Customer Segmentation Using Supervised 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 (20241011610018)
Doulat Biswal (20241011610070)
Krishna (2024101161000103)
Harsh Aggarwal (20241011610081)

Submitted to

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

Introduction

Customer segmentation involves dividing customers into groups. Groups are based on similar characteristics. This allows businesses to better understand their customers. They can create targeted marketing strategies.

Targeted strategies

Segmentation creates customer focused marketing.

Understand customers

Segmentation builds deeper insights.

Background & Motivation

Traditional segmentation methods are manual and inefficient. AI and supervised learning automate the segmentation process. They offer faster and more accurate results.

Automated

Fast

Accurate

Project Objectives

This project focuses on implementing a supervised learning model. The goal is to segment customers effectively. We aim to classify customers into value-based segments.

Automate segmentation

Apply in real-world scenarios.

Visualize important features

Understand key segmentation drivers.

Dataset Overview

The dataset used is the Marketing Campaign Dataset from Kaggle. It contains demographic and behavioral attributes. These include age, education, income, and marital status.

Demographics

- Age

- Education

- Income

Behavioural Data

- Spending on products

- Number of children

- Marital Status

Methodology

The methodology includes data preprocessing, feature engineering, and model selection. We used a Decision Tree Classifier for its interpretability.

Preprocessing

Feature Engineering

Model Selection

Model Evaluation

Code Highlights

The code involves data loading, cleaning, and label encoding. We created a target variable by computing total spending. The DecisionTreeClassifier was used from sklearn.

- **Data Loading**

    - **Model Training**

        - **Evaluation**

```python
# 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, classificati

# 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_Revenu

# 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 spe
df['Total_Spend'] = df[['MntWines', 'MntFruits', 'MntMea
```

Output

```
First 5 rows:
     ID  Year_Birth  Education Marital_Status   Income  Kidhome  Teenhome  \
0  5524        1957  Graduation         Single  58138.0        0         0
1  2174        1954  Graduation         Single  46344.0        1         1
2  4141        1965  Graduation       Together  71613.0        0         0
3  6182        1984  Graduation       Together  26646.0        1         0
4  5324        1981         PhD        Married  58293.0        1         0

  Dt_Customer  Recency  MntWines  ...  NumWebVisitsMonth  AcceptedCmp3  \
0  04-09-2012       58       635  ...                  7             0
1  08-03-2014       38        11  ...                  5             0
2  21-08-2013       26       426  ...                  4             0
3  10-02-2014       26        11  ...                  6             0
4  19-01-2014       94       173  ...                  5             0

   AcceptedCmp4  AcceptedCmp5  AcceptedCmp1  AcceptedCmp2  Complain  \
0             0             0             0             0         0
1             0             0             0             0         0
2             0             0             0             0         0
3             0             0             0             0         0
4             0             0             0             0         0

   Z_CostContact  Z_Revenue  Response
0              3         11         1
1              3         11         0
2              3         11         0
3              3         11         0
4              3         11         0
```


Top 10 Feature Importances
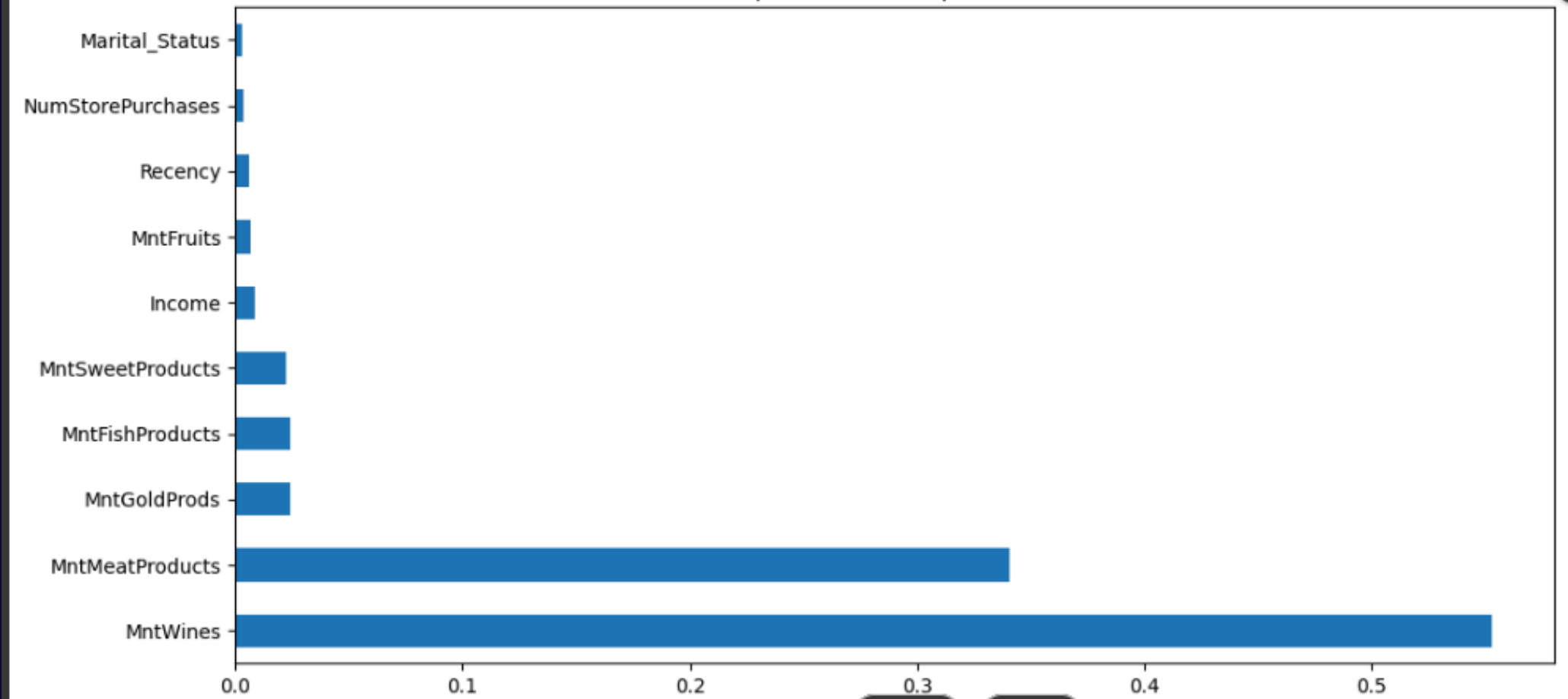
```
Classification Report:
              precision    recall  f1-score   support

        High       0.95      0.93      0.94       221
         Low       0.97      1.00      0.98       224
      Medium       0.93      0.92      0.92       220

    accuracy                           0.95       665
   macro avg       0.95      0.95      0.95       665
weighted avg       0.95      0.95      0.95       665
```

Conclusion

The model successfully segmented customers using supervised learning. Key features included income, age, and product spending. The project demonstrates the value of AI in marketing.

2 Key Role

1

Good Accuracy

Marketing Value

3

Thank You

Thank you for your time and attention. Are there any questions?