

Customer Segmentation Using Unsupervised Learning



This presentation explores customer segmentation using supervised learning techniques. It covers the project's background, objectives, methodology, and results. We aim to provide a clear understanding of how AI can enhance marketing strategies.

Submitted by

Kunal Prajapati (202410116100108)
Doulat Biswal (202410116100070)
Krishna (2024101161000103)
Harsh Aggarwal (202410116100081)

Submitted to

Mr. Apoorv Jain
Assistant Professor
MCA Department
KIET Group of Institutions

Introduction to Customer Segmentation

- Customer segmentation divides customers into groups by shared traits.
- Enhances understanding of customer needs and targeting.
- Uses Unsupervised Learning: KMeans Clustering to find patterns automatically.





Background & Motivation

■ **Limitations of Manual Segmentation**
Inefficient for large datasets

■ **Value of AI**
Discovers hidden patterns in big data

■ **Clustering Benefits**
Groups customers without predefined labels

■ **Business Impact**
Personalised marketing and services

Project Objectives

Segment Customers

Using KMeans Clustering

Group on Key Features

Income, Age, Spending behaviour

Visualize Clusters

Via Principal Component Analysis (PCA)

Business Insights

Analyse cluster profiles

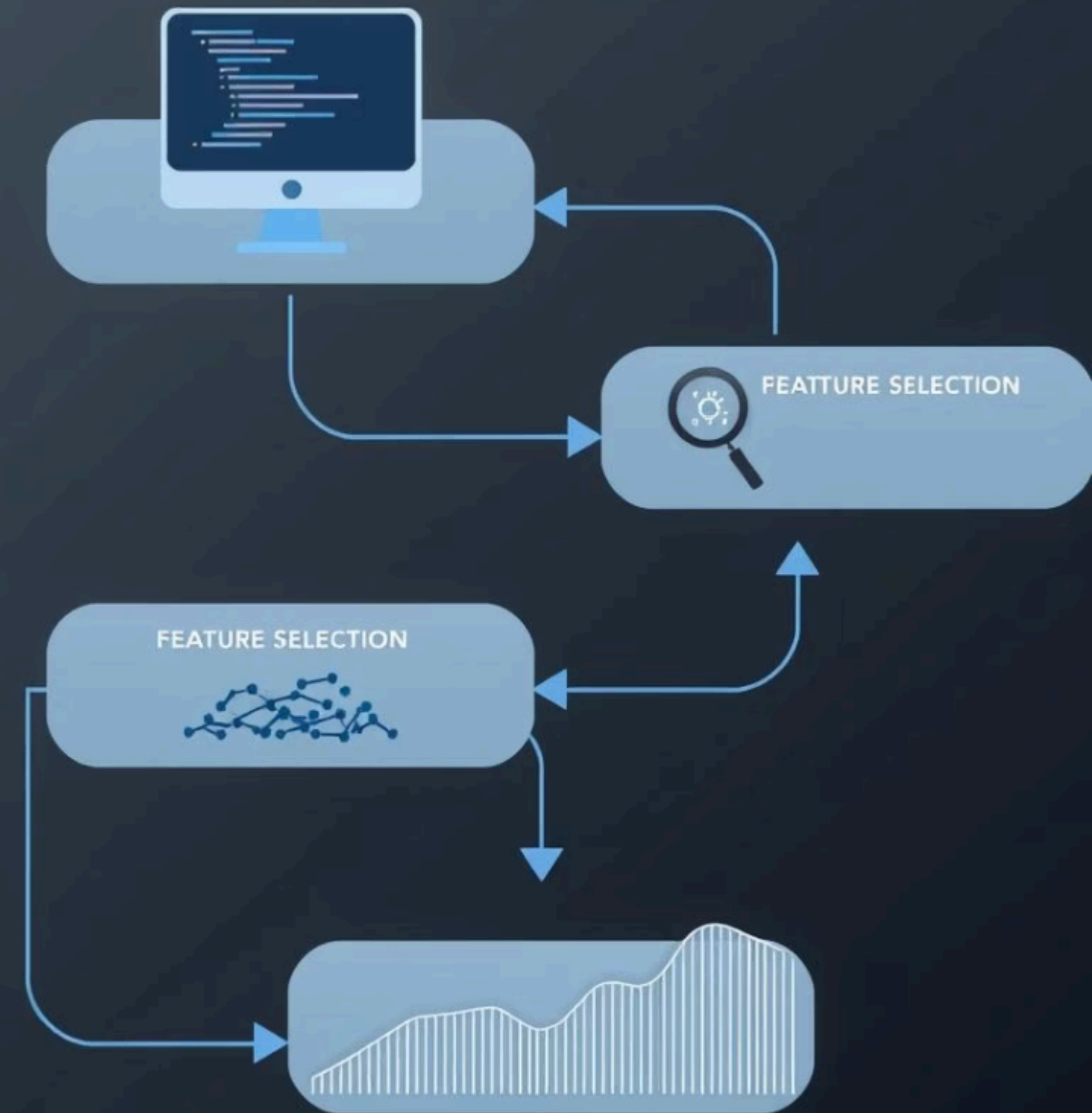


Dataset Overview

- Source: Kaggle - Marketing Campaign
- Demographics: Age, Education, Marital Status
- Spending on Wine, Fruits, Meat, etc.
- Total_Spend engineered by summing expenditures



Data Science



Methodology



Preprocessing

- Drop unneeded columns
- Handle missing values
- Encode categorical data

Feature Selection

Choose Income, Recency, Education, Total Spend

KMeans Clustering

Apply 3 clusters, assign labels

Visualization

PCA to 2D, plot clusters

Code Highlights

Label Encoding

For categorical variables

StandardScaler

Data scaling for clustering

KMeans Model

n_clusters set to 3

PCA

Two-dimensional visualization

Cluster Analysis

Grouped data to profile clusters

```
# Step 1: Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report

# Step 2: Load the dataset
df = pd.read_csv('/content/marketing_campaign.csv', sep=';')
print("First 5 rows:\n", df.head())

# Step 3: Drop unnecessary columns (like ID, date fields)
df.drop(['ID', 'Dt_Customer', 'Z_CostContact', 'Z_Revenue'], axis=1, inplace=True)

# Step 4: Handle missing values
df.dropna(inplace=True)

# Step 5: Encode categorical variables
categorical_cols = ['Education', 'Marital_Status']
le = LabelEncoder()
for col in categorical_cols:
    df[col] = le.fit_transform(df[col])

# Step 6: Create a target variable
# Let's define 'High-Value Customers' based on total spend
df['Total_Spend'] = df[['MntWines', 'MntFruits', 'MntMeatProducts', 'MntSeafood', 'MntGroceries', 'MntTotal']].sum(axis=1)
```

Output Visualization



Cluster Summary:					
	Income	Recency	Education	Marital_Status	Kidhome \
Cluster					
0	34438.377551	48.788776	2.222449	3.702041	0.910204
1	77690.590994	49.814259	2.454034	3.776735	0.041276
2	57782.665718	48.716927	2.587482	3.721195	0.092461

	Teenhome	Total_Spend
Cluster		
0	0.362245	137.813265
1	0.035647	1387.679174
2	1.061166	669.401138

Conclusion and Future Work

Successful Segmentation

Three distinct clusters identified

PCA Visualization

Clear cluster separation observed

Improvement Ideas

Explore hierarchical clustering & DBSCAN



Thank You!