Application Development for Customer Segmentation Using an Unsupervised Learning Algorithm

M. Nirmala*[1], M. Shah Makzoom[2]

*[1]Department of Computer Applications, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India

[2]Department of Computer Applications, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India

A R T I C L E I N F O

A B S T R A C T

Making wise selections is a requirement for any business to produce healthy revenue. Every customer in a firm is unique in comparison to other possible customers, and each one has various preferences for and objections to the product. The clients are divided into groups based on shared traits including gender, age, interests, and spending patterns. Targeting a certain group of customers with specialized goods, services, and marketing methods is made simpler by the customer segmentation process. This study focuses on segmenting customers based on their income and spending scores. The segmentation is done using the KMeans clustering algorithm, an unsupervised learning mechanism, and the total data is divided into three clusters. The model is then implemented using Python and Flask web framework.

Keywords: Customer, Segmentation, KMeans Algorithm, Clustering, Flask Framework

I. INTRODUCTION

Every organization relies heavily on data, which is important for expansion and generating revenue. We can identify the client's interests using grouped or ungrouped data. Consumer segmentation is a challenge since it requires breaking out a vast array of data into different groups according to factors like age, demographics, spending, income, gender, etc. These teams are also referred to as clusters. This allows us to learn which products are doing well and what demographic is buying them, among other things. Also,

we can significantly increase the product's supply to increase revenue. The ideal use of unsupervised classification is customer segmentation. Find client segments in the dataset using clustering to target the possible user base. Algorithms for machine learning open the door to tailored suggestions. A system may examine relationships and trends in enormous volumes of historical data using machine learning, a subset of artificial intelligence.

This paper is arranged as follows: Section 2 deals with the Literature survey associated with this paper.

Section 3 explains the concepts of unsupervised learning and KMeans algorithm, Section 4 discusses the experimental setting requirements. Section 5 explains the implementation part and integrating the model with Flask Framework. Section 6 concludes the paper with results.

II. LITERATURE SURVEY

By better utilizing their marketing funds and gaining a competitive edge over their rival businesses, businesses can apply the customer segmentation technique. Additionally, it aids a company in planning the marketing budget, identifying client retention, identifying new market prospects, and enhancing marketing efficiency. By better utilizing their marketing funds and gaining a competitive edge over their rival businesses, businesses can apply the customer segmentation technique [1].

According to sentiment analysis, recommender systems have become a crucial area of study, as Nidhi Kapoor [2] noted.

Liu [3] divided the three articles into three categories: collaborative filtering (CF), content-based, and context-based.

The studies by Tripathi et al. [4] provided statistics about developments in recommender system research and provided practitioners and scholars with insight and future directions on the recommender system utilizing sentimental analysis.

In order to categorize customers, Hossain [5] mentioned a number of factors, including age, gender, family, education level, and income. Big Data has many different properties, including size, noise, dimensionality, algorithm calculations, cluster structure, and a quick summary of the different clustering techniques categorized under partitioning, hierarchical, density, grid-based, and model-based algorithms.

Chinedu et al [6] explored the necessity of segmentation of the customers using clustering algorithms as the core functionality of CRM. The commonly used K-Means and Hierarchical Clustering were studied and the two approaches were integrated.

This paper [7] talks about how to use clustering algorithms in consumer segmentation to Applications of Unsupervised Learning make a marketing plan for a credit card company. It demonstrates how the PCA can be used as a reliable check tool for K-means and hierarchical clustering during the clustering process.

Using a collaborative filtering methodology, this research [8] provided a recommendation system. According to test results, the suggested strategy offers sound advice. The suggested action can be used in various fields to advertise items like movies, music, and other goods.

This paper [9] talks about how businesses can use formed clusters to find specific customers and show them content on social media platforms and in marketing campaigns that they are really interested in.

A [10] solution is put forth to divide the client group into the premium and standard categories using the machine learning techniques NEM, LiRM, and LoRM..

III. UNSUPERVISED LEARNING

A type of machine learning called unsupervised learning is used to derive conclusions from unlabelled data. There is no way to predict the right response. It recognizes the patterns in the data and responds to their presence or absence in each individual piece of data. It makes assumptions based on patterns.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 2

128

For instance, based on their purchase patterns, clients are divided into three separate categories.
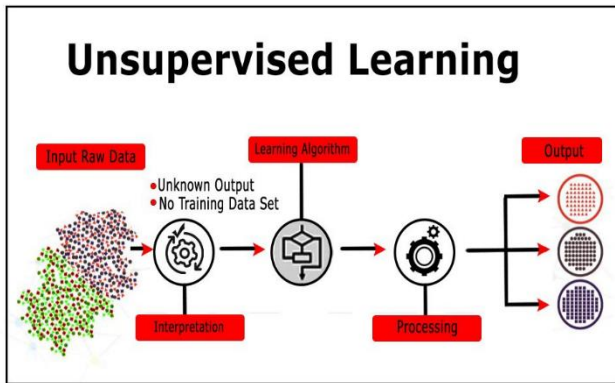

Figure 1 : Unsupervised Learning

- Learning associations
- Clustering
- Density estimation
- Dimensionality reduction
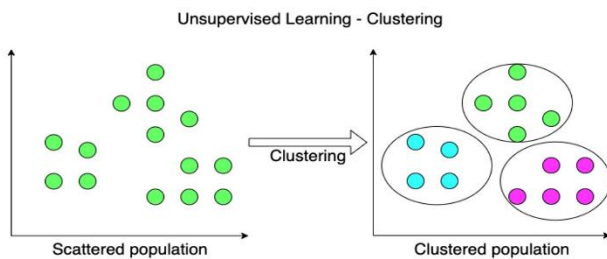- Feature selection
- Outlier/novelty detection.

Figure 2 : Unsupervised Learning – Clustering

K-Means Clustering Algorithm

Table 1 : KMeans Algorithm Steps

Steps	Descriptions
1	Decide on K clusters.
2	Data cluster centre selection. Choose a cluster centre from among K data points at random. Choose your cluster centres so that they are as widely apart as you can make them

3	Measure the distance between each data point and the cluster centre. Either the provided distance function or the Euclid mean distance formula may be used to compute the distance. $$p(a,b) = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$ $$P(a, b) = |x2 - x1| + |y2 - y1|$$
4	Each data point should be assigned to a cluster. The cluster whose centre is closest to the given data point receives the associated data point.
5	The centers of freshly generated clusters should be recalculated. By calculating the mean of all the data points continued in a cluster, the cluster's center can be determined.
6	Repeat Steps 3 through 5 repeatedly until one or more of the following stopping criteria is reached: Newly created clusters' centers remain unchanged. There are still data points in the same cluster. There are no more iterations possible.

Below is a sample for eight points (with (x, y) representing locations) into three clusters A1(2, 10) A2(2, 5) A3(8, 4) A4(5, 8) A5(7, 5) A6(4, 4) A7(1, 2) A8(4, 9). Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2).

Table 2 : Sample Eight Point Tabulation with 3 initial clusters

	Point	(2, 10) Dist Mean 1	(5, 8) Dist Mean 2	(1, 2) Dist Mean 3	Cluster
A1	(2, 10)	0	5	9	1
A2	(2, 5)	5	6	4	3
A3	(8, 4)	12	7	9	2
A4	(5, 8)	5	0	10	2
A5	(7, 5)	10	5	9	2
A6	(6, 4)	10	5	7	2
A7	(1, 2)	9	10	0	3
A8	(4, 9)	3	2	10	2

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 2

129

IV. SYSTEM SPECIFICATIONS

Hardware Requirements

PROCESSOR   : Intel Core i5
RAM   : 12GB
HARD DISK DRIVE   : 512GB

Software Requirements

OPERATING SYSTEM : WINDOWS 11
APPLICATION   : JUPYTER NOTEBOOK
                        PYTHON
                        VISUAL STUDIO CODE
                        FLASK Framework

V. IMPLEMENTATION

The dataset was only produced for instructional purposes, specifically to teach market basket analysis and consumer segmentation ideas. Information from the data set includes the customer's annual income and spending score.

Based on predetermined criteria, including customer behaviour and purchasing power, spending scores are assigned.

One of the most popular techniques for examining data and gaining a thorough understanding of its structure is clustering.

The task of locating the subtitles and subgroups in the entire dataset can be described as such. Several subgroups with similar data are grouped together.
Due to some similarities, a cluster is a group of aggregated data points. Market basket analysis uses clustering to divide up the consumer base based on their behaviours and transactions.

The dataset contains 302 rows and 2 columns called as INCOME and SPEND. The data set is clean and no missing values are present.

The Elbow point graph is drawn to infer the number of clusters for the given problem. The number of cluster classification for this problem is identified as 3.
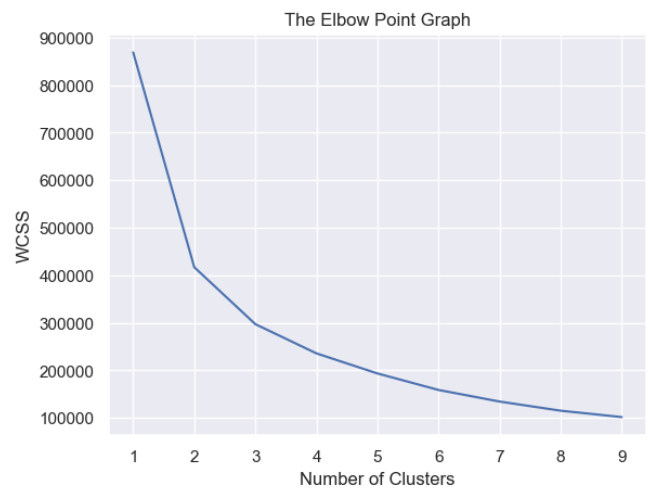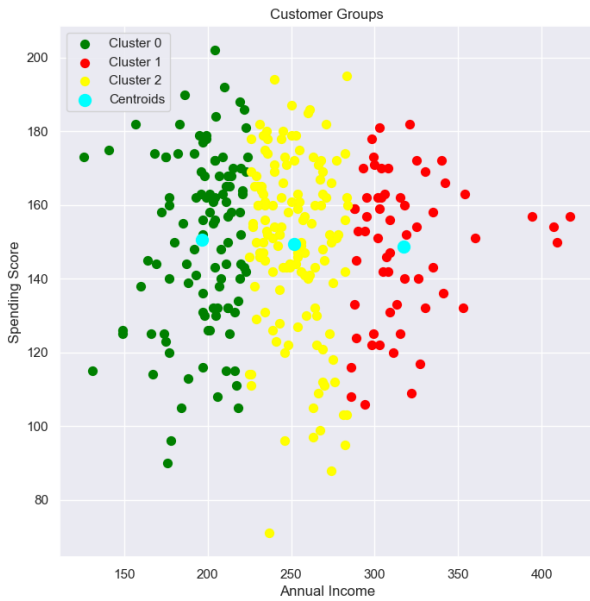
Figure 3 : Elbow Method

For the given dataset after applying the K means cluster algorithm the segmentation of the data set based upon clusters are given below as 3 Clusters - 0, 1, 2

```
[2 2 0 2 1 0 1 2 0 0 2 2 2 0 2 0 1 2 2 2 2 2 2 2 0 1 0 0 1 0 0 0 0 2 0 0 1
 2 2 1 1 2 0 2 1 1 2 2 0 2 2 1 2 0 2 0 0 2 0 1 2 1 0 0 0 0 2 0 0 2 2 0 2
 0 2 2 0 0 2 2 1 1 1 2 2 2 0 0 2 2 0 0 1 0 2 1 2 1 2 2 2 0 2 0 0 2 2 2 1
 0 1 0 2 0 0 0 0 2 1 2 2 2 0 0 0 2 0 2 0 2 1 2 1 2 0 0 0 2 1 1 0 0 0 2 2 2
 2 0 2 0 2 2 0 0 2 0 0 0 2 1 0 0 0 1 2 2 2 0 2 2 2 0 0 0 2 1 0 2 1 2 1 2 2
 1 2 2 2 0 1 0 0 2 0 1 2 2 2 2 0 2 2 2 0 0 2 2 0 1 2 1 2 1 2 2 1 0
 2 1 2 0 2 0 1 1 2 1 1 2 1 1 1 1 1 2 2 2 0 2 0 2 1 2 2 2 1 2 1 1 1 2 1 2 0 2
 2 2 2 2 0 0 1 0 1 2 2 2 2 2 2 0 0 2 1 0 1 0 0 0 0 1 0 2 1 0 0 1 2 0 0 0
 0 0 2 2 0 0 2]
```

Figure 4 : Clustered Output for the Data Set

Visualizing the clusters based on Annual Income and Spending Score and the number of cluster classification depicted is 3 shown in Green, yellow and red dots. The teal green dot among every cluster shows the centroid point of the specific cluster.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 2

130

Figure 5 : Cluster Visualization with 3 Segments

Table 3 : Customer Values Segmentation based on Cluster

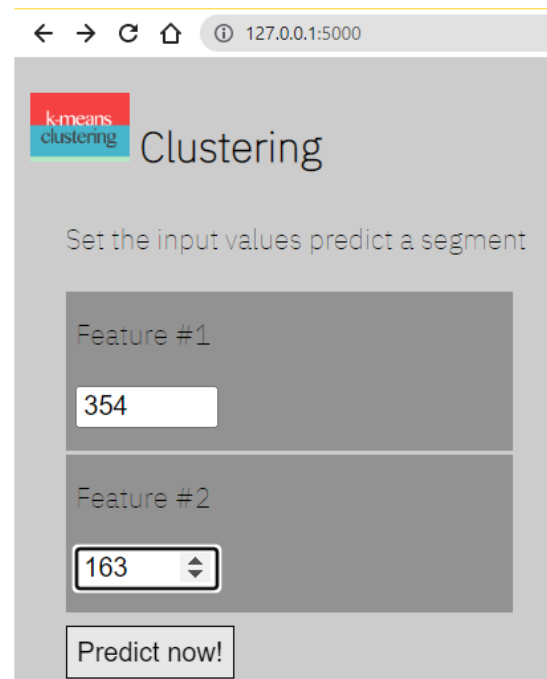| Cluster	Count
Cluster 01	107
Cluster 02	60
Cluster 03	135

Steps for Integrating an Unsupervised Machine Learning application with Flask Framework

- Create the Machine Learning Model
- Save the centroids as Pickle file Locally
- Develop the web Application in Flask and integrate with the model created.
- Create a Template folder which contains index.html and predict.html and appropriately link the html files with the application file
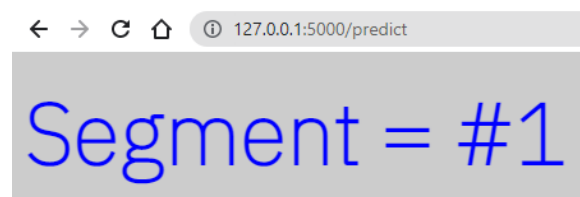
VI. Results and Conclusions

The machine learning model has been created using the KMeans clustering Unsupervised learning Algorithm. The annual income and the spending score of the customer column is taken into consideration for the concept of customer segmentation process. After

the model is created, Save centroids as pickle file locally as freezed_centroids.pkl and create an app.py which contains 2 routing files one as a home page where the details of Annual Income and Spending score is given as Input is referred in Figure 6, Figure 8, Figure 10 and their respective output is given in Figure 7, Figure 9, Figure 11. With the help of k means clustering algorithm we are able to classify the data points to specific cluster and the segmentation of customer based on annual income and spending score is segregated.

Figure 6 : Input Values for Customer Segment 1

Figure 7 : Result of Customer Segment 1

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 2
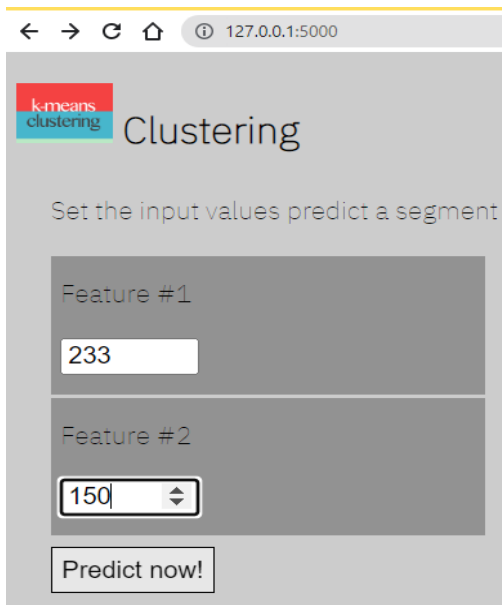
131

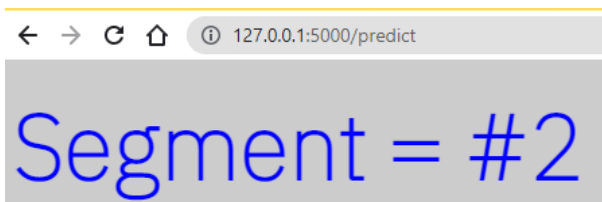Figure 8 : Input Values for Customer Segment 2
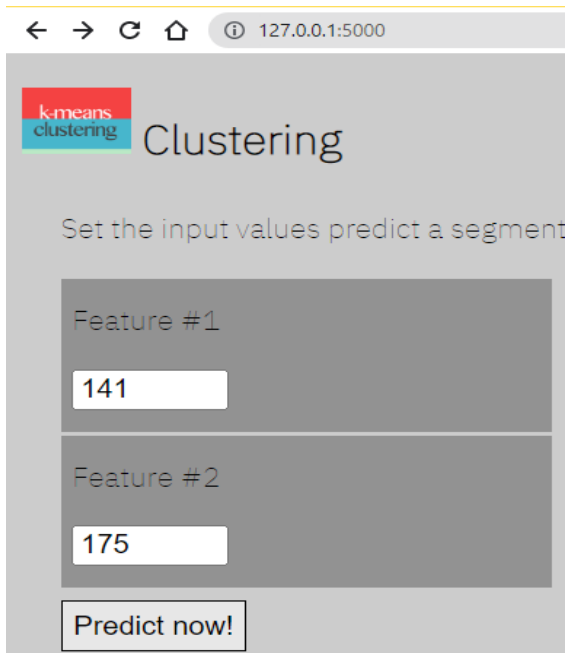

Figure 9 : Result of Customer Segment 1

Figure 10 : Input Values for Customer Segment 0

Figure 11 : Result of Customer Segment 0

VII. REFERENCES

[1]. Tushar Kansal; Suraj Bahuguna; Vishal Singh; Tanupriya Choudhury Customer, "Segmentation based on RFM model and Clustering Techniques With K-Means Algorithm", IEEE 2018

[2]. Muhammad Iqbal Dzulhaq, Kartika Wulan Sari, Syaipul Ramdhan, Rahmat Tullah, Sutarman," Customer Segmentation Based on RFM Value Using K-Means Algorithm", ICIC 2019

[3]. Chaohua Liu," Customer Segmentation and Evaluation Based On RFM, Cross- selling and Customer Loyalty", IEEE 2011

[4]. Tripathi, Aditya Bhardwaj, Poovammal," Approaches to Clustering in Customer Segmentation", IJET 2018

[5]. A.S.M. Shahadat Hossain," Customer Segmentation using Centroid Based and Density Based Clustering Algorithms", 2017 3rd International Conference on Electrical Information and Communication Technology (EICT).

[6]. Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu, "Application of K-Means Algorithm for Efficient Customer Segmentation", International Journal of Advanced Research in Artificial Intelligence, 10.14569/IJARAI.2015.041007, October 2015

[7]. Azad Abdulhafedh, "Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation", Journal of City and Development. 2021, 3(1), 12-30. DOI: 10.12691/jcd-3-1-3,Feb 02, 2021.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 2

132

[8]. K. Windler, U. Juttner, S. Michel, S. Maklan, and E.K. ¨Macdonald, "Identifying the right solution customers: A managerial methodology," Industrial Marketing Management, vol. 60, pp. 173 –186, 2017.

[9]. Prof. Nikhil Patankar, Soham Dixit, Akshay Bhamare, Ashutosh Darpel and Ritik Raina, "A Case Study on Customer Segmentation by using Machine Learning Methods", IEEE, Recent Trends in Intensive ComputingYear: December 2021.

[10].Sukru Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods",IEEE, Year: 2018.

VIII. Biographies and Photographs

M.Nirmala, received the B.Sc Degree in Computer Science from Bharathiar University, MCA from Bharathidasan University and M.Phil from Bharathiar University. She has a total teaching experience of 20+ years. Her area of interest includes Networking, Simulation & Modeling, Machine Learning and Deep Learning. Received the Best Paper Award and Published in IEEE explore. She has published more than 18 International Journals, 12 International Conferences and authored 4 Books.

M. SHAH MAKZOOM received his BCA in Computer Applications from Nesamony Memorial Christian College and currently pursuing Final year MCA from Hindusthan College of Engineering and Technology, an Autonomous college affiliated to Anna University. His area of interest includes Machine Learning and IOT. Currently doing projects in IOT.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 2

133