iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36105

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Customer Segmentation using Data Mining Techinique

Medha
Computer Science and Engineering Department, ABES Institute of Technology

Abstract: Data mining technique can help implementing customer relationship management dimensions namely customer identification, customer attraction, customer retention and customer development . Various Data mining functions can be used to implement CRM elements . This review paper is focused on one of the element of customer identification that is customer segmentation and its implentation using K-means clustering technique. It discuses in breif about the K-means Algorithm.
Keywords: Data mining, Customer Relationship Management, Customer identification, Customer attraction, Customer retention Customer development, K-means clustering

I. INTRODUCTION

Customer Segmentation is the process of dividing customes into several groups that share common interest and characteristics in different ways that are relevant to marketing such as gender, age, monthly income and spendings. Customer segmentation is done under the notion that every individual has different requirements .These requirements when addressed with a specific marketing effort can help gaining more profit. Customer segmentation is one of the element of Customer identification along with Target customer analysis. Customer identification is one of the four dimensions of CRM i.e Customer relationship management.

CRM is a strategy and process of acquiring, retaining, and partnering with a group of customers to create superior value for the company. It involves the integration of marketing, sales, customer service, and the supply chain functions to achieve greater efficiencies in increasing customer value.

Data mining techniques can help in implementing customer segmentation by detecting and extracting hidden customer characteristics and behaviours from large data in the databases . The generative aspect of data mining consists of the building of a model from data so that the data can be used for various purposes.

II. RELATED WORK

Khalili-Damghani proposed a hybrid soft computing approach on the basis of clustering, rule extraction, and decision tree methodology to predict segmentation of new customers of customer-centric companies. This approach was applied in two case studies in the fields of insurance and telecommunication, respectively, aiming to predict potentially profitable leads and to outline the most influential features available to customers during prediction.[1]

Yizhang Jiang applied data mining methods such as c-means , transfer learning , and multiview learning in brain CT, EEG image segmentation, and multiview clustering research.[2]

Cheng & Chen used a RFM model and K-means to perform customer relationship management, and experimental results demonstrate that the model they proposed is an effective method in customer value analysis.[3]

Jun Wu, Li Shi, Wen and Sang in their paper " An Empirical Study on Customer Segmentation by Purchase Behaviors using a RFM Model and K-Means Algorithm " used K-means algorithm to cluster according to R-F-M attributes. The accuracy of their algorithm depends on initialization conditions and the number of clusters.[4]

The K-means algorithm is the most widely algorithm to effectively identify valuable customers .The literature shows that one of the major applications of K-means is customer segmentation.

According to Bruce,Larzan and Timothy in their paper " Approaches to Customer Segmentation " Segmentation can also be done based on Supervised classification using Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). SVM correct classification rate was generally about 2% higher than the LDA rate, irrespective of size of training set . Clustering , Conjoint Analysis and Latent class models can also be used for segmentation [5]

Dennis, Marsland and Cockett (2001) in their paper " Data Mining for shopping Centres - Customer Knowledge Management Framework " depicted the use of Clustering analysis for data mining of shopping centres where in shoppers were classified accordingly by the importance of shop and the Service.[6]

According to E.W.T. Ngai, Li Xiu, D.C.K. Chau in their paper "Application of data mining techniques in customer relationship management: A literature review and classification " ,Logistic regression Decision Trees, Pattern Based Cluster and Self organizing Maps can also be used for Customer Segmentation.[7]

©IJRASET: All Rights are Reserved

Data mining technique can help implementing customer relationship management dimensions namely customer identification, customer attraction, customer retention and customer development . This review paper is focused on one of the element of customer identification that is customer segmentation and its implentation using clustering technique.

III.IMPLEMENTATION

A. K-means Clustering Algorithm

K-Means Clustering is an unsupervised learning algorithm. Unsupervised learning algorithm is an algorithm that learns itself from the unlabeled dataset . It finds the hidden patterns and relevant information from the data and act on that data. K-means follows an iterative approach, dividing the unlabeled dataset into k different clusters such each datapoint belong to only one cluster and has similar properties as the other datapoints in that cluster. Each cluster in the model is associated with a centroid and hence it is centroid based algorithm.

The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters in which they are present. It tries to make the data points inside the cluster as similar as possible while also keeping the different clusters as far as possible.

The steps involved in K-means clustering are as follows ,

1) We determine the number of clusters to be created i.e k.
2) Algorithm select the the k datapoints from the cluster that serves as the initial centroids of our clusters.
3) Each datapoint is assigned to nearest centroid.
4) After each datapoint is being assigned to a cluster, the algorithm calculates a new mean value of each cluster and calculates a new centroid.
5) Each datapoint is again checked and is assigned to the nerest new centroid formed.
6) Above two steps are repeated until clusters that are formed in a given iteration is similar to the clusters that were formed in the previous iteration.

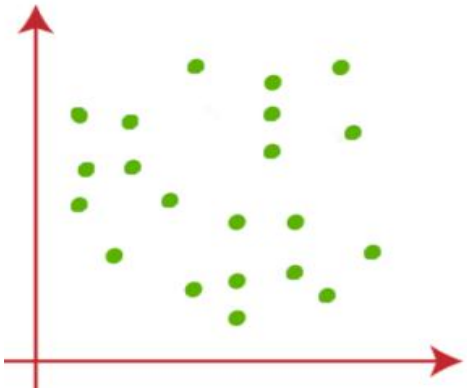


Fig. 1 A set of datapoints

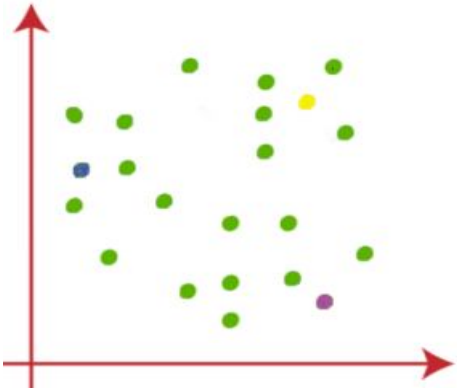


Fig. 2 Initial centers chosen

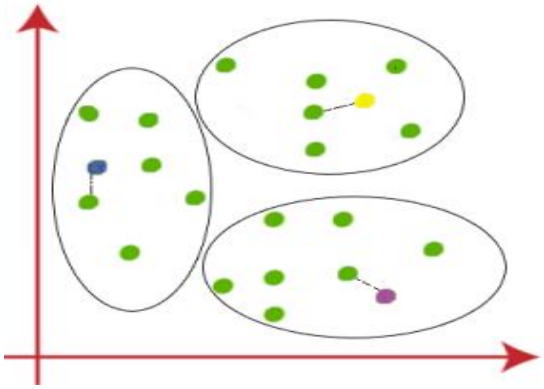


Fig. 3 Iterative process of calculating new centroids

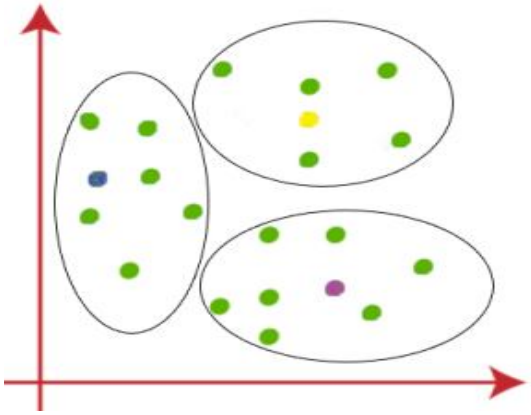


Fig. 4 Final clusters are created

©IJRASET: All Rights are Reserved

There are three methods that can be used in order to determine the optimal value of k are,

a) Elbow method: A Elbow method is based on the concept of WCSS value which is given as,

$$WCSS = \sum Pi \in C1 \text{ distance}(Pi\ C1)2 + \sum Pi \in C2 \text{distance}(Pi\ C2)2 + \sum Pi \in C3 \text{ distance}(Pi\ C3)2$$

Basically a graph if being plotted between number of clusters and sum of squared distance and that value K is taken where the curve starts forming an elbow. Therefore, this method is known as elbow method.

b) Silhouette method: Silhouette method is used to measure the quality of the clustering operation. It is performed for different values of K. If we obtain a high average silhouette width, it means that we have good clustering for that particular value of K . Coefficient can be calculated as,

$$\frac{b^i - a^i}{\max(a^i, b^i)}$$

Where,

ai represents the average distance from all the observation in same cluster

bi represents the average distance from all the observation in closest cluster

c) Gap Statistic method: Researchers at Stanford University – R. Tibshirani, G.Walther and T. Hastie published the Gap Statistic Method in the year 2001.It can be for any of the clustering method like Kmeans, hierarchical clustering etc. The clusGap function can be used for providing gap statistic as well as standard error for a given output. It is basically used to compare the intracluster variation for different values of K .

IV. CONCLUSIONS

Customer segmentation can be implemented using K-means algorithm. Through the process of customer segmentation companies can gain deep understanding of customers needs and preferences. Valuable segments of customers can be targetted in order to reap maximum benefit . Companies can make more efficient stratergies to retain more customers. By using this, they can also minimize the risk of investment.

REFERENCES

[1] K. Khalili-Damghani, F. Abdi, and S. Abolmakarem, “Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: real case of customer-centric industries,” Applied Soft Computing. vol. 73, pp. 816–828, 2018.

[2] Y.Jiang,F.L.Chung,S.Wang et.al., “Collaborative fuzzy clustering from multiple weighted views,” IEEE Transaction on Cybernetics vol. 45, no. 4, pp. 688–701, 2015.

[3] C.-H. Cheng and Y.-S. Chen, “Classifying the segmentation of customer value via RFM model and RS theory,” Expert Systems with Applications, vol. 36, no. 3, pp. 4176–4184, 2009.

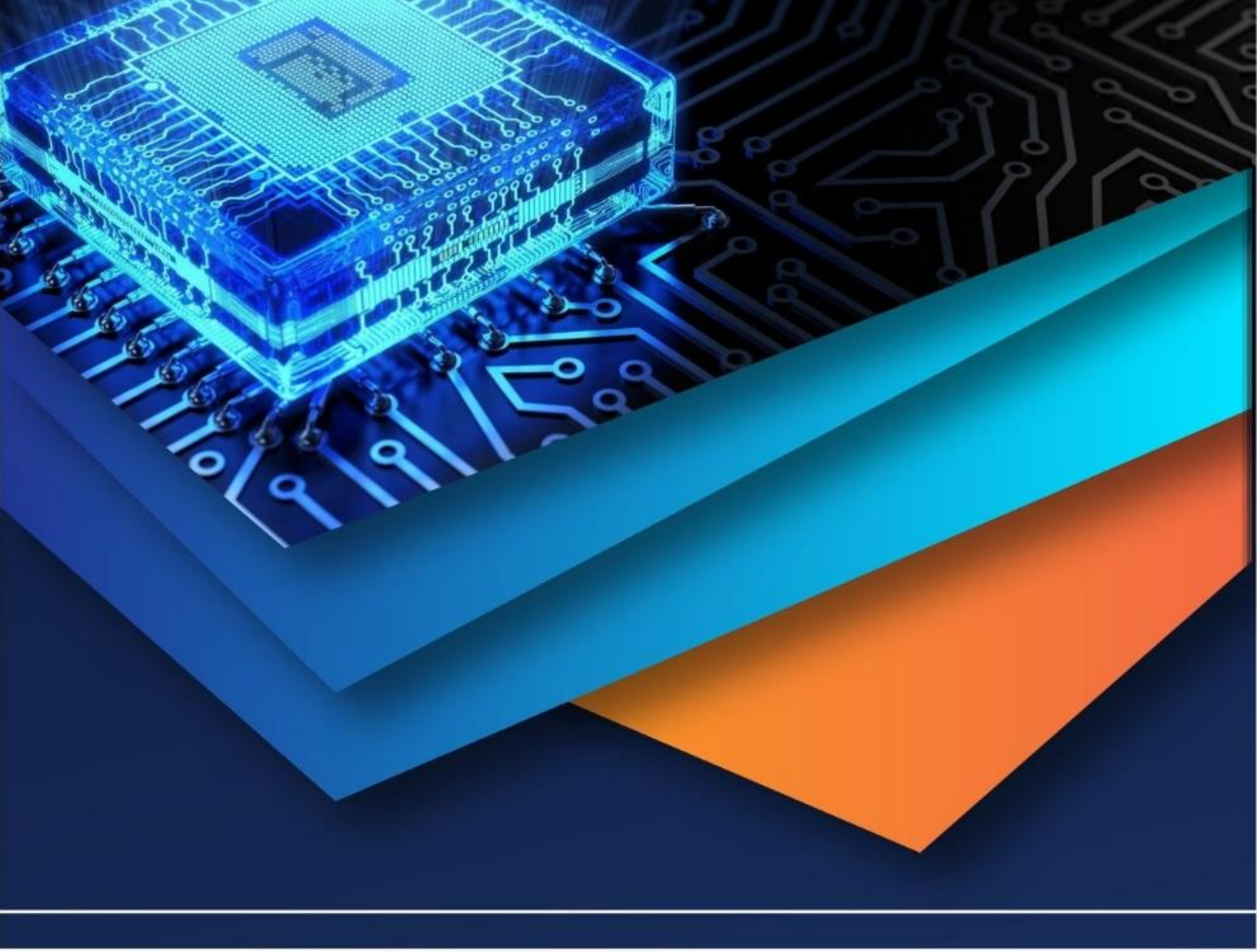
[4] Jun Wu, Li Shi,Wen-Pin Lin,Sang-Bing Tsai,Yuanyuan Li, Liping Yang, and Guangshu Xu “ An Empirical Study on Customer Segmentation by Purchase Behaviors using a RFM Model and K-Means Algorithm “ . https://doi.org/10.1155/2020/8884227, https://www.hindawi.com/journals/mpe/2020/8884227/

[5] Bruce, Larzan and Timothy “ Approaches to Customer Segmentation “. “https://www.researchgate.net/publication/230557972_Approaches_to_Customer_Segmentation”

[6] Dennis, Marsland and Cockett (2001) “ Data Mining for shopping Centres - Customer Knowledge Management Framework “.https://www.emerald.com/insight/content/doi/10.1108/13673270110411797/full/html

[7] E.W.T. Ngai, Li Xiu, D.C.K. Chau “Application of data mining techniques in customer relationship management: A literature review and classification “. https://www.sciencedirect.com/science/article/abs/pii/S0957417408001243

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)