranda, "Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering," 2018 International Conference on Information Management and Technology (ICIMTech), 2018, pp. 412-416, doi:10.1109/ICIMTech.2018.8528086.

[3] Karim Hamdi and Ali Zamiri, 2016. Identifying and Segmenting Customers of Pasargad Insurance Company Through RFM Model (RFM). International Business Management, 10: 4209-4214.

[4] Ş. Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1-6, doi: 10.1109/IDAP.2018.8620892.

[5] Nimbalkar, Divya & Shah, Paulami. (2013). Data mining using RFM Analysis. 10.13140/RG.2.2.24229.04328.