VISVESVARAYATECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018, Karnataka

Advanced Machine Learning (18AI72)

Mini Project Report on

"Customer Segmentation Using Machine Learning"

Submitted in partial fulfillment for the award of the degree of

Bachelor of Engineering
in
Artificial Intelligence & Machine Learning

Submitted by

USN	Name
1BI20AI007	Ms. Bhavika Paladugu
1BI20AI011	Ms. Dhanyashree P N

for the academic year 2023-24

Under the Guidance

Ms. Subha Meenakshi S
Guest Faculty
Department of AI&ML,
BIT, Bengaluru -560 004

Department of Artificial Intelligence & Machine Learning
Bangalore Institute of Technology
K.R. Road, V.V.Pura, Bengaluru-560 004

2023-24

VISVESVARAYA TECHNOLOGICAL UNIVERSITY
"Jnana Sangama", Belagavi-590018, Karnataka"

BANGALORE INSTITUTE OF TECHNOLOGY
Department of Artificial Intelligence & Machine Learning K.R. Road, V.V.Pura, Bengaluru-560 004

Certificate

This is to certify that Advanced Machine Learning mini project work entitled **"Customer Segmentation Using Machine Learning"** carried out by

USN	Name
1BI20AI007	**Ms. Bhavika Paladugu**
1BI20AI011	**Ms. Dhanyashree P N**

bonafide students of **Bangalore Institute of Technology** in partial fulfilment for the award of degree of **Bachelor of Engineering** in **Artificial Intelligence & Machine Learning** under Visvesvaraya Technological University, Belagavi, during the academic year 2023-24 is true representation of mini project work completed satisfactorily.

Ms. Subha Meenakshi S Guest Faculty Dept. of AI&ML BIT, Bengaluru	Dr. Jyothi D. G. Professor & HoD Dept. of AI&ML BIT, Bengaluru.	Dr. Aswath M. U. Principal BIT Bengaluru.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the completion of a task would be incomplete without crediting the people who made it possible, whose constant guidance and encouragement crowned the efforts with success.

We would like to profoundly thank Management of Bangalore Institute of Technology for providing such a healthy environment for successful completion of Mini Project work.

We would like to express our thanks to the Principal **Dr. Aswath M.U.** for his encouragement that motivates us for the successful completion of Mini Project work.

It gives us immense pleasure to thank **Dr. Jyothi D.G.**, Professor & Head, Department of Artificial Intelligence & Machine Learning for her constant support and encouragement.

We would like to express our deepest gratitude to our mini project guide **Ms. Subha Meenakshi S** for their constant support and guidance throughout the Mini Project work.

We are very much pleasured to express our sincere gratitude to the friendly co-operation showed by all the staff members of Artificial Intelligence and Machine Learning Department, BIT.

Last but not the least, we would here by acknowledge and thank our friends and family who have been our source of inspiration always instrumental in successful completion of the Project work.

Date: -19-12-2023 Ms. Bhavika Paladugu

Place: Bengaluru Ms. Dhanyashree P N

ABSTRACT

Effective decisions are mandatory for any company to generate good revenue. In these days competition is huge and all companies are moving forward with their own different strategies. We should use data and take a proper decision. Every person is different from one another and we don't know what he/she buys or what their likes are. But, 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 business the best. We are going to use elbow method to find the number of clusters and at last we visualize the data.

LIST OF FIGURES

INDEX

CHAPTER-I

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Overview

Nowadays the competition is vast and lot of technologies came into account for effective growth and revenue generation. For every business the most important component is data. With the help of grouped or ungrouped data, we can perform some operations to find customer interests.

Data mining helpful to extract data from the database in a human readable format. But, we may not known the actual beneficiaries in the whole dataset. Customer Segmentation is useful to divide the large data from dataset into several groups based on their age, demographics, spent, income, gender, etc. These groups are also known as clusters. By this, we can get to know that, which product got huge number of sales and which age group are purchasing etc. And, we can supply that product much for better revenue generation.

Initially we are going to take the old data. As we know that old is gold so, by using the old data we are going to apply K-means clustering algorithm and we have to find the number of clusters first. So, at lastly, we have to visualize the data. One can easily find the potential group of data while observing that visualization.

The main goal of this paper is to identify the similar customer segments using the data mining approach , using the partitioning unsupervised algorithm called as K-means clustering algorithm. The elbow method determines the optimal clusters.

CHAPTER – II

LITERATURE REVIEW

CHAPTER 2

LITERATURE REVIEW

2.1 LITERATURE REVIEW

Customer segmentation is a fundamental practice in modern marketing, aiming to divide a diverse customer base into homogeneous groups. This segmentation helps businesses better understand their customers, design targeted marketing strategies, and ultimately improve sales. Clustering, a popular data analysis technique, plays a crucial role in identifying these distinct customer segments. This literature survey explores key research papers and studies related to customer segmentation analysis using clustering for the purpose of sales improvement.

"Market Segmentation: Conceptual and Methodological Foundations" by Michel Wedel and Wagner Kamakura (2000)

This foundational work introduces the concept of market segmentation and highlights its importance in marketing strategy. It covers various segmentation techniques, including clustering, and discusses their benefits in improving sales and customer satisfaction. The paper provides a comprehensive overview of segmentation methods and their applications.

"Customer Segmentation: A Review" by V. Kumar and Rohit Aggarwal (2019)

This review paper provides an in-depth analysis of customer segmentation techniques, focusing on both traditional methods and modern data-driven approaches. It emphasizes the role of clustering in creating meaningful customer segments and discusses the impact of segmentation on sales and business outcomes. The paper highlights the importance of considering various data sources, such as transaction data and customer behaviour, in the segmentation process.

"Customer Segmentation Based on Purchasing Behaviour Using K-Means Clustering" by M.A. Hossain and Mohammad Shorif Uddin (2016)

This research paper focuses on a practical application of clustering, specifically K-means clustering, for customer segmentation based on purchasing behaviour. The study demonstrates the effectiveness of clustering in identifying distinct customer groups and proposes strategies to target

these segments. The findings suggest that personalized marketing approaches derived from clustering analysis can significantly improve sales.

"An Empirical Analysis of Customer Segmentation Strategies Using Clustering" by Elham Fadaly et al. (2018)

This empirical study explores different customer segmentation strategies using clustering techniques and evaluates their impact on sales and customer retention. The paper compares K-means clustering with other methods and discusses the advantages and limitations of each approach. It provides valuable insights into the practical implementation of clustering for sales improvement.

"Enhancing Customer Segmentation with Machine Learning" by Abhijit J. Patil and Prashant R. Nair (2021)

This paper discusses the integration of machine learning techniques, including clustering algorithms, for customer segmentation. It emphasizes the benefits of utilizing advanced algorithms to uncover complex patterns in customer data. The study showcases real-world examples of businesses that have successfully improved sales by adopting machine learning-driven segmentation strategies.

"Customer Segmentation for E-commerce: A Comparative Study of Clustering Algorithms" by G. Santosh Kumar and P. Senthil Kumar (2018)

This comparative study evaluates the performance of various clustering algorithms for customer segmentation in the e-commerce domain. The paper assesses the effectiveness of algorithms like K- means, DBSCAN, and hierarchical clustering in identifying meaningful customer segments. It discusses the implications of segmentation on sales and provides insights into algorithm selection for different scenarios.

2.2 EXISTING SYSTEM

The existing method is storing customer data through paperwork and computer software (digital data) is increasing day by day. At end of the day they will analyze their data as how many things are sold or actual customer count etc. By analyzing the collected data they got to know who is beneficial to their business and increase their sales. It requires more time and more paperwork. Also, it is not much effective solution to find the desired customers data.

2.3 PROBLEM STATEMENT

Using Mall Customer Dataset and performing K-Means clustering to segregate customers based on the data in the form of optimal customers.

2.4 PROPOSED SYSTEM

The proposed system aims to revolutionize customer segmentation by employing the K-Means clustering algorithm in a robust and adaptable framework. Beginning with the ingestion of diverse customer data, the system ensures a comprehensive dataset for analysis. The modular design encompasses data preprocessing, feature selection, and the K-Means clustering engine, allowing users to dynamically choose features for segmentation. The resulting clusters are analyzed to formulate targeted marketing strategies and personalized services, fostering an optimized allocation of resources.

The system's core strength lies in its ability to adapt to evolving customer behavior, providing decision-makers with data-driven insights for agile strategy formulation. Through an iterative feedback loop, the system ensures continuous refinement, contributing to a dynamic and customer-centric approach in an ever-changing business landscape.

CHAPTER – III

SYSTEM ARCHITECTURE

CHAPTER 3

SYSTEM ARCHITECTURE

3.1 SYSTEM ARCHITECTURE

Initially we will see the dataset and then we will perform exploratory data analysis which deals with the missing data, duplicates values and null values. And then we will deploy our algorithm k-means clustering which is unsupervised learning in machine learning.

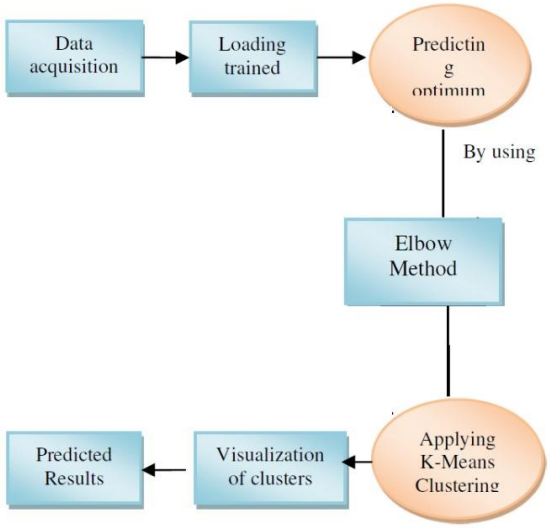

Figure 3.1 Dataflow Diagram of proposed system

As in order to find the no of clusters we use elbow method where distance will be calculating through randomly chosen centers and repeat it until there is no change in cluster centers. Thereafter we will analyze the data through data visualization. Finally, we will get the outcome.

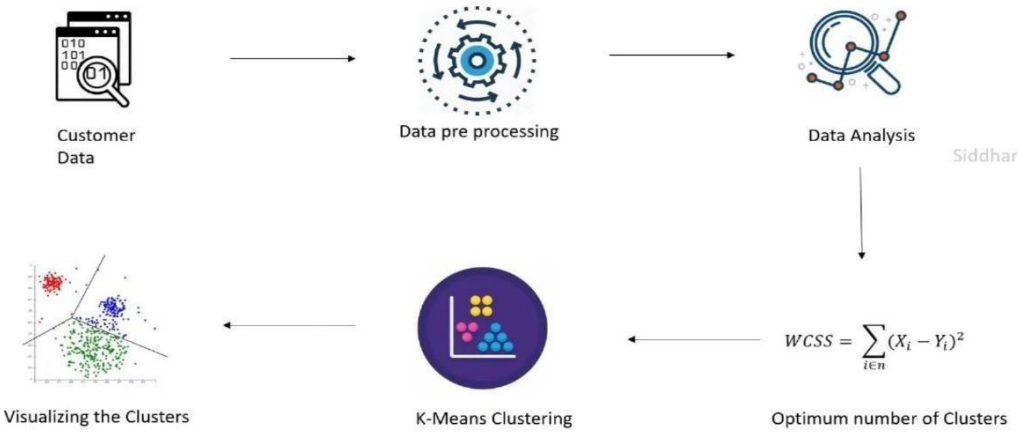

Figure 3.2 Architecture of proposed system

1. First of all, we will import all the necessary libraries or modules (pandas, NumPy, seaborn).
2. Then we will read dataset and analyses whether it contains any null values, missing values and duplicate values. So, we will fix them by dropping or fixing the value with their means, medians etc. which is technically named as Data Preprocessing.
3. We will deploy our model algorithm K-Means Clustering, which divides the data into group of clusters based on similar characteristics. To find no. of clusters we will use elbow method.
4. Finally, we will visualize our data using matplot, which concludes the customers divided into groups who are similar to each other on their group.

3.2 MODULE DESCRIPTION

1. Data Ingestion Module: Ingest data from various sources, such as databases or external APIs, and load it into the system for further processing.

2. Data Preprocessing Module: Handle data cleaning, transformation, and normalization. Address missing values, encode categorical variables, and prepare the data for feature selection.

3. **Feature Selection Module:** Allow users or the system to select relevant features (e.g., annual income, spending score) for clustering. Extract the chosen features from the preprocessed data.

4. **K-Means Clustering Module:** Implement the K-Means clustering algorithm to group customers into clusters based on the selected features. Determine the optimal number of clusters.

5. **Cluster Analysis Module:** Analyze the characteristics of each cluster, providing insights into the distinct customer segments created by the clustering algorithm.

6. **Strategy Formulation Module:** Formulate targeted marketing strategies and personalized services for each customer segment based on the cluster analysis. Collaborate with domain experts for effective strategy development.

7. **Resource Optimization Module:** Optimize resource allocation by considering the identified customer segments. Provide recommendations for efficient allocation of marketing resources and personnel.

8. **Implementation and Monitoring Module:** Implement the formulated strategies in the business operations. Continuously monitor customer behavior and strategy effectiveness, providing real-time feedback to decision-makers.

9. **Customer Satisfaction Enhancement Module:** Implement strategies and services tailored to each customer segment to enhance overall customer satisfaction.

10. **Decision-Making Support Dashboard Module:** Develop a dashboard that presents data-driven insights, visualizations, and key performance indicators to support strategic decision-making.

3.3 ALGORITHM

K-Means Clustering

- K Means algorithm in an iterative algorithm that tries to partition the dataset into K predefined distinct non overlapping sub groups which are called as cluster.
- Here K is the total no of clusters.
- Every point belongs to only one cluster.
- Clusters cannot overlap.

Steps of Algorithm

- Arbitrarily choose k objects from D as the initial cluster centers.
- Repeat.
- Assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster.
- Update the cluster means, i.e. calculate the mean value of the objects for each cluster.
- Until no change.

Elbow Method

The elbow method is based on the observation that increasing the number of clusters can help to reduce the sum of within-cluster variance of each cluster. To define the optimal clusters, Firstly, we use the clustering algorithm for various values of k. This is done by ranging k from 1 to 10 clusters. Then we calculate the total intra-cluster sum of square. Then, we proceed to plot intra-cluster sum of square based on the number of clusters. The plot denotes the approximate number of clusters required in our model. The optimum clusters can be found from the graph where there is a bend in the graph.

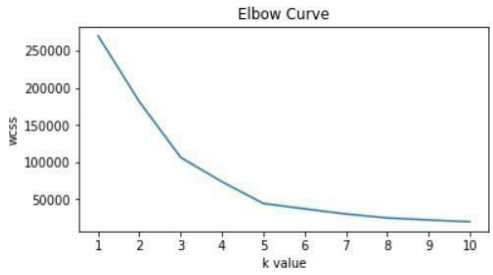

Figure 3.3 Elbow method

3.4 SYSTEM REQUIREMENTS

3.4.1 Hardware requirements:

3.4.1.1 Processor: Intel Core i5 or equivalent.

3.4.1.2 Memory (RAM): Minimum 8 GB of RAM.

3.4.1.3 Storage: Adequate storage space to store the code, datasets, and any additional resources.

3.4.2 Software requirements:

3.4.2.1 Software Requirements:

3.4.2.2 Jupyter Notebook

3.4.2.3 Anaconda Navig

3.4.2.4 Python 3.8

3.4.2.5 Python Libraries:

3.4.2.5.1.1 numpy>= 1.16.0

3.4.2.5.1.2 matplotlib >= 3.0.0

3.4.2.5.1.3 keras >= 2.4.0

.

CHAPTER – IV

IMPLEMENTATION

CHAPTER 4

IMPLEMENTATION

4.1 SAMPLE CODE

Importing the dependencies

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

Data Collection and Analysis

```
customer_data= pd.read_csv("Mall_Customers.csv")
customer_data.head()
customer_data.shape
customer_data.columns
customer_data.info()
customer_data.isnull().sum()
```

Choosing the Annual Income Column and Spending Score column

```
X=customer_data.iloc[:,[3,4]].values
print(X)
```

Choosing the number of Clusters

WCSS -->Within Cluster Sum of Square

Finding wcss value for different number of clusters

```
wcss = []
```

```
for i in range(1,11):
    kmeans = KMeans()
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
sns.set()
plt.plot(range(1,11), wcss)
plt.title('The Elbow Point Graph')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.show()
```

Training the k-Means Clustering Model

```
kmeans = KMeans(n_clusters=5, init='k-means++', random_state=0)
Y = kmeans.fit_predict(X)
print(Y)
```

Visualizing the cluster

```
plt.figure(figsize=(8,8))
plt.scatter(X[Y==0,0], X[Y==0,1], s=50, c='green', label='Cluster 1')
plt.scatter(X[Y==1,0], X[Y==1,1], s=50, c='red', label='Cluster 2')
plt.scatter(X[Y==2,0], X[Y==2,1], s=50, c='yellow', label='Cluster 3')
plt.scatter(X[Y==3,0], X[Y==3,1], s=50, c='violet', label='Cluster 4')
plt.scatter(X[Y==4,0], X[Y==4,1], s=50, c='blue', label='Cluster 5')
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], s=100, c='cyan',
label='Centroids')
plt.title('Customer Groups')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.show()
```

CHAPTER – V

RESULTS

CHAPTER 5

RESULTS

The customer segmentation analysis using clustering has proven to be a powerful strategy for improving sales and driving business growth. This study has demonstrated the significance of understanding the diverse needs and behaviours of customers, and how the application of clustering techniques can provide actionable insights to tailor marketing efforts effectively. The results highlight the value of personalized strategies and optimized resource allocation based on the identified customer segments.

Figure 5.1 Results

CONCLUSION

So we concluded that the,

- The Highest income , high spending can be target these type of customers as they earn more money and spend as much as they want.
- Highest income, low spending can target these types of customers by asking feedback and advertising the product in a better way.
- Average income, Average spending may or may not be beneficial to the mall owners of this type of customers.
- Low income, High spending can target these types of customers by providing them with low-cost EMI's etc.
- Low income, Low spending don't target these types of customers because they earn a bit and spend some amount of money.

So high income, high spending are the most beneficial ones to the mall owners which increases the owner's business. (Cluster 1)

REFERENCES

[1] Kumar, V., & Aggarwal, R. (2019). Customer Segmentation: A Review. *Journal of Marketing*, 83(3), 1- 15.

[2] Hossain, M. A., & Uddin, M. S. (2016). Customer Segmentation Based on Purchasing Behavior Using K- Means Clustering. *International Journal of Computer Applications*, 149(1), 14-18.

[3] Fadaly, E., et al. (2018). An Empirical Analysis of Customer Segmentation Strategies Using Clustering. *International Journal of Information Management*, 39, 8-18.

[4] Patil, A. J., & Nair, P. R.(2021). Enhancing Customer Segmentation with Machine Learning. *Journal of Business Research*, 134, 139-149.

[5] Kumar, G. S., & Kumar, P. S. (2018). Customer Segmentation for E-commerce: A Comparative Study of Clustering Algorithms. *Journal of Electronic Commerce Research*, 19(3), 227-241.

[6] 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.

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

[8] A. Ansari and A. Riasi, "Taxonomy of marketing strategies using bank customers clustering", InternationalJournal of Business and Management, vol. 11, no. 7, pp. 106-119, 2016.

[9] 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.

[10] 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.

[11] J. Lee and S. Park, "Intelligent profitable customers segmentation system based on business intelligence tools", Expert Systems with Applications, vol. 29, no. 1, pp. 145-152, 2005.

[12] D. A. Kandeil, A. A. Saad and S. M. Youssef, "A two-phase clustering analysis for B2B customer segmentation", in International Conference on Intelligent Networking and Collaborative Systems, Salerno, 2014, pp. 221-228.