VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM

"Jnana Sangama", Belgaum-590018

A Report On

"Customer Segmentation using Machine Learning"

Submitted by

NANDINI MUNDRA 4NI16IS055

SUSHANT SRIVASTAVA 4NI16IS109

TANMOY DEBNATH 4NI16IS111

VINAY PANDHARIWAL 4NI16IS116

Under the Guidance of

Mr. SUHAAS K.P., Assistant Professor

Dept of IS&E

NIE, Mysuru

The National Institute of Engineering

(Autonomous Institution)

MYSURU-570008

Department of Information Science and Engineering

MYSURU-570008

2019-2020

THE NATIONAL INSTITUTE OF ENGINEERING
MYSURU-570008

Department of Information Science and Engineering

Certificate

This is to certify that the Minor project work entitled "Customer Segmentation using Machine Learning" is a work carried out by **NANDINI MUNDRA, SUSHANT SRIVASTAVA, TANMOY DEBNATH AND VINAY PANDHARIWAL** in partial fulfilment for the Minor project prescribed by National Institute of Engineering, Autonomous Institution under Vishvesvaraya Technological University, Belagavi for the seventh semester B.E Information Science & Engineering. It is certified that all correction/suggestions indicated for Internal Assessment have been incorporated. The Minor project report has been approved as it satisfies the academic requirements in respect of the Minor project work prescribed for the Seventh Semester.

Signature of Guide	Signature of Prof. and Head, Department of IS&E	Signature of the Principal
(Mr. Suhaas K. P.)	**(Dr. P Devaki)**	**(Dr. G. Ravi)**

Name of the Examiner	Signature with Date
1.	1.
2.	2.

ACKNOWLEDGEMENT

We would like to take this opportunity to express our profound gratitude to all those people who are directly or indirectly involved in the completion of the project. We thank each and every one who encouraged us in every possible way.

We would like to thank **Dr. G. RAVI,** principal, NIE, Mysuru for letting us to be a part of this prestigious institution and letting us to explore our abilities to the fullest.

We would like to extend our sincere gratitude to **Dr. P DEVAKI,** Professor and Head, Department of IS&E, NIE, Mysuru for being a source of inspiration and instilling an enthusiastic spirit in us throughout the process of project making.

We wish to express our heartfelt gratitude towards **Mr. Suhaas K.P,** Minor project Guide, Assistant Professor, Department of IS&E, NIE, Mysuru for his consistent guidance and valuable knowledge.

We are extremely pleased to thank our family members and friends for their continuous support, inspiration and encouragement, for their helping hand and also last but not the least We are grateful to all the members who supported us directly or indirectly in our academic process.

NANDINI MUNDRA: 4NI16IS055

SUSHANT SRIVASTAVA: 4NI16IS109

TANMOY DEBNATH: 4NI16IS111

VINAY PANDHARIWAL: 4NI16IS116

ABSTRACT

Customer Relationship Management (CRM) has always played a crucial role as a market strategy for providing organizations with the quintessential business intelligence for building, managing and developing valuable long-term customer relationships. A number of business enterprises have come to realize the significance of CRM and the application of technical expertise to achieve competitive advantage. The zeitgeist of modern era is innovation, where everyone is embroiled into competition to be better than others. Today's business run on the basis of such innovation having ability to enthrall the customers with the products, but with such a large raft of products leave the customers confounded, what to buy and what to not and also the companies are nonplussed about what section of customers to target to sell their products. This is where machine learning comes into play, various algorithms are applied for unravelling the hidden patterns in the data for better decision making for the future. This elude concept of which segment to target is made unequivocal by applying segmentation. The process of segmenting the customers with similar behaviors into the same segment and with different patterns into different segments is called customer segmentation. This project explores the importance of Customer Segmentation as a core function of CRM as well as the various models for segmenting customers using clustering techniques. The available clustering models for customer segmentation, in general, and the major models of K-Means and Hierarchical Clustering, in particular, are studied and the virtues and vices of the techniques are pointed out. Finally, the possibility of developing a hybrid solution by the combination of the above two techniques, having the ability to outperform the individual models, is discussed. A python program has been developed and the program is been trained by applying standard scaler onto a dataset.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter: 1

INTRODUCTION

In the contemporary day and age, the importance of treating customers as the principal asset of an organization is increasing in value. Organizations are rapidly investing in developing strategies for better customer acquisition, maintenance and development. The concept of business intelligence has a crucial role to play in making it possible for organizations to use technical expertise for acquiring better customer insight for outreach programs. In this scenario, the concept of CRM garners much attention since it is a comprehensive process of acquiring and retaining customers, using business intelligence, to maximize the customer value for a business enterprise.

One of the two most important objectives of CRM is customer development through customer insight. This objective of CRM entails the usage of an analytical approach in order to correctly assess customer information and analysis of the value of customers for better customer insight. Keeping up with the changing times, organizations are modifying their business flow models by employing systems engineering as well as change management and designing information technology (IT) solutions that aid them in acquiring new customers.

Due to the diverse range of products and services available in the market as well as the intense competition among organizations, customer relationship management has come to play a significant role in the identification and analysis of a company’s best customers and the adoption of best marketing strategies to achieve and sustain competitive advantage. One of the most useful techniques in business analytics for the analysis of consumer behavior and categorization is customer segmentation. Customer Segmentation helps organizations in identifying or revealing distinct groups of customers who think and function differently and follow varied approaches in their spending and purchasing habits. Customers vary in terms of behavior, needs, wants and characteristics and the main goal of clustering techniques is to identify different customer types and segment the customer base into clusters of similar profiles so that the process of target marketing can be executed more efficiently. This study aims to explore the avenues of using customer segmentation, as a business intelligence tool within the CRM framework as well as the use of clustering techniques for helping organizations redeem a clearer picture of the valuable customer base. The concepts of customer relationship management, customer segmentation as a core function of CRM as well as the approach of segmenting customers using clustering techniques are discussed.

Chapter: 2

CUSTOMER RELATIONSHIP MANGEMENT, CUSTOMER SEGMENTATION AND USE OF CUSTOMER SEGMENTATION.

2.1 Customer Relationship Management

Customer Relationship Management is an important business approach for developing and securing steady, long-term customer associations. The modern marketing approach promotes the usage of CRM as part of the organization's business strategy for enhancing customer service satisfaction. CRM enables business enterprises in customer value analysis as well as the targeting of those customers that prove of greater value. It also helps business organizations in developing high-quality and long-term customer company relationships that increase loyalty and profits. An accurate evaluation of customer profitability and the targeting of high value customers are important factors that contribute to the success of CRM. CRM being a customer-centric strategy, it is important for business organizations to be familiar with their customer base in terms of its characteristics and behavior. These insights into the customer data can then become useful when it is employed in such Information Technology (IT) solutions that provide valuable outputs for better targeting the profitable customers. CRM plays a major role in targeting the customer base, once it is identified using essential segmentation strategies. The CRM strategy is a closed circular structure with four dimensions. Customer identification, customer attraction, customer retention, and customer development. Thus, customer identification which lays the foundation of this structure clearly implies that the act of grouping or segmenting customers according to their behavior and characteristics-customer segmentation, emerges as a core function of CRM.

2.2 Customer Segmentation

As the market is widening, the rate of competition between all business entities is rapidly growing. Hence, these business enterprises are increasing their expenditure on their marketing strategies to achieve competitive advantage. In this context, the significance of employing Information Technology (IT) solutions to marketing campaigns emerges as a pivotal step in a modern approach to business. Customer Segmentation is a popular technique of partitioning the customer base into externally distinct and internally uniform groups in order to create varied marketing strategies for targeting each group according to its characteristics. Generally speaking, it is defined as the process whereby the consumers of a business enterprise are divided into groups according to their preferences, characteristics and purchasing behavior.

By studying and analyzing large volumes of collected customer data, businesses can improve their marketing decisions based on the customers" preferences. Maximum profits can be generated for any business entity if the resources are utilized judiciously in order to cultivate the most loyal and useful group of customers once customer segmentation and clustering have enabled the allocation of customers to such groups. The total customer set can be divided and grouped into clusters based on their buying behavior, frequency, demographics etc. Hence, instead of studying each customer individually, firms can group similar customers together so that their needs can be better understood.

2.3 Use of Customer Segmentation

Target Marketing and Customer Segmentation are so closely related that they are often used interchangeably. Target marketing refers to the grouping of buyers based on certain characteristics which the firms intend to serve. It has been referred to as a personal branding strategy in the context of a specific customer. There are three steps that ought to be followed in order to devise segmentation-based marketing strategies. First, customers in the selected market are segmented into different groups based on their characteristics. Secondly, the segments are studied for their properties and the different ways in which marketing tactics can be applied to that specific group. At last, required comparisons on the competing brands and studies about customer behavior to their products can be completed. Hence, a segmentation model which is useful will be able to effectively increase the profitability and competitive value for a company. The next section delves into the concept of using clustering for customer segmentation and the various algorithms involved.

Chapter: 3

THE FAMOUS DATA MINING MYTH

Why men don’t buy beer and diapers at the same time, and what we can still learn from urban legends.

Wal-Mart, the world’s largest retailer, supposedly found out that there are certain times at which beer and diapers sell particularly well together – when on Friday evenings young men make a last dash to the supermarket to get beer and their wives call after them, “Pick up some diapers, too.”

“Some of the ways Wal-Mart managers found to exploit their findings are legendary. One such legend is the story, “diapers and beer”. Wal-Mart discovered through data mining that the sales of diapers and beer were correlated on Friday nights. It determined that the correlation was based on working men who had been asked to pick up diapers on their way home from work. On Fridays the men figured they deserved a six-pack of beer for their trouble; hence the connection between beer and diapers. By moving these two items closer together, Wal-Mart reportedly saw the sales of both items increase geometrically.”

A version with a slightly different view of the roles involved suggests that the men are sent to the supermarket for the diapers and, because there’s no time left to go to a bar, take beer home with them.

In all versions of the story, Wal-Mart then puts the diapers closer to the beer and makes a fortune.

What the diapers-and-beer example, should tell us is this: There are algorithms which we can use for automated recognition of data associations.

Chapter: 4

USING CLUSTERING TECHNIQUES FOR CUSTOMER SEGMENTATION

Clustering, which is also called as cluster analysis, is an important subject in data mining. The data set is partitioned into some groups and the data points in each group, that is, the cluster are more related to each other than to those in other clusters. These data points are clubbed together by detecting correspondences according to the attributes found in raw data, however the main purpose is to find the suitable number of clusters which are relevant as well as insightful for analysis purposes. This process is a repeatable and iterative task where vast amounts of raw data are scanned for similarities and patterns. The unorganized data is searched for knowledge that is important and then data points are assigned. For getting favorable results, a specific clustering algorithm along with certain parameters may be best suited in a market domain.

Clustering is a type of data mining technique used in a number of applications, involving areas such as machine learning, classification and pattern recognition. There are various clustering algorithms which differ from each other in terms of the approach they follow in order to do the grouping of the objects according to their characteristics. These are stated as follows – In Partitioned Based Clustering all the data points are taken as a single cluster in the beginning. These data points are then separated into clusters by iteratively positioning these objects between the clusters. Some of the partitioning algorithms are KMeans, K-Medoids and K-Modes. For Hierarchical Clustering, one of the two present approaches can be followed for implementation. One is Agglomerative (bottom-up) approach and another one is Divisive (top-down) approach.

In the agglomerative process, each observation begins in its own cluster or segment and then pairs of these formed segments are combined when moving up the hierarchy. In the divisive approach, all the observations begin in one cluster and then are repeatedly split into different clusters. The results are in the form of dendrograms. In Density Based Clustering, the clusters are defined as regions of higher density than the rest of the dataset. Objects are differentiated as core, noise and border points. Grid Based Clustering algorithms partition the data set into grid structures containing a number of cells. Grid algorithm uses subspace and hierarchical clustering techniques. STING, CLIQUE are some of the grid-based clustering techniques.

4.1 K-Means Clustering

K-Means is one of the most widely used clustering algorithms, and is simple and efficient. The aim of K-Means algorithm is to divide M points in N dimensions into K clusters (assume k centroids) fixed a priori. These centroids should be placed in a wise fashion so that the results are optimal which otherwise can differ if locations of the centroids change. So, they should be placed as far as possible from each other. Each data point is then taken and associated with the nearest centroid until no data points are pending. This way an early grouping is done and at this point, k new centroids have to be recalculated as these will be the centers of the clusters formed earlier. After having calculated these centroids, the data points are then allocated to the clusters to the nearest centroids. In this iteration, the centroids change their position stepwise until no further modifications have to be done and the location of the centroids remain intact.

The K-Means algorithm is relatively simple. The „K" cluster points, which will be the centroids, are placed in the space among the data points. Each data point is assigned to the centroid for which the distance is the least. After each data object has been assigned, centroids of the new groups are re-calculated. The above two steps are repeated until the movement of the centroid ceases. This means that the objective function of having the least squared error is completed and it cannot be improved further. Hence, we get K clusters as a result. K-Means algorithm aims at minimizing an objective function, which here, is the squared-error. It is an indicator of the distance of the data points from their respective cluster centers. The process in this algorithm always terminates but the relevance or the optimal configuration cannot be guaranteed even when the condition on the objective function is met. The algorithm is also sensitive to the selection of the initial random cluster centers. That is why it runs multiple times to reduce this effect but for a large number of data points, it tends to perform very well even though it is iterative.

A variety of cluster validity indices are used, major of them are Dunn, Davies Bouldin, Silhouette, Sum of Squares within Cluster (SSWC), C, Calinski-Harabasz. SSWC is simple and the most widely used criterion to gauge the validity of the clusters. Smaller values of SSWC mean better clusters. Here we apply K-Means Clustering algorithm on a relatively small dataset and the results are depicted. The dataset is based on customer information for a mall and has 5 attributes named Customer, Genre, Age, Annual Income and Spending Score. It consists of 200 observations, each of which refers to a unique customer and the spending scores are decided and calculated by the company, based on their spending habits. Hence annual income and spending scores are the key indicators in this data. The age attribute of the customers can also be experimented with, to analyze which age group works best for a business. Any business would always keep the monetary values of any customer as top indicators. Thus, the annual income and spending scores of the customers will be best suited for clustering. As K-Means algorithm requires the number of clusters as input, below we will use the elbow method to get the optimal number of clusters which can be formed. It works on the principal that after a certain number of „K" clusters, the difference in SSE (Sum of Squared Errors) starts to decrease and diminishes gradually. Here, the WCSS(Within-Cluster-Sum-of-Squared-errors) metric is used as an indicator of the same. Hence, the „K" value, specifies the number of clusters. In Figure 1, it can be observed that an elbow point occurs at K=5. After K=5, the difference in WCSS is not so visible. Hence, we will choose to have 5 clusters and provide the same as input to the K-Means algorithm.

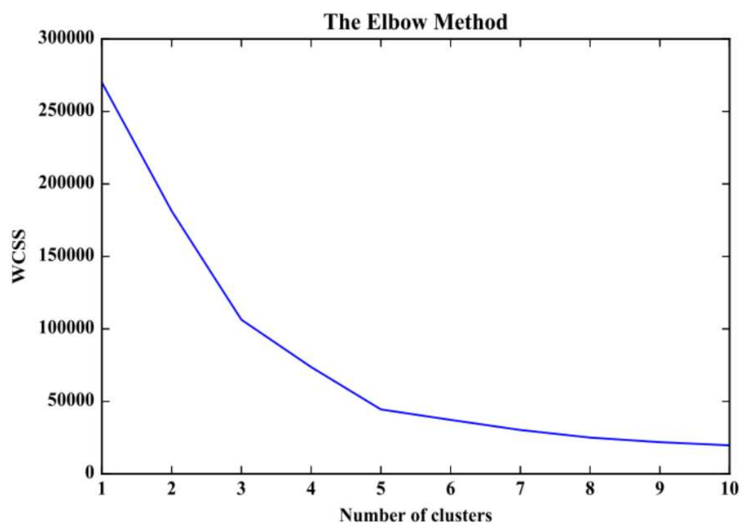


Fig 4.1: Finding the optimal number of clusters.

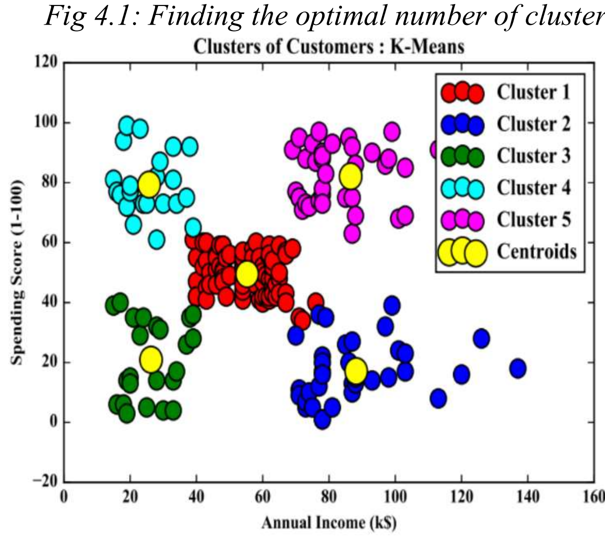


Fig 4.2: Clusters formed using K-means clustering.

As shown in Figure 2, the scatter plot of the clusters is created with Annual income plotted against X-axis and Spending Score against Y-axis. The data points under each cluster are represented using distinct colors and the centroids are also highlighted, as shown above. K-means has been used in all the major domains; have used it in their research about finding clusters in retail industry, have taken a dataset of the transactions in a supermarket in their comprehensive analysis of data clustering algorithms. Have also applied K-means algorithm when doing customer segmentation. Chosen K-means as the partitioning algorithm for the customer data. also chose a form of the K-means algorithm as a part of their clustering technique. Hence K-means is widely accepted and largely used in customer segmentation. One of the major problems in cluster analysis is the selection of the number of clusters „K". Hence, this algorithm does not give the optimal number of clusters but instead this has to be given as input while performing clustering.

The ideal number of clusters are the ones in which variation is minimum among the clusters and maximum between the clusters. One major advantage of K-Means clustering is that the computational speed of this algorithm is higher than other hierarchical methods of clustering and it is also easy to implement. 0When the differences between the categories is small and the data is scaled, this algorithm tends to perform better but one major downside is that it depends upon the initial cluster centers. So, if these are chosen incorrectly, then the results tend to be inaccurate and unstable. The time complexity of K-Means is given by O(nkl), where n represents the number of data points, k gives the number of clusters and l is the number of iterations undergone before the algorithm converges. The space complexity is given by O(k+n).

4.2 Hierarchical Clustering

Hierarchical clustering is a method of cluster analysis which builds a hierarchy of data points as they move into a cluster or out of it. Strategies for this algorithm generally fall into two categories.

1. Agglomerative - This is a bottom-up approach where each observation begins as an initial cluster and then merges into clusters as they move up the hierarchy. Divisive technique is a top-down approach where there is only one cluster initially and is then split into finer cluster groups as they move down the hierarchy. This merging and splitting of clusters take place in a greedy manner and the hierarchical algorithm yields a dendrogram which represents the nested grouping of patterns and the levels at which groupings change.

2. Divisive - These are quite rarely used in market research and hence the agglomerative approach is the one that is widely followed by the practitioners. Here in each step, the two closest clusters are merged based on a specific linkage criterion. The linkage criterion defines the distance between the two clusters. The grouping of data points depends very much on the choice of the linkage criterion. Some of them are complete, single, ward or average. The time complexity of linkage metrics-based hierarchical clustering is high and it is generally given by $O(n^3)$. It is sometimes reduced to $O(n^2 logn)$ or $O(n^2)$ for some algorithms like CLINK and SLINK. The algorithm for the agglomerative hierarchical clustering approach proceeds by taking each observation in a cluster of its own. A pair of clusters with the shortest distance between them is chosen.

The above two clusters are replaced with a new cluster by merging the original clusters in the previous step. Previous two steps are repeated until only one cluster remains and that cluster will contain all the observations. One major advantage of Hierarchical clustering is that we do not need to know the exact number of clusters beforehand and we can choose the formation of the clusters as they merge. The dataset containing 200 mall customers and their information is taken, same as used for the K-Means algorithm, and agglomerative hierarchical clustering is applied. We can visualize the cluster formation in the form of a tree-diagram called a dendrogram, which allows us to illustrate the hierarchical organization of entities just like in Figure 3.

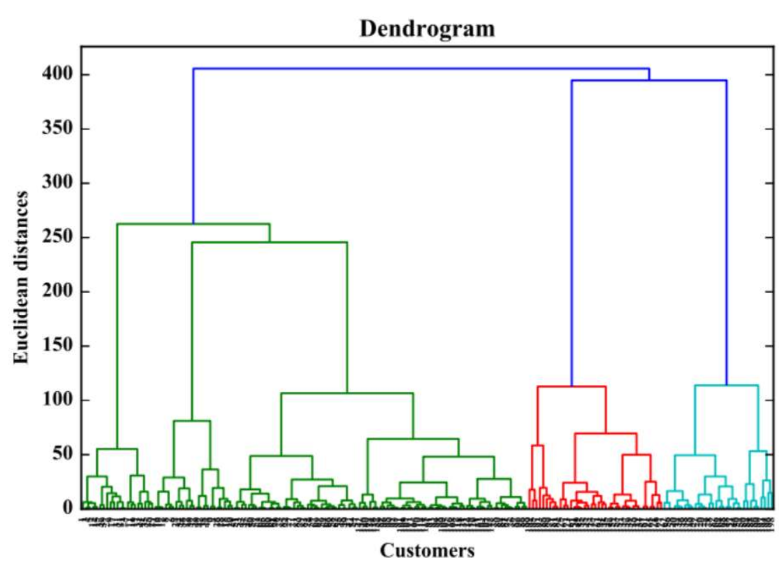


Fig 4.3: Visualizing the clusters using dendrogram.

We have the option of choosing the required number of clusters from the dendrogram itself by selecting the range of maximum distance and then placing a cut-off line at that position. It simply indicates that the distance between the formed clusters is maximum and distinction can be made among them. Hence, according to Figure 3, for satisfactory results, we can choose five clusters (K=5). The clusters are depicted as follows in Figure 4.

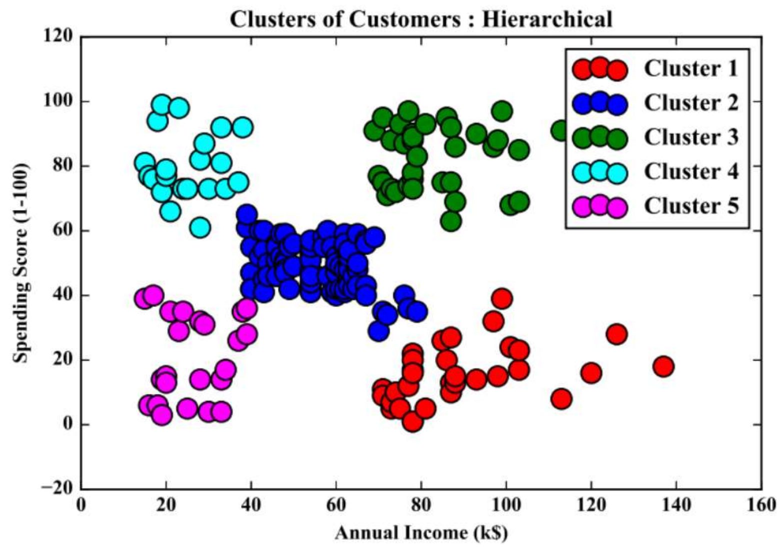


Fig 4.4: Clusters formed using K-means clustering.

Each color represents a different cluster and data points are plotted against Annual Income on the X-axis and the Spending Score on Y-axis. Hierarchical clustering has been extensively used for segmentation purposes due to its ability to produce results in a visual way. It helps in determining the number of clusters for any analysis. It can be used for varied datasets like categorical, spatial and time series with numerical data set being the most common as it consists of data as just real numbers. It was also used and compared with other clustering algorithms in bank customer segmentation. Have made use of a hierarchical pattern-based clustering. The main advantage of Hierarchical clustering is that the output is in the form of a hierarchy(dendrogram) which tells us exactly at which point the clusters merged or split. Hence it is easy to choose and decide on the number of clusters that we wish to take by looking at the dendrogram. However, for a large number of observations its computational speed is very low as compared to the nonhierarchical methods of clustering. Hence, the size and order of the data have an impact on the final results obtained. Nevertheless, one does not need to plug the number of clusters as the input to the algorithm and hence we can have different partitioning groups, the choice of which completely depends on the end user.

Chapter: 5

PRACTICAL APPROACH FOR HIERARCHAL CLUSTERING

5.1 Agglomerative Hierarchical Clustering

The Agglomerative Hierarchical Clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. It's also known as AGNES (Agglomerative Nesting). It's a "bottom-up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

5.2 The mathematics of clustering

The mathematics behind clustering, in very simple terms involves minimizing the sum of square of distances between the cluster centroid and its associated data points:

$$Minimize \sum_{j=1}^{k} \sum_{i=1}^{n} (x_{ij} - c_j)^2$$

- K = number of clusters
- N= number of data points
- C=centroid of cluster j
- (xij — cj)– Distance between data point and centroid to which it is assigned

5.3 How does it work?

1. Make each data point a single-point cluster → forms N clusters
2. Take the two closest data points and make them one cluster → forms N-1 clusters
3. Take the two closest clusters and make them one cluster → Forms N-2 clusters.
4. Repeat step-3 until you are left with only one cluster.

There are several ways to measure the distance between clusters in order to decide the rules for clustering, and they are often called Linkage Methods. Some of the common linkage methods are:

- **Complete-linkage**: the distance between two clusters is defined as the longest distance between two points in each cluster.
- **Single-linkage**: the distance between two clusters is defined as the shortest distance between two points in each cluster. This linkage may be used to detect high values in your dataset which may be outliers as they will be merged at the end.

- **Average-linkage**: the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster.
- **Centroid-linkage:** finds the centroid of cluster 1 and centroid of cluster 2, and then calculates the distance between the two before merging.

The choice of linkage method entirely depends on you and there is no hard and fast method that will always give you good results. Different linkage methods lead to different clusters.

The point of doing all this is to demonstrate the way hierarchical clustering works, it maintains a memory of how we went through this process and that memory is stored in **Dendrogram**. A Dendrogram is a type of tree diagram showing hierarchical relationships between different sets of data.

As already said a Dendrogram contains the memory of hierarchical clustering algorithm, so just by looking at the Dendrogram you can tell how the cluster is formed.

Note:

1. Distance between data points represents dissimilarities.
2. Height of the blocks represents the distance between clusters.

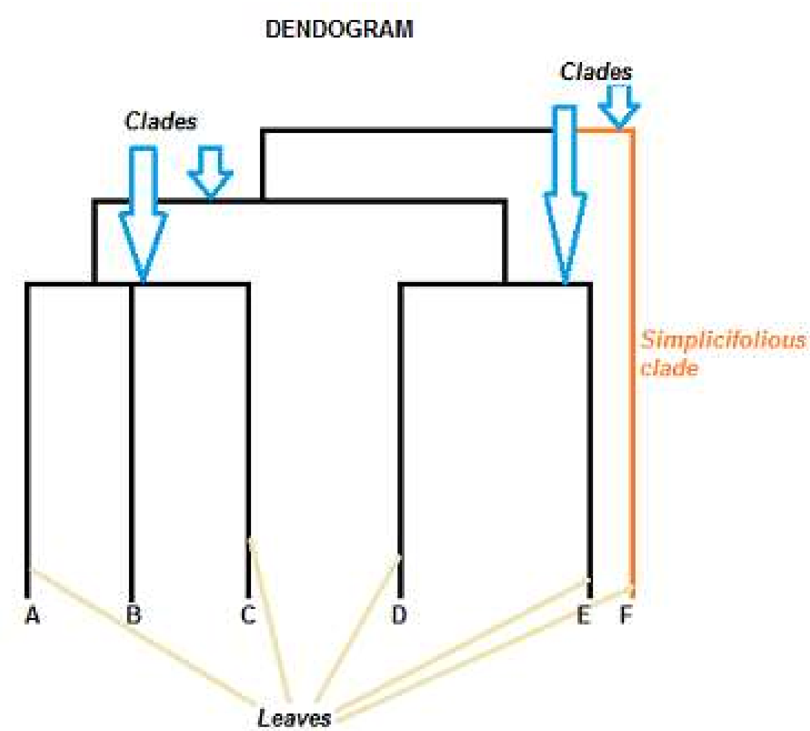


Fig 5.1: Parts of Dendrogram.

A dendrogram can be a column graph (as in the image below) or a row graph. Some dendrograms are circular or have a fluid-shape, but the software will usually produce a row or column graph. No matter what the shape, the basic graph comprises the same parts:

- The ***Clades*** are the branch and are arranged according to how similar (or dissimilar) they are. Clades that are close to the same height are similar to each other; clades with different heights are dissimilar — **the greater the difference in height, the more dissimilarity.**
- Each clade has one or more ***leaves***.
- Leaves A, B, and C are more similar to each other than they are to leaves D, E, or F.
- Leaves D and E are more similar to each other than they are to leaves A, B, C, or F.
- Leaf F is substantially different from all of the other leaves.

A clade can theoretically have an infinite amount of leaves. However, the more leaves you have, the harder the graph will be to read with the naked eye.

One question that might have intrigued you by now is how do you decide when to stop merging the clusters?

You cut the dendrogram tree with a horizontal line at a height where the line can traverse the maximum distance up and down without intersecting the merging point.

For example, in the below figure L3 can traverse maximum distance up and down without intersecting the merging points. So, we draw a horizontal line and the number of vertical lines it intersects is the optimal number of clusters.

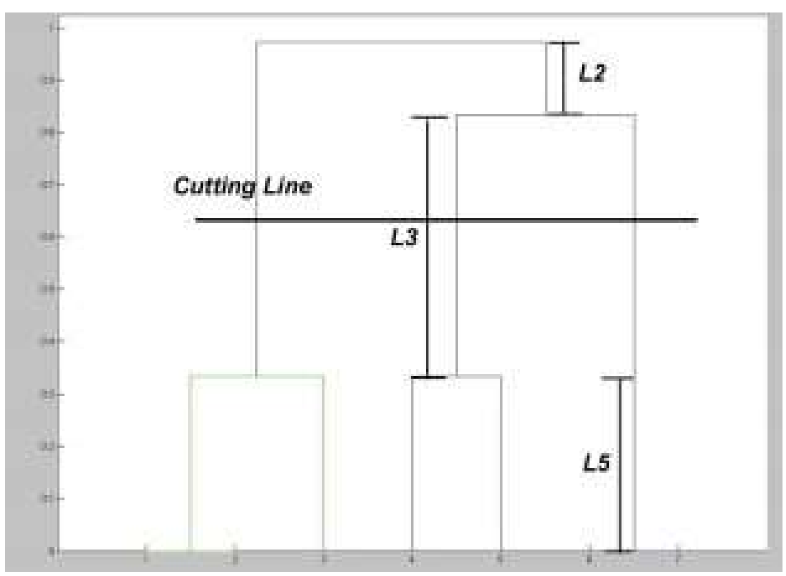


Fig 5.2: Example.

Number of Clusters in this case = 3.

Chapter: 6

PRACTICAL IMPLEMENTATION OF HIERARCHIAL CLUSTERING

6.1 Measuring the goodness of Clusters

Well, there are many measures to do this, perhaps the most popular one is the Dunn's Index. Dunn's index is the ratio between the minimum inter-cluster distances to the maximum intra-cluster diameter. The diameter of a cluster is the distance between its two furthermost points. In order to have well separated and compact clusters you should aim for a higher Dunn's index.

Now let's implement one use case scenario using Agglomerative Hierarchical clustering algorithm. The data set consist of customer details of one particular shopping mall along with their spending score.

Let's start by importing 3 basic libraries:

```
import numpy as np

import matplotlib.pyplot as plt

import pandas as pd
```

Load the dataset:

```
dataset = pd.read_csv('/.../Mall_Customers.csv')
```

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Fig 6.1: Dataset used.

So, our goal is to cluster customers based on their spending score. Out of all the features, CustomerID and Genre are irrelevant fields and can be dropped and create a matrix of independent variables by select only Agenda Annual Income.

```
X = dataset.iloc[:, [3, 4]].values
```

Next, we need to choose the number of clusters and for doing this we'll use Dendrograms.

```
import scipy.cluster.hierarchy as sch

dendrogrm = sch.dendrogram(sch.linkage(X, method = 'ward'))

plt.title('Dendrogram')

plt.xlabel('Customers')

plt.ylabel('Euclidean distance')

plt.show()
```

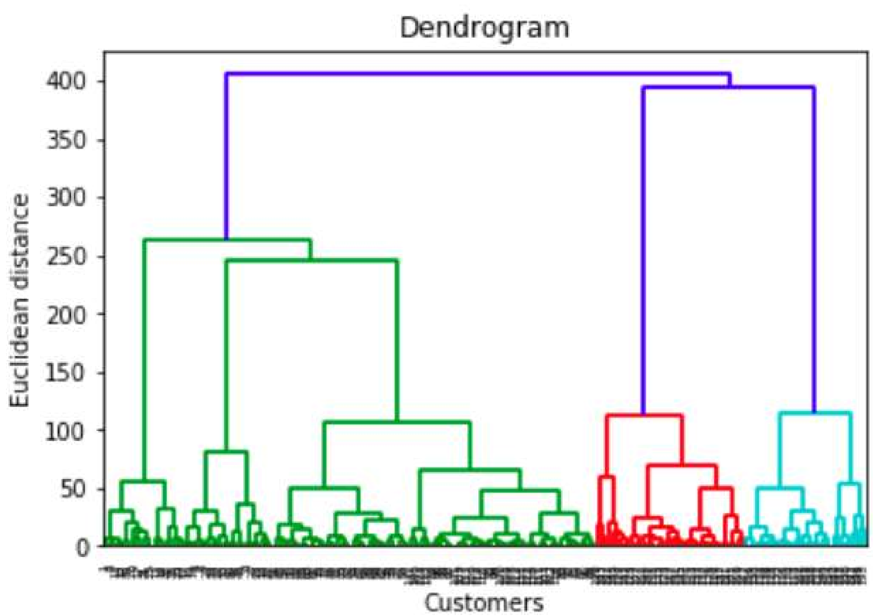


Fig 6.2: Example Dendrogram.

As we have already discussed to choose the number of clusters, we draw a horizontal line to the longest line that traverses maximum distance up and down without intersecting the merging points. So, we draw a horizontal line and the number of vertical lines it intersects is the optimal number of clusters. In this case, it's 5. So, let's fit our Agglomerative model with 5 clusters.

```
from sklearn.cluster import AgglomerativeClustering

hc = AgglomerativeClustering(n_clusters = 5, affinity = 'euclidean', linkage =
'ward')

y_hc = hc.fit_predict(X)
```

Visualize the results.

```
plt.scatter(X[y_hc == 0, 0], X[y_hc == 0, 1], s = 50, c = 'red', label = 'Careful')

plt.scatter(X[y_hc == 1, 0], X[y_hc == 1, 1], s = 50, c = 'blue', label = 'Standard')

plt.scatter(X[y_hc == 2, 0], X[y_hc == 2, 1], s = 50, c = 'green', label = 'Target')

plt.scatter(X[y_hc == 3, 0], X[y_hc == 3, 1], s = 50, c = 'cyan', label = 'Careless')

plt.scatter(X[y_hc == 4, 0], X[y_hc == 4, 1], s = 50, c = 'magenta', label = 'Sensible')

plt.title('Clusters of customers')

plt.xlabel('Annual Income (k$)')

plt.ylabel('Spending Score (1-100)')

plt.legend()

plt.show()
```

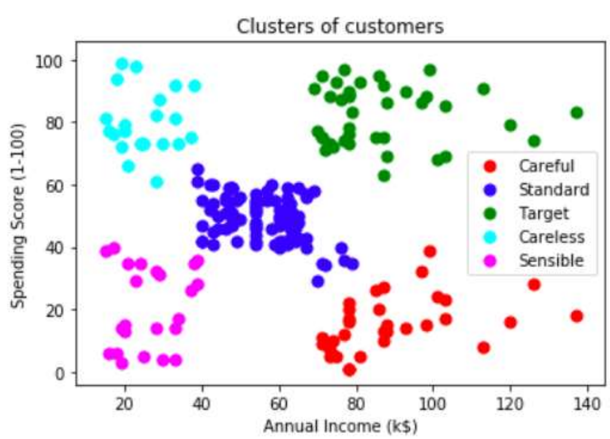


Fig 6.3: Resultant Cluster.

CONCLUSION

Due to increasing commercialization, consumer data is increasing exponentially. When dealing with this large magnitude of data, organizations need to make use of more efficient clustering algorithms for customer segmentation. These clustering models need to possess the capability to process this enormous data effectively. Each of the above discussed clustering algorithms come with their own set of merits and demerits. The computational speed of K-Means clustering algorithm is relatively better as compared to the hierarchical clustering algorithms as the latter require the calculation of the full proximity matrix after each iteration. K-Means clustering gives better performance for a large number of observations while hierarchical clustering has the ability to handle fewer data points.

The major hindrance produces itself in the form of selecting the numbers of clusters „K‟ for the K-Means process, which have to be provided as an input to this non-hierarchical clustering algorithm. This limitation does not exist in the case of hierarchical clustering since it does not require any cluster centers as input. It depends on the user to choose the cluster groups as well as their number. Hierarchical clustering also gives better results as compared to K-Means when a random dataset is used. The output or results obtained when using hierarchical clustering are in the form of dendrograms but the output of K-Means consists of flat structured clusters which may be difficult to analyze. As the value of k increases, the quality(accuracy) of hierarchical clustering improves when compared to K-Means clustering.

As such, partitioning algorithms like K-Means are suitable for large datasets while hierarchical clustering algorithms are more suitable for small datasets. Both K-Means and Hierarchical clustering have drawbacks that make them unsuitable when used individually. For business use, data visualization forms a major part of efficient data analysis and hierarchical clustering aids in doing so. Furthermore, when the performance aspect is taken into account, K-Means tends to deliver better results. With the vices and virtues of the two techniques pointed out, it comes to light that an amalgam of the best of these algorithms could outperform the individual models. In summary, different clustering algorithms, owing to their properties towards different kinds of data can be used in succession such that the advantages of these techniques could be harnessed in full. However, the selection process of these suitable techniques as well as their judicious implementation could require a considerable time investment in studying and processing the data along with an adequate understanding of the goals and requirements.

REFERENCE

[1] E. Ngai, L. Xiu and D. Chau, “Application of data mining techniques in customer relationship management: A literature review and classification”, Expert Systems with Applications, vol. 36, no. 2, pp. 2592-2602, 2009.

[2] J. Peppard, “Customer Relationship Management (CRM) in financial services”, European Management Journal, vol. 18, no. 3, pp. 312-327, 2000.

[3] A. Ansari and A. Riasi, “Taxonomy of marketing strategies using bank customers clustering”, International Journal of Business and Management, vol. 11, no. 7, pp. 106-119, 2016.

[4] M. Ghzanfari, et al., “Customer segmentation in clothing exports based on clustering algorithm”, Iranian Journal of Trade Studies, vol. 14, no. 56, pp. 59-86, 2010.

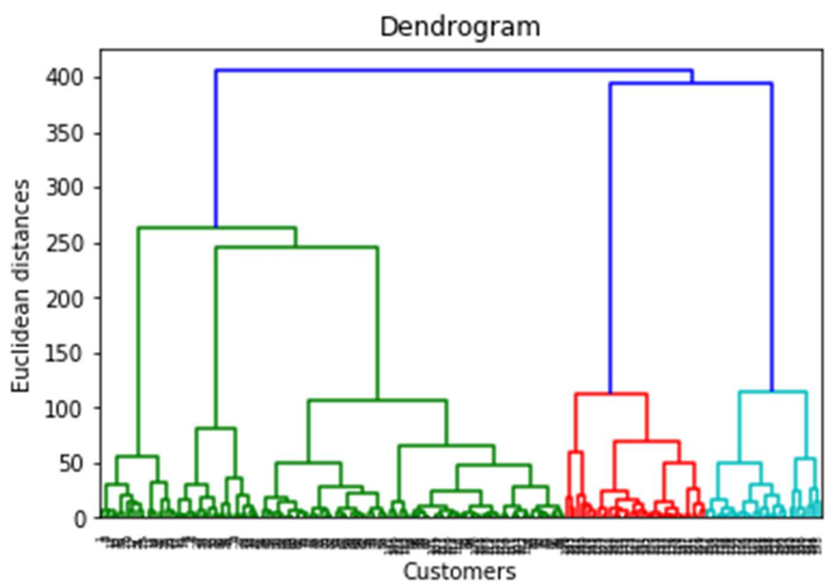
[5] C. Rygielski, J. Wang and D. Yen, “Data mining techniques for customer relationship management”, Technology in Society, vol. 24, no. 4, pp. 483-502, 2002.

[6] Approaches to Clustering in Customer Segmentation Shreya Tripathi1*, Aditya Bhardwaj2 , Poovammal E3’

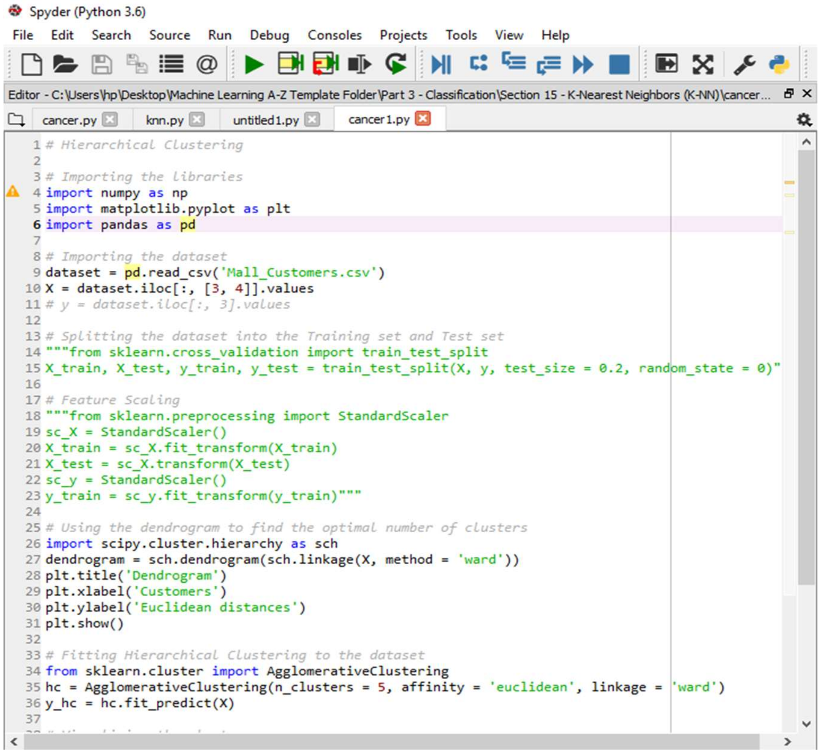
[7] Google.com.

SCREENSHOT GALLERY

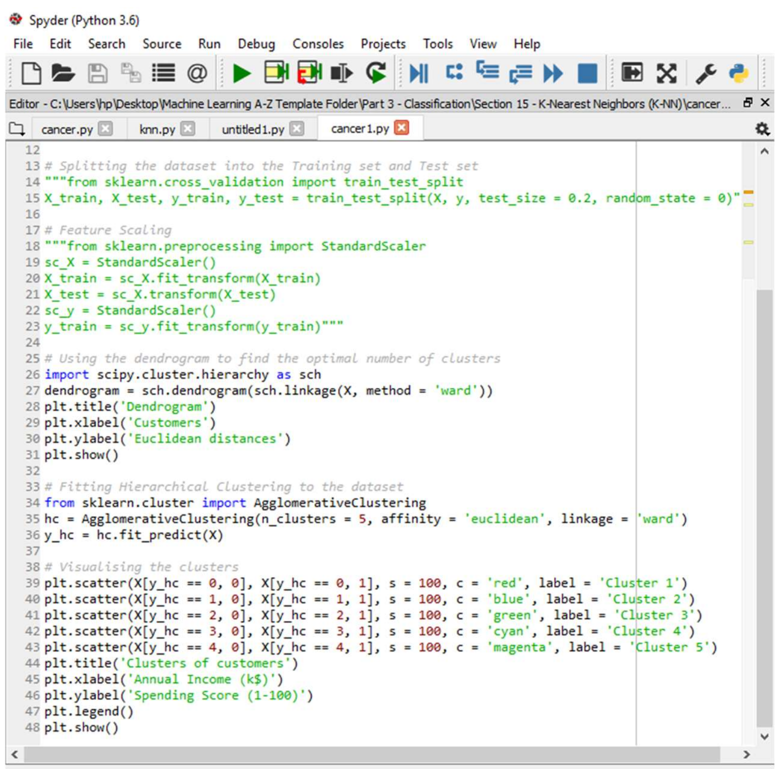
Screenshot 1: Final Clusters.



Screenshot 2: Final Dendrogram based on the dataset.



Screenshot 3: Code for dendrogram Screenshot.



Screenshot 4: Cluster Code Screenshot.