**Customer Segmentation using K-Means Clustering Algorithm**

**Term Project Report**

***Submitted by***

###### SUBHASH KUMAR WARY (202002022110)

###### AKHER UDDIN AHMED (202102022100)

***Under the supervision of***

**Dr. Ranjan Maity**

**(Assistant Professor)**

**B. Tech. 6th Semester**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**



केन्द्रीय प्रौद्योगिकी संस्थान कोकराझार

**CENTRAL INSTITUTE OF TECHNOLOGY KOKRAJHAR**

**(Deemed to be University under MoE, Govt. of India)**

**April 2024**

## **ABSTRACT**

## Customer segmentation is a separation of a market into multiple distinct groups of consumers who share the similar characteristics. Segmentation of market is an effective way to define and meet customer needs. Unsupervised Machine Learning Techniques K-Means Clustering Algorithm is used to perform Market Basket Analysis. Market Basket Analysis is carried out to predict the target customers who can be easily converged, among all the customers. In order to allow the marketing team to plan the strategy to market the new products to the target customers which are similar to their interests.

## Key words: Target Customers, Clusters, Unsupervised Learning, K-Means, Market Basket Analysis

# **CONTENTS**

1 [Introduction](#_bookmark8)………………………………………………………………………… 1

1. [Literature Review](#_bookmark11)…………………………………………………………………... 2
   1. Customer Segmentation 2

2.2 Clustering and K-Means Algorithm 2-3

1. Methodology………………………………………………………………………… 4
   1. Clustering 4
   2. K Means Clustering Algorithm 4

3.3 Elbow Method 4-5

1. [Results](#_bookmark43) and Conclusion……………………………………………………………...6
2. [Future Scope](#_bookmark64)………………………………………………………………………… 7
3. References…………………………………………………………………………… 8

**1. Introduction**

Management and maintain of customer relationship have always played a vital role to provide business intelligence to organizations to build, manage and develop valuable long term customer relationships. The importance of treating customers as an organizations main asset is increasing in value in present day and era. Organizations have an interest to invest in the development of customer acquisition, maintenance and development strategies. The business intelligence has a vital role to play in allowing companies to use technical expertise to gain better customer knowledge and Programs for outreach. By using clustering techniques like k-means, customers with similar means are clustered together. Customer segmentation helps the marketing team to recognize and expose different customer segments that think differently and follow different purchasing strategies. Customer segmentation helps in figuring out the customers who vary in terms of preferences, expectations, desires and attributes. The main purpose of performing customer segmentation is to group people, who have similar interest so that the marketing team can converge in an effective marketing plan. Clustering is an iterative process of knowledge discovery from vast amounts of raw and unorganized data. Clustering is a type of exploratory data mining that is used in many applications, such as machine learning, classification and pattern recognition.

1

**2. Literature Review**

**2.1. Customer Segmentation**

Over the years, as there is very strong competition in the business world, the organizations have to enhance their profits and business by satisfying the demands of their customers and attract new customers according to their needs. The identification of customers and satisfying the demands of each customer is a very complex task. This is because customers may be different according to their demands, desires, preferences and so on. Instead of “one-size-fits-all” approach, customer segmentation clusters the customers into groups sharing the same properties or behavioural characteristics. According to, [1] customer segmentation is a strategy of dividing the market into homogenous groups. The data used in customer segmentation technique that divides the customers into groups depends on various factors like, demographical conditions, data geographical conditions and economic conditions as well as behavioural patterns. The customer segmentation technique allows the business to make better use of their marketing budgets, gain a competitive edge over their rival companies, demonstrating the better knowledge of the needs of the customer. It also helps an organization in, increasing their marketing efficiency, plan the marketing budget, determining new market opportunities, making better brand strategy, identifying customers retention.

According to [1], Decision makers use many variables to segment customers. Demographic variables such as age, gender, family, education level and income are the easiest and common variables for segmentation. Socio- cultural, geographic, psychographic and behavioural variables are the other major variables that are used for segmentation.

[2], presented various clustering algorithms taking into account the characteristics of Big Data such as size, noise, dimensionality, algorithm calculations, cluster shape and presented a brief overview of the various clustering algorithms grouped under partitioning, hierarchical, density, grid-based and model-based algorithms.

[4] explored the necessity of segmentation of the customers using clustering algorithms as the core functionality of CRM. The mostly used K-Means and Hierarchical Clustering were studied and the advantages and disadvantages of these techniques were highlighted. At last, the idea of creating a hybrid approach is addressed by integrating the above two strategies with the potential to surpass the individual designs.

[5], Merged clustering of fuzzy c-means and genetic algorithms to cluster, steel industry customers, by using the LRFM variables (length, recency, frequency, monetary value) system, customers were divided into two clusters

**2.2 Clustering and K-Means Algorithm**

Clustering algorithms generates clusters such that within the clusters are similar based on some characteristics. Similarity is defined in terms of how close the objects are in space.

According to [1], K-means algorithm in one of the most popular centroid based algorithms. Suppose data set, D, contains n objects in space. Partitioning methods distribute the objects in

2

D into k clusters, C1,……Ck , that is, Ci ⊂ D and Ci ∩Cj = ∅ for (1 ≤ i, j ≤ k). A centroid-based partitioning technique uses the centroid of a cluster, Ci, to represent that cluster. Conceptually, the centroid of a cluster is its centre point. The difference between an object p ∈ Ci and ci, the representative of the cluster, is measured by dist(p,ci), where dist(x,y) is the Euclidean distance between two points x and y.

Algorithm: The k-means algorithm for partitioning, where each cluster’s centre is represented by the mean value of the objects in the cluster. Input: k: the number of clusters, D: a data set containing n objects. Output: A set of k clusters. Method: (1) arbitrarily choose k objects from D as the initial cluster centres; (2) repeat (3) (re)assigns each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster; (4) update the cluster means, that is, calculate the mean value of the objects for each cluster; (5) until no change.

3

**3. Methodology**

**3.1 Clustering**

Clustering is one of the most common methods used in exploring data to obtain a clear understanding of the data structure. It can be characterized as the task of finding the subtitles and subgroups in the complete dataset. Similar data is clustered in many subgroups. A cluster refers to a collection of aggregated data points due to some similarities. Clustering is used in Market basket analysis used to segment the customers based on their behaviours and transactions.

**3.2 K Means Clustering Algorithm**

K Means Clustering is the most common and simplest Machine learning algorithm and it follows an iterative approach which attempts to partition the dataset into different “k” number of predefined and non-overlapping subgroups where each data point belongs to only one subgroup according to their similar qualities.

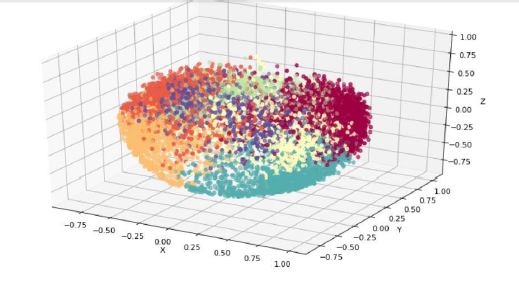


Fig.: K-Means Clustering Algorithm

**3.3 Elbow Method**

Elbow method is a tool used for analysing the clusters formed from our dataset and helps to interpret the appropriate number of optimal clusters in dataset. From this method the optimal number of clusters for our dataset is found to be seven.

4

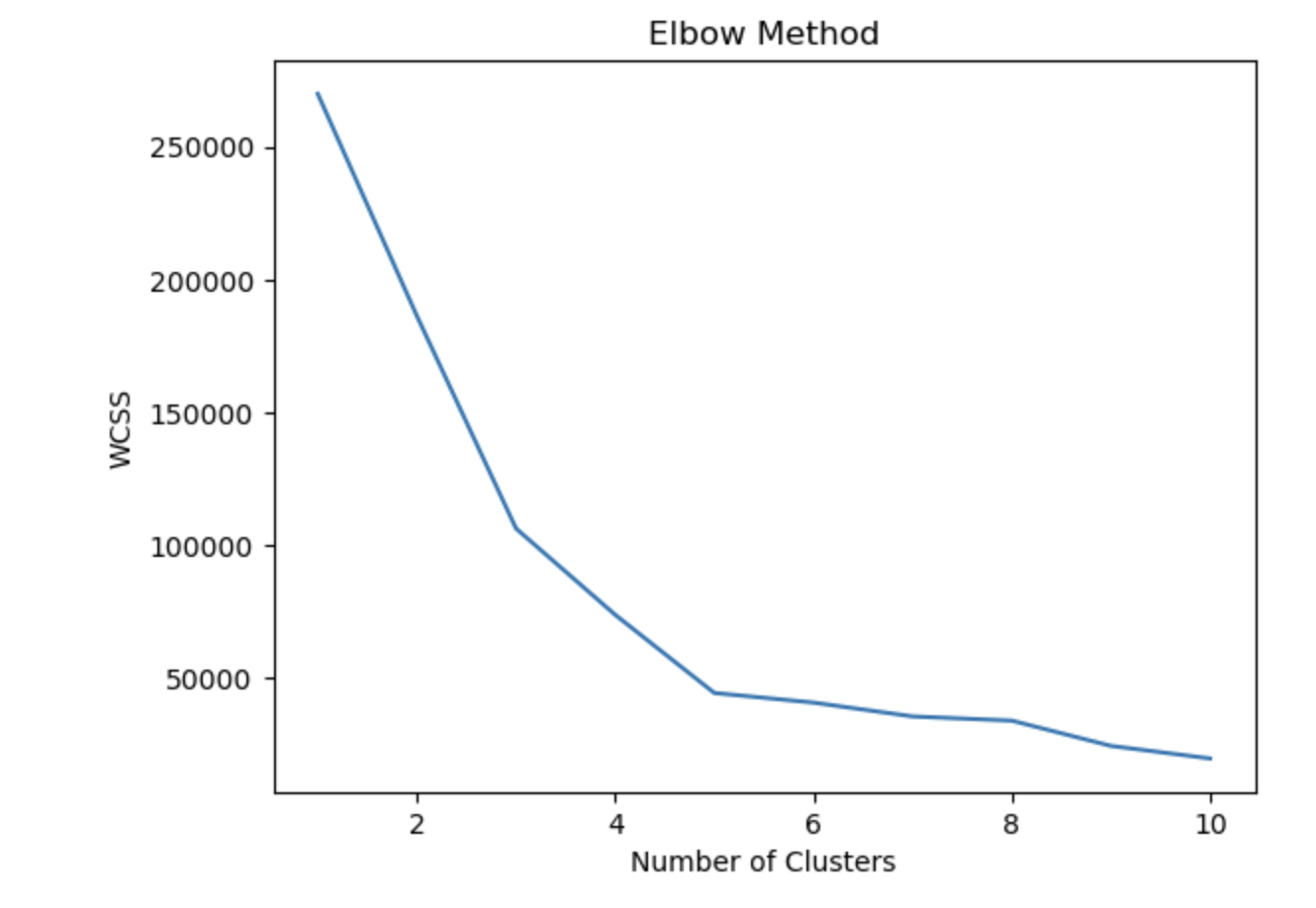


Fig.: Elbow Method

5

**4. Results and Conclusion**

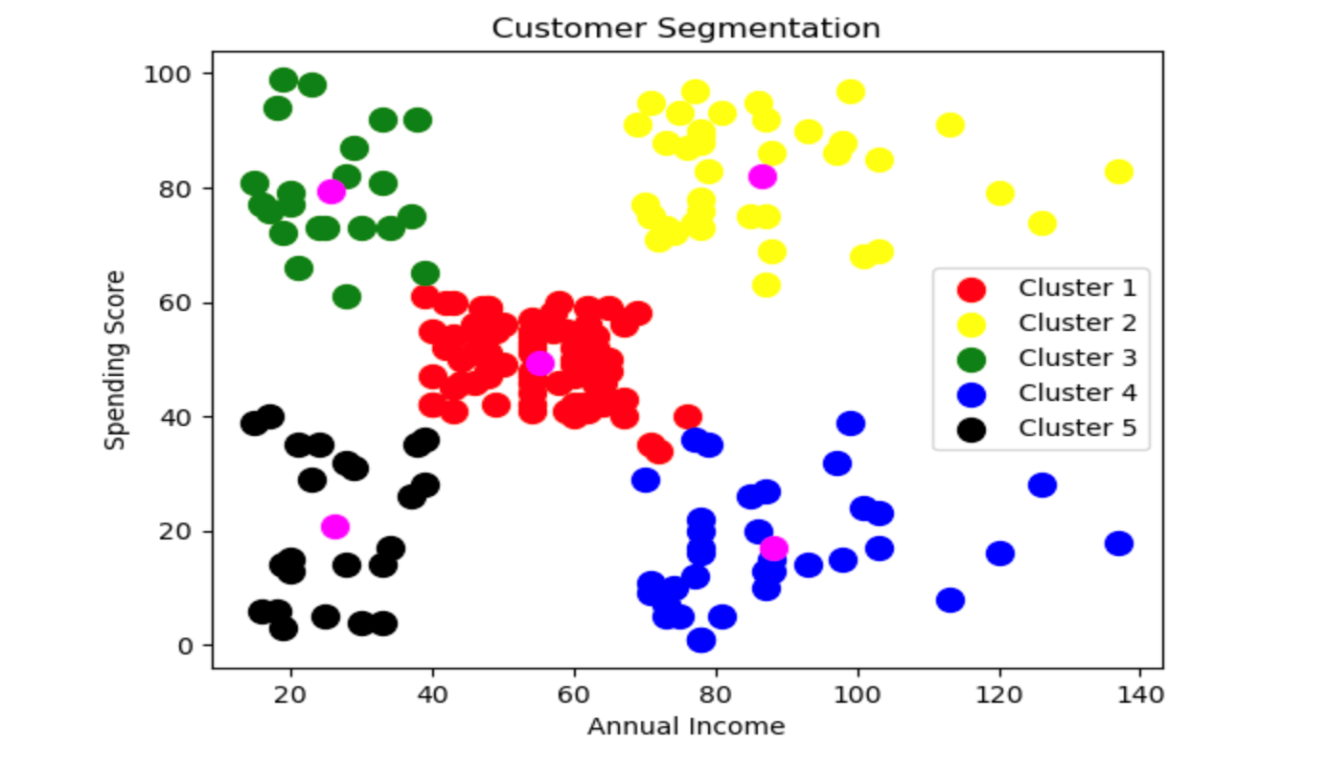
****

Fig.: Cluster of Customer

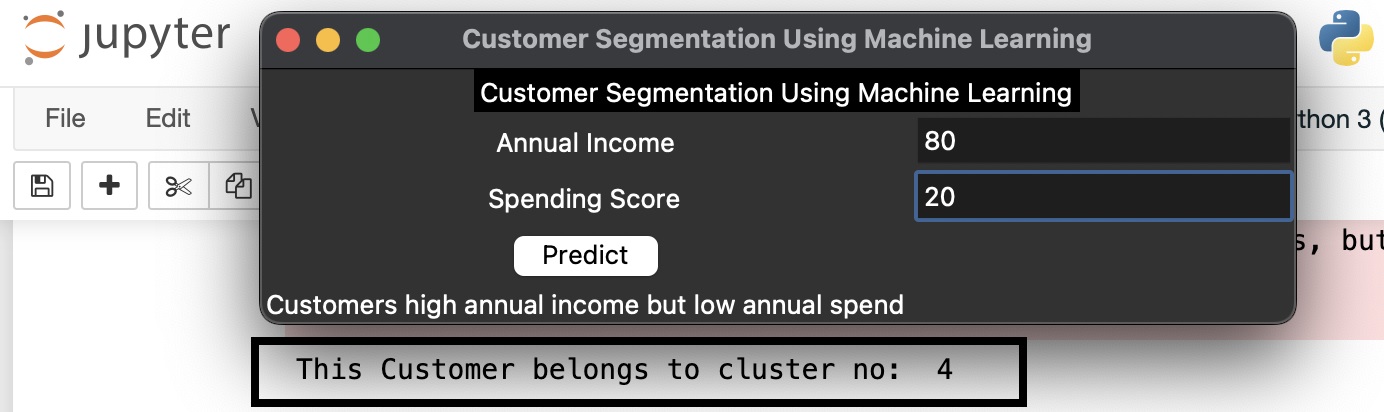


Fig.: Output/Result

Cluster 1: Customers with medium annual income and medium annual spend.

Cluster 2: Customers with high annual income and high annual spend.

Cluster 3: Customers with low annual income and high annual spend.

Cluster 4: Customers high annual income but low annual spend.

Cluster 5: Customers with low annual income and low annual spend.

Firstly, we have started with data pre-processing. Then, we applied K-Means clustering algorithms. Then, we divided the data into five clusters because five clusters can be easily used to determine the behaviours of customers. However, each of the clusters have their own characteristics.

6

**5. Future Scope**

* We can use different distance metrics: K-means clustering uses Euclidean distance as the default distance metric. However, other distance metrics such as Manhattan, Cosine and Minkowski can be used to increase the accuracy of the clustering.
* We can use different initialization methods: K-means clustering uses random initialization by default. However, other initialization methods such as K-means and k-medoids can be used to increase the accuracy of the clustering.
* We can use different cluster sizes: K-means clustering uses a fixed number of clusters.

7

**6. References**

[1]

D. P. Yash Kushwaha, “Customer Segmentation using K-Means Algorithm,” 8th Semester Student of B.tech in Computer Science and Engineering.

[2]

E. A. Onur DOĞAN1, “CUSTOMER SEGMENTATION BY USING RFM”.

[3]

C. M. S. R. a. K. V. N. T. Sajana, “A Survey on,” in Indian Journal of Science and Technology, Volume 9, Issue 3,, Jan 2016..

[4]

A. B. P. E. Shreya Tripathi, “Approaches to,” in International Journal of Engineering and Technology, Volume 7, 2018.

[5]

A. R. Azarnoush Ansari, ““Customer Clustering Using a,” in International Journal of Business and Management, Volume 11, Issue 7,, 2016.

[6]

“mini-batch-k-means-clustering-algorithm,” [Online]. Available: https://www.geeksforgeeks.org/ml-mini-batch-k-means-clustering-algorithm/.

[7]

“/hierarchical-clustering-customer-segmentation,” [Online]. Available: https://www.coursera.org/projects/hierarchical-clustering-customer-segmentation.

8