JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CUSTOMER SEGMENTATION FOR MARKETING USING MACHINE LEARNING K-MEANS CLUSTERING ALGORITHM

[1]N. Sri Lakshmi, [1]V. Sai Pramod, [1]Shakeera, [1]A. Praveen, [2]Shaik. Kabeer
[1]U.G Student, [2]P.G Student
[1]Department of Information Technology,[2] Department of MCA
SRK Institute of Technology, Vijayawada, Andhra Pradesh, India.

Abstract: In this present work, the objective is to segment a large group of credit card users according to similarities and usage characteristics. Customer segmentation is then used by marketing department to tailor products and services specific to these customer segments. With the help of machine learning technique one can sort out the data and can find the target group by applying several algorithms to the dataset. Without this, it will be very difficult and no better techniques are available to find the group of people with similar character and interests in a large dataset. Here, the customer segmentation using k-means clustering helps to group the data with same attributes which exactly helps to marketing the best. We are going to use elbow method to find the number of clusters and at last we visualize the data. In this project K-means machine learning algorithm is used to form clusters of different customers. In this competitive market, it is of great significance to divide customer groups to develop customer centered products which would help them to derive successful marketing strategies and also understand what a customer wants. This helps in unravelling the hidden patterns in the data for better decision making for the future and which segment of customers to be targeted.

Keywords: **Clustering, Elbow Method, K-Means algorithm, Customer Segmentation.**

1. INTRODUCTION

Customer Segmentation means grouping the customers based on the marketing groups which shares the similarity among customers. To be more exact, it means segmenting customers sharing the common characteristics which is the best way of marketing. Customer Segmentation is gathering information about each customer and analyzing to identify the different patterns for creating the segments. Information is collected using data which is published related to market categories. The basic information which includes the age, gender, bank balance, purchases, instalments etc., We use k-means, an unsupervised learning algorithm to divide the unlabeled dataset into k different clusters in such a way that customers belong to one cluster have similar properties. The best way of marketing is to understand and analyze what a customer wants. Segmentation of the customers would make the marketing easier by understanding precisely what a customer needs and also in reducing the time and increasing the overall sales. This is the motivation for taking up this project. I. S. Dhillon and D. M. Modha [1] reviewed about the Concept decompositions for large sparse text data using clustering, Machine Learning. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu [2] studied about an efficient K-means clustering algorithm. MacKay and David [3] studied about An Example Inference Task: Clustering.

Jiawei Han, Micheline Kamber, Jian Pei [4] reviewed about the Data Mining Concepts and Techniques. D. Aloise, A. Deshpande, P. Hansen, and P. Popat [5], studied about The Basis of Market Segmentation. S. Dasgupta and Y. Freund [6] studied about the Random Trees for Vector Quantization. V. Vijilesh, A. Harini, M. Hari Dharshini, R. Priyadarshini [7] studied about the Customer segmentation using machine learning. Jaswanth Reddy Vulchi [8] reviewed about customer segmentation using k-means clustering. Yash Kushwaha, Deepak Prajapati [9] studied about the Customer segmentation using K-means algorithm.

2. SYSTEM ANALYSIS:

The present work consists of two phases, in the first phase is the existing system, which stores customer data through paperwork and analyzing manually. But the digital data is increasing day by day. By analyzing the collected customer data manually, they get to know about who is beneficial to their business and increase their sales. The second phase is the process of segmentation is performed on data of credit card users. We collect the data which includes the details about age, gender,

income, credit limit, purchases, frequency of purchases etc., of the credit card user. We then apply k-means algorithm to cluster the credit card users. Later we visualize the data.

2.1 ANALYSIS MODEL:

2.1.1 WATERFALL MODEL – DESIGN:

The waterfall approach was the first SDLC Model to be used widely in Software Engineering to ensure the success of the project. In the "The Waterfall" approach, the whole process of software development is divided into separate phases. In this Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially.

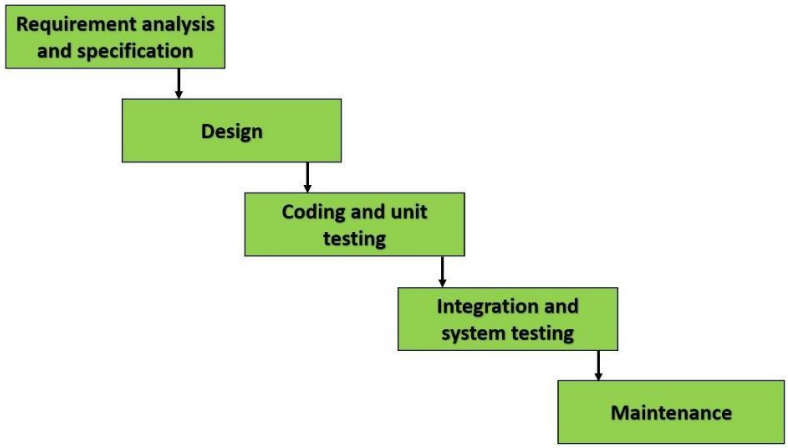

Fig.1: Waterfall Model diagram

2.1.2 MODULES:

The main modules in this project are importing libraries and dataset, visualizing and exploring the data, finding the optimal number of clusters, applying the k means method and visualizing the results.

Firstly, we will import different libraries to perform different functions. We import numpy to perform mathematical operations, pandas, matplot lib for representation and visualization of the results, sea born for visualization of the graphs.

Then we import the dataset. The dataset used in this project is the details about credit card users. The details involves customer ID, balance, balance frequency, purchases, one off purchases, installments purchases, cash advance, purchases, frequency, one off, purchases, frequency, installments purchases, frequency, cash, advance, frequency, cash advance trx, purchases trx, credit limit, payments, minimum payments, prc full payment, tenure.

3. FEASIBILITY STUDY:

3.1 FEASIBILITY STUDY:

Feasibility Study is a high-level capsule version of the entire process intended to answer a number of questions like: What is the problem? Is there any feasible solution to the given problem? Is the problem even worth solving? Feasibility study is conducted once the problem clearly understood. Feasibility study is necessary to determine that the proposed system is Feasible by considering the technical, Operational, and Economical factors. By having a detailed feasibility study the management will have a clear-cut view of the proposed system.

The following feasibilities are considered for the project in order to ensure that the project is variable and it does not have any major obstructions. Feasibility study encompasses the following things.

- Technical Feasibility
- Economic or financial feasibility
- Operational feasibility

In this phase, we study the feasibility of all proposed systems, and pick the best feasible solution for the problem. The feasibility is studied based on three main factors as follows.

3.2 TECHNICAL FEASIBILITY:

In this step, we verify whether the proposed systems are technically feasible or not. i.e., all the technologies required to develop the system are available readily or not.

Technical Feasibility determines whether the organization has the technology and skills necessary to carry out the project and how this should be obtained. The system can be feasible because of the following grounds.

- All necessary technology exists to develop the system.
- This system is flexible and it can be expanded further.
- This system can give guarantee of accuracy, ease of use, and reliability.
- Our project is technically feasible because, all the technology needed for our project is readily available.

3.3 ECONOMIC FEASIBILITY:

In this step, we verify which proposal is more economical. We compare the financial benefits of the new system with the investment. The new system is economically feasible only when the financial benefits are more than the investments and expenditure.

Economic Feasibility determines whether the project goal can be within the resource limits allocated to it or not. It must determine whether it is worthwhile to process with the entire project or whether the benefits obtained from the new system are not worth the costs. Financial benefits must be equal or exceed the costs. In this issue, we should consider:

- The cost to conduct a full system investigation.
- The cost of h/w and s/w for the class of application being considered.
- The development tools.
- The cost of maintenance etc.

Our project is economically feasible because the cost of development is very minimal when compared to financial benefits of the application.

4.4 OPERATIONAL FEASIBILITY:

In this step, we verify different operational factors of the proposed systems like manpower, time etc., whichever solution uses less operational resources, is the best operationally feasible solution. The solution should also be operationally possible to implement. Operational Feasibility determines if the proposed system satisfied user objectives could be fitted into the current system operation.

- The methods of processing and presentation are completely accepted by the clients since they can meet all user requirements.
- The clients have been involved in the planning and development of the system.
- The proposed system will not cause any problem under any circumstances
- Our project is operationally feasible because the time requirements and personnel requirements are satisfied. We are a team of four members and we worked on this project for three working months.

5. UML DIAGRAMS:

Class-based Modeling, or more commonly class-orientation, refers to the style of object-oriented programming in which inheritance is achieved by defining classes of objects; as opposed to the objects themselves. The most popular and developed model of OOP is a class-based model, as opposed to an object-based model. In this model, objects are entities that combine state (i.e., data), behavior and identity. The structure and behavior of an object are defined by a class, which is a definition, or blueprint, of all objects of a specific type. An object must be explicitly created based on a class and an object thus created is considered to

be an instance of that class. An object is similar to a structure, with the addition of method pointers, member access control, and an implicit data member which locates instances of the class (i.e. actual objects of that class) in the class hierarchy.

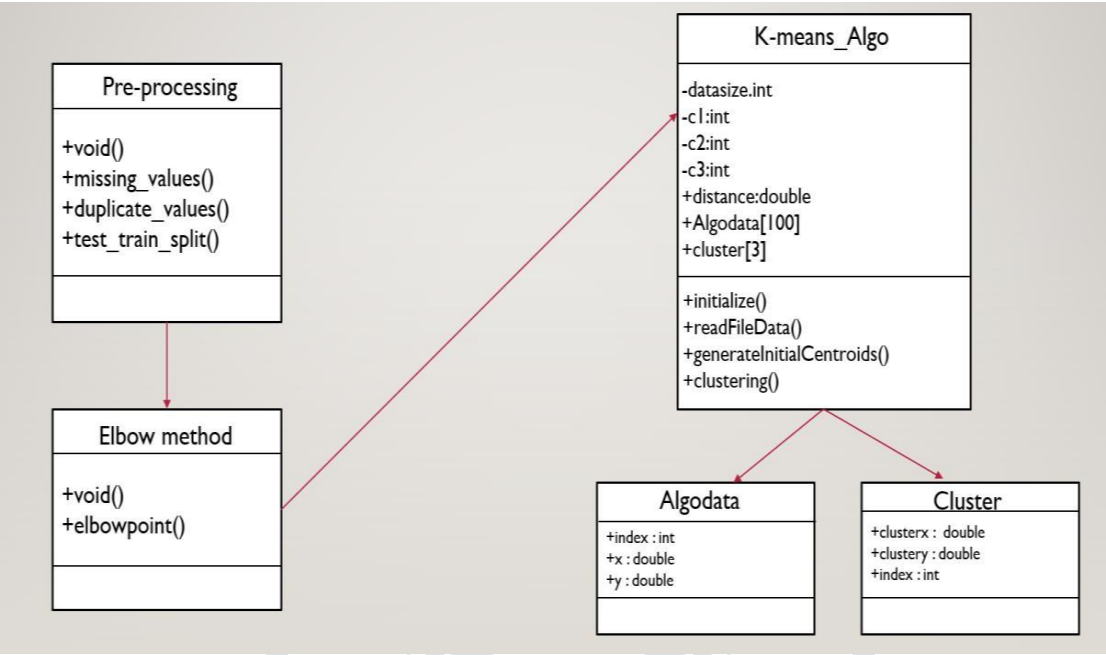

Fig.2: Class Diagram

Class diagram represents the objects, methods and their relationships. Here firstly we perform the preprocessing techniques on the data. Using these methods we remove the missing values, duplicate values and split the data into training data and test data. Next, we find out the elbow point using the following elbow point () method. Then, we will apply the k-means algorithm. We use the above represented variables. Algo data represents the data used in the algorithm and cluster represents the data related to clustering.

Sequence diagram represents the sequence of messages between objects in an interaction here we collect the data first and perform preprocessing as said before. We perform them in a sequence firstly handling missing values, then handling duplicate values and later we split data into test data and training data. The result obtained after preprocessing is sent to perform k-means algorithm. We follow a sequence first we find out cluster centers and perform classification and finally we get the result of customers being segmented into diff clusters.

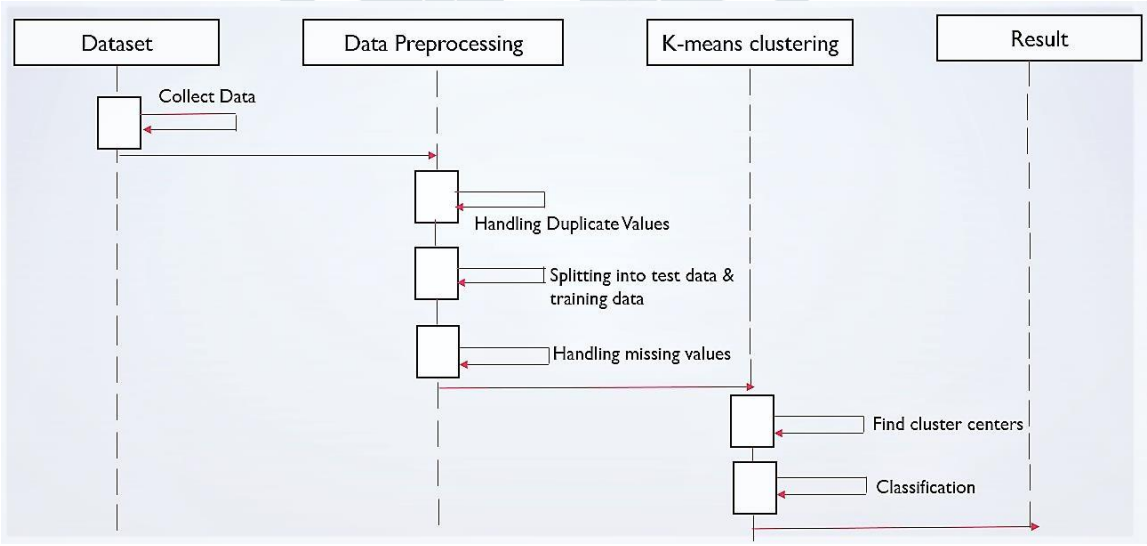

Fig.3: Sequence Diagram

Based on the above study, envolpe a code was developed by us in the Anaconda prompt pass all the commands implemented the customer segmentation for marketing.

6. RESULTS AND DISCUSSIONS

The present work deals with the Customer Segmentation For Marketing The testing is carried out using Anaconda prompt Jupiter Notebook software.

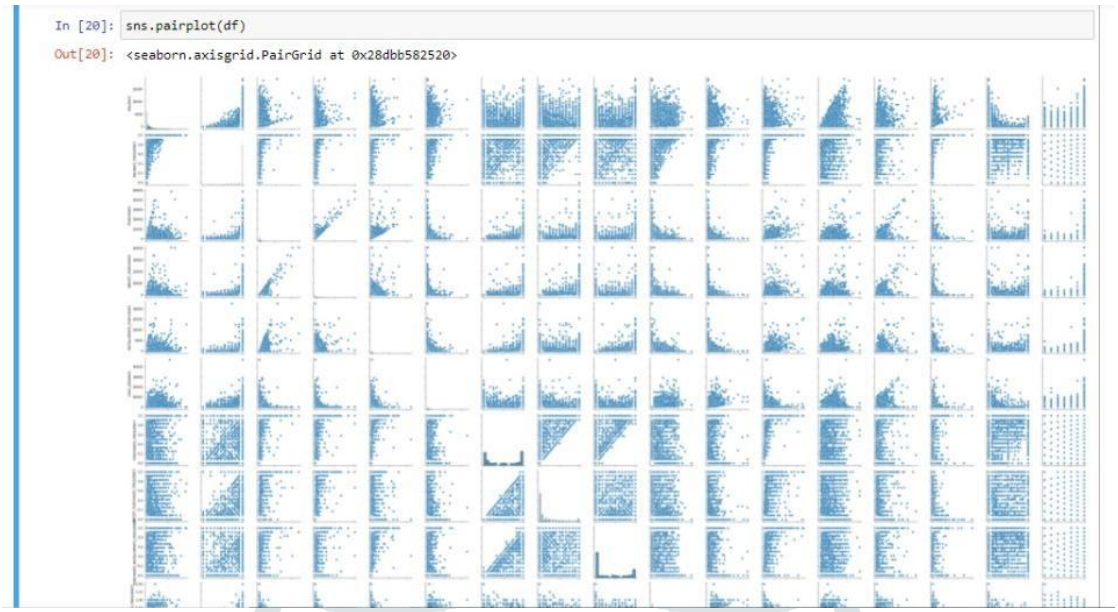

Fig.4: Data analysis

Fig.4 shows the data analysis of the client, which shows the usage of client and the monthly payments of the clients.

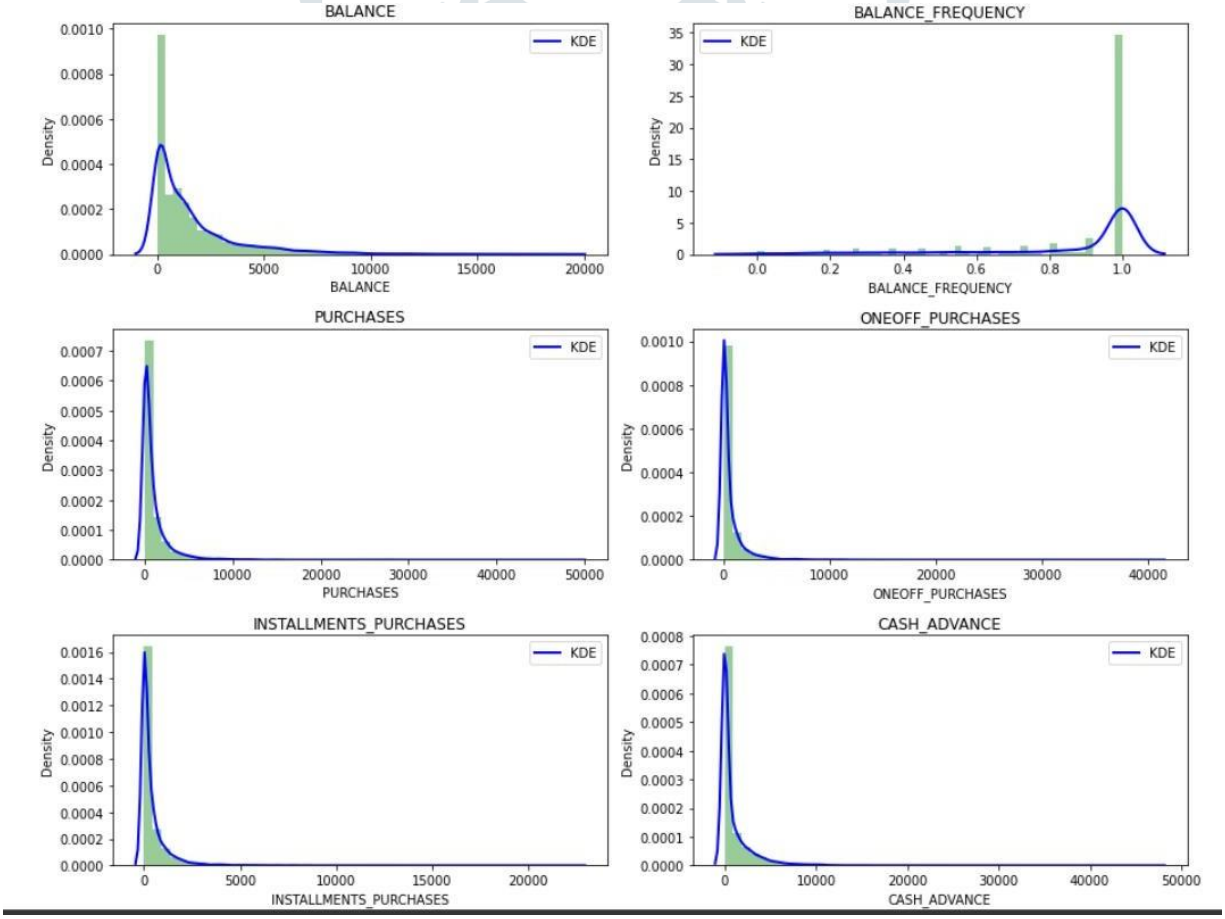

Fig.5: Data visualization

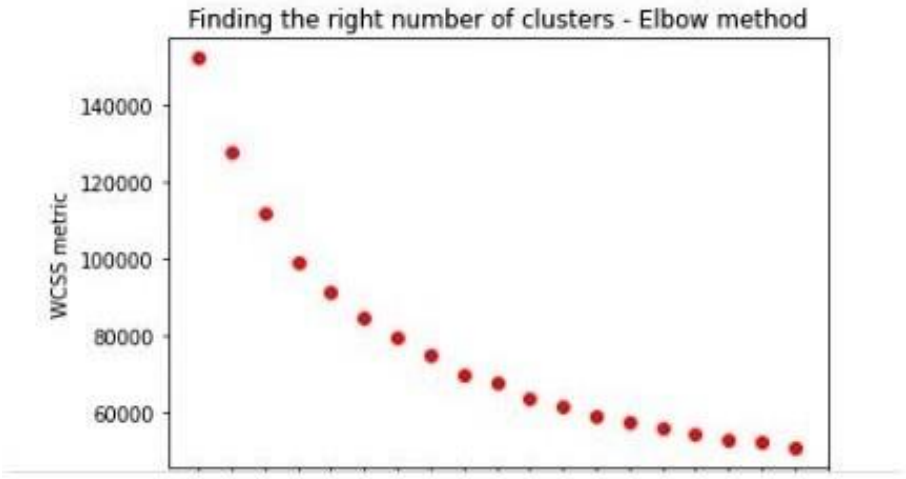

Fig.6: Elbow method

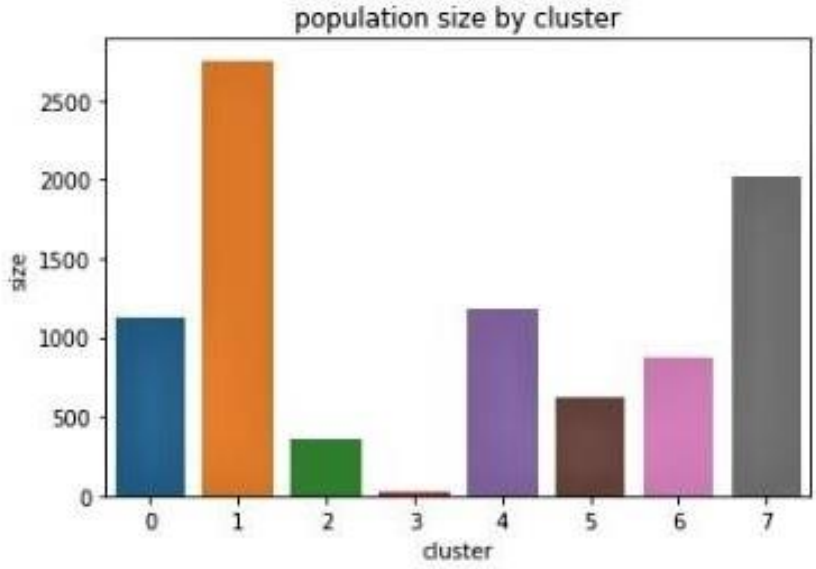

Fig.7: K-means method

7. CONCLUSIONS:

In this study, this paper presented an implementation of the k-means clustering algorithm for customer segmentation using data collected from credit card users. It partitions the customers into mutually exclusive groups or clusters. First customers cluster (revolvers) who use credit cards as a loan (most lucrative sector): highest balance and cash advance, low purchase frequency, high cash advance frequency, high cash advance transactions, and low percentage of full payment. Second customers cluster (Credit purchasers) with high purchase frequency who use payment installment facility the most (highest installment frequency), pay in full whenever possible, and do not use costly cash advance service. Third customers cluster (Active cash buyers): Those are active buyers paying in full. Cluster with highest purchase frequency, second highest purchase transactions and one-off purchases, highest % of payment in full. Fourth customer cluster (VIP/Prime): high credit limit and the highest percentage of full payment, target for increased credit limit and increase spending habits

REFERENCES

1. I. S. Dhillon and D. M. Modha, "Concept decompositions for large sparse text data using clustering," Machine Learning, vol. 42, issue 1, pp. 143-175, 2001.
2. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 881-892, 2002.
3. MacKay and David, "An Example Inference Task: Clustering," Information Theory, Inference and learning algorithms Cambridge University Press, pp. 284-292, 2003.
4. Jiawei Han, Micheline Kamber, Jian Pei "Data Mining Concepts and Techniques", Third Edition.
5. D.Aloise, A. Deshpande, P. Hansen, and P. Popat, "The Basis Of Market Segmentation" Euclidean sum-of-squares clustering," Machine Learning, vol. 75, pp. 245- 249, 2009.
6. S. Dasgupta and Y. Freund, "Random Trees for Vector Quantization,"
7. V.Vijilesh, A.Harini, M.Hari Dharshini, R.Priyadarshini "Customer segmentation using machine learning",2021

8. Jaswanth Reddy Vulchi “Customer segmentation using k-means clustering”
9. Yash Kushwaha, Deepak Prajapati “Customer segmentation using K-means algorithm”