

Department of Computer Science and Engineering

Customer Segmentation Engine: An Intelligent Clustering-Based System for Automated Market Grouping Using Machine Learning

**Mrs. Divya M, Department of
CSE
Rajalakshmi Engineering
College
Chennai, India Chennai, India
divya.m@rajalakshmi.edu.in**

SAKTHI SHALINI R 220701241

Problem Statement and Motivation

- **Problem Statement**

In today's competitive market, businesses often struggle to understand the diverse needs and behaviors of their customers. Customer segmentation is crucial for targeting the right audience with personalized products, services, and marketing campaigns. However, traditional methods of customer segmentation, such as manually grouping customers based on basic criteria like age or income, are time-consuming, limited, and prone to errors. These methods fail to capture the deeper, more complex patterns in customer behavior, leading to missed opportunities in marketing strategies and customer retention efforts.

The problem is that businesses lack an automated, scalable system that can intelligently group customers based on meaningful attributes without relying on predefined rules or assumptions. Without such a system, businesses are unable to fully understand their customers and make data-driven decisions that can improve customer satisfaction and increase revenue.

Motivation:- The motivation behind this project is to provide businesses with a more efficient and intelligent way to segment their customers. Using machine learning, specifically KMeans clustering, businesses can automatically group customers based on their actual behaviors and preferences, not just on basic data points. This approach allows businesses to uncover hidden patterns that are not obvious through manual analysis. By implementing a data-driven, machine learning-based system, companies can make smarter decisions in marketing, product development, and customer support. The system can continuously adapt as new customer data becomes available, providing up-to-date insights. Additionally, customer segmentation can help businesses better target high-value customers, optimize advertising campaigns, and personalize services, ultimately leading to improved customer satisfaction and higher profits.

Existing System

- An existing project that closely matches this idea is a customer segmentation system using machine learning, especially the KMeans clustering algorithm. This project is commonly built using datasets like the Mall Customer Segmentation dataset, which includes key customer details such as Gender, Age, Annual Income, and Spending Score. The goal of the project is to automatically group customers into different segments based on their similarities. For example, one group might be young people with high spending habits, another might be older customers with lower spending, and so on. These groups help businesses understand different customer types and plan personalized marketing strategies. The process starts with data cleaning, where unnecessary columns like CustomerID are removed. Then, the remaining features are prepared for analysis. To decide how many groups to create, the Elbow Method is used. This method shows a graph to help pick the best number of customer segments. Once that is done, the KMeans algorithm is applied to divide the customers into meaningful clusters. After clustering, the results are shown using colorful 2D and 3D scatter plots, which make it easier to see how customers are grouped. Each cluster or group shows a different type of customer based on income and spending habits. This existing system shows how machine learning can help businesses move away from manual grouping and use smart, data-driven methods to understand their customers better. It is easy to apply, scalable for different types of businesses, and useful for improving sales, promotions, and customer satisfaction.

Objectives

- **Automate Customer Grouping:** Use machine learning to automatically segment customers based on behavior (e.g., spending habits, income).
- **Improve Marketing Strategies:** Help businesses target specific customer groups with personalized marketing campaigns.
- **Enhance Decision-Making:** Provide data-driven insights to assist businesses in making smarter decisions for products, promotions, and customer engagement.
- **Use KMeans Clustering:** Apply the KMeans algorithm to classify customers into meaningful segments based on key features.
- **Visualize Customer Segments:** Create 2D and 3D visualizations to help businesses easily interpret and act on the segmentation results.
- **Scalability and Adaptability:** Ensure the system can handle large datasets and adapt as new customer data is added.
- **Optimize Customer Insights:** Uncover hidden patterns in customer behavior to improve service offerings and customer satisfaction.
- **Enhance Customer Retention:** Identify customer segments that are at risk of leaving and create targeted retention strategies based on their behavior.
- **Maximize Revenue:** Help businesses focus on high-value customers by segmenting them according to spending habits, enabling targeted upselling and cross-selling opportunities.
- **Reduce Marketing Costs:** By targeting specific customer groups effectively, reduce the overall cost of marketing campaigns while increasing conversion rates.
- **Support Business Growth:** Enable businesses to scale their marketing and customer engagement efforts as they expand, ensuring that the segmentation process can handle increasing amounts of data efficiently.

Abstract

In this project, we create an intelligent system to automatically group customers based on their behaviors using machine learning. By applying the KMeans clustering algorithm, customers are divided into segments based on key factors like Annual Income and Spending Score. This helps businesses find hidden patterns in customer behavior that might be missed with manual methods.

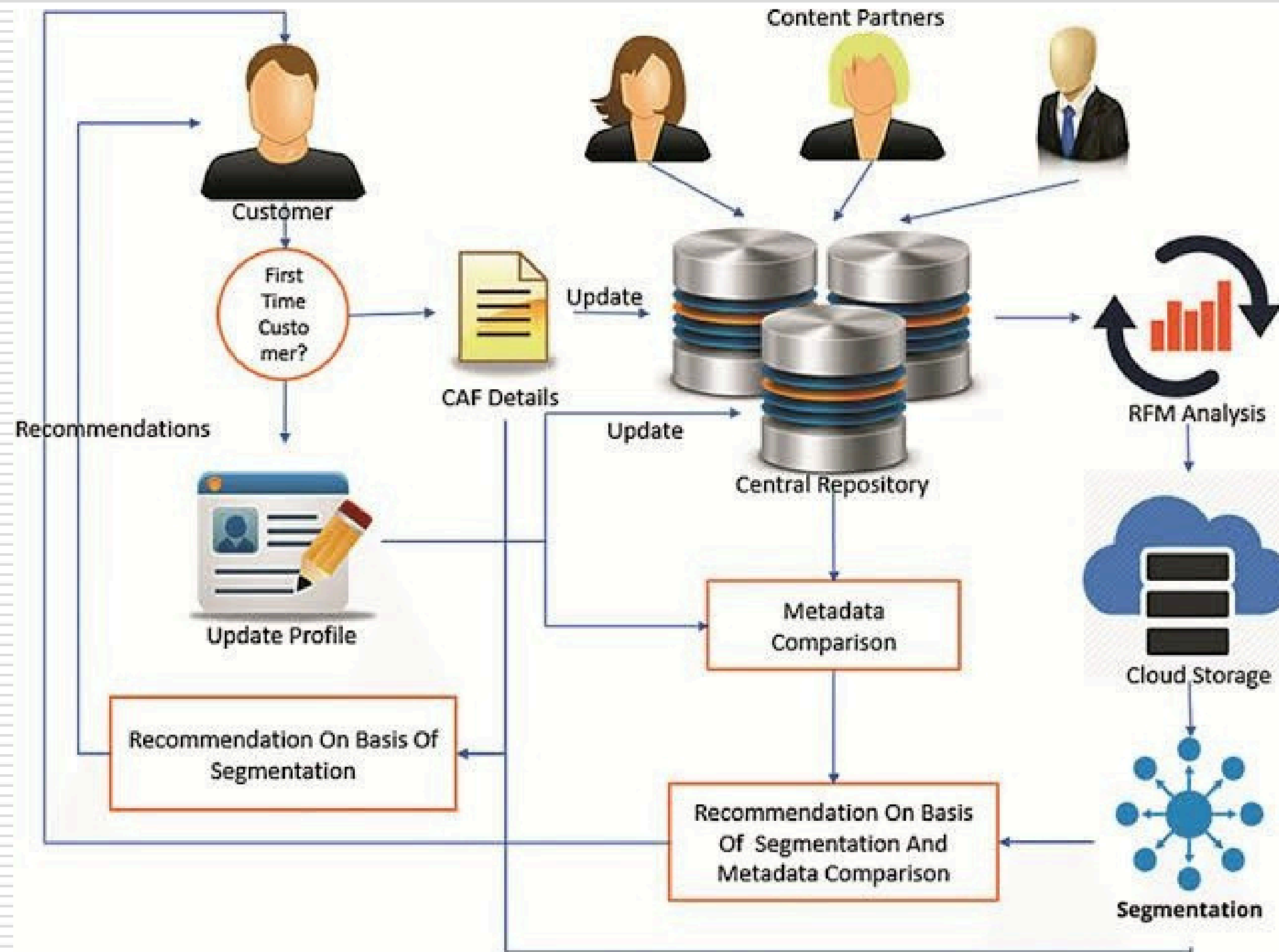
The project starts with data cleaning and selecting the most relevant features, like Gender, Age, Income, and Spending Score. The Elbow Method is used to determine the best number of customer groups. After that, KMeans is applied to classify customers into different segments. The results are then shown using 2D and 3D scatter plots for easy analysis and decision-making.

This approach helps businesses understand which customers are high-value, price-sensitive, or have other key traits, allowing them to create more targeted marketing strategies. The system is flexible and can be applied to different business needs, making it a powerful tool for customer analysis and improving marketing efforts. Furthermore, this system can adapt to new data over time. As businesses collect more customer information, the system can update and refine the customer segments. This means that companies can continuously improve their marketing efforts and stay aligned with evolving customer behaviors. In conclusion, this machine learning-based customer segmentation system provides businesses with an automated, scalable solution for understanding their customers. It simplifies the process of identifying key customer groups and enhances the overall effectiveness of marketing strategies. By leveraging data-driven insights, businesses can not only improve customer satisfaction but also boost sales and profitability.

Proposed System

The proposed system is a smart, machine learning–based customer segmentation tool that helps businesses understand their customers better. It works by taking customer data such as age, annual income, and spending score, and then grouping similar customers together using the KMeans clustering algorithm. This method automatically finds patterns in the data that may not be visible through normal analysis. For example, it can group customers into categories like young high spenders, low-income careful buyers, or middle-aged average spenders. The system starts by cleaning the data and removing any unnecessary columns like CustomerID. It then uses a method called the Elbow Method to decide how many groups (or clusters) to create. After that, it applies clustering to different combinations of features — like age and spending score, income and spending score, and all three together — to find the most useful groupings. Once the clusters are created, the system shows the results using colorful 2D and 3D graphs so it's easy to see how customers are grouped. This helps businesses quickly understand who their top spenders are, who might need different offers, and how to personalize marketing campaigns for each group. The system is efficient, easy to scale, and can be updated with new data to improve results over time. It saves businesses time and effort, while making their marketing strategies more effective.

System Architecture



List of Modules

Data Handling & Analysis

- `numpy` – for numerical operations
- `pandas` – for loading and manipulating data

Data Visualization

- `matplotlib.pyplot` – for basic plots and charts
- `seaborn` – for advanced and styled data visualizations
- `mpl_toolkits.mplot3d` – for 3D scatter plots

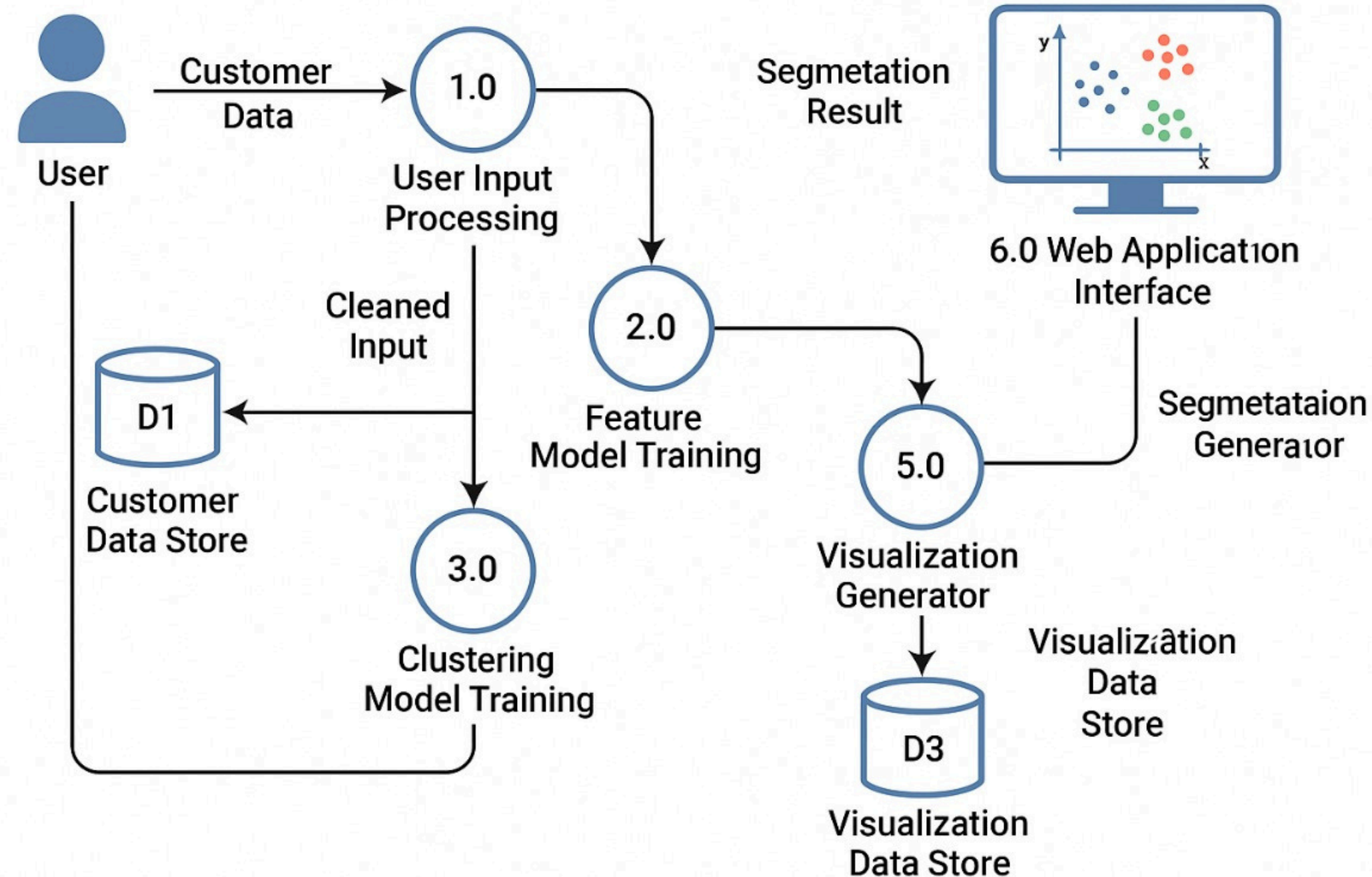
File Upload (Google Colab specific)

- `google.colab.files` – for uploading datasets to Colab

Machine Learning

- `sklearn.cluster.KMeans` – for clustering customers using the KMeans algorithm

Functional Description for each modules with DFD and Activity Diagram



Customer Segmentation System Using KMeans

Implementation & Results of Module

	CustomerID	Gender	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

	CustomerID	Age	Annual Income (k\$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

Conclusion & Future Work

Conclusion This project successfully used machine learning, specifically the KMeans clustering algorithm, to group customers into different segments based on their income, age, and spending behavior. By analyzing the data, we were able to identify patterns that help businesses understand their customers better—like who spends more, who is budget-conscious, and how age or income affects shopping habits. This kind of automated segmentation is much faster and more accurate than doing it manually, and it helps companies make smarter marketing decisions.

Future Work

- Use more advanced algorithms (like DBSCAN or hierarchical clustering) to improve the accuracy of segmentation.
- Add more customer features such as location, online activity, or product preferences for deeper analysis.
- Build a web dashboard or app so businesses can upload data and get segmentation results instantly.
- Integrate real-time data updates, so customer segments automatically refresh as new data comes in.
- Apply deep learning models to handle more complex and large datasets in the future.

References

1. MacQueen, J. (1967). *Some Methods for Classification and Analysis of Multivariate Observations*. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, No. 14, pp. 281–297). University of California Press. – Introduced the K-Means clustering algorithm used in this project.
2. Scikit-learn Developers. (2024). *Clustering: KMeans*. Scikit-learn Documentation. Retrieved from: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
Official documentation for the KMeans implementation used in the project.



Thank You