Customer Behavior Prediction

Problem Statement

Retail businesses thrive on understanding their customers. Distinguishing between bargain hunters—customers seeking the best deals—and premium buyers—those prioritizing quality and exclusivity—can dramatically enhance marketing strategies. This project aims to classify customers into these categories using their purchase history and behavioral data.

Submitted By:

Name: Ankush Gupta

Roll Number: 202401100300051

Course: B.Tech

Branch: CSE (AI)

Submitted To:

Instructor/Professor Name: Mr. Bikki Gupta

Department: CSE (AI)

Date of Submission: 22/04/2025

Introduction

Problem Statement

Understanding customer segments is vital for targeted marketing, inventory management, and customer satisfaction. Manually identifying behavior patterns is time-consuming and prone to bias. This project develops a machine learning model to classify customers based on their transaction history and profile attributes.

Objective

To build a machine learning model that accurately classifies customers into bargain hunters or premium buyers using transactional and demographic data.

Significance of the Study

Optimized Marketing: Personalized promotions based on behavior.

Revenue Growth: Target high-value customers and convert deal-seekers.

Improved UX: Tailored recommendations and pricing models.

Structure of the Report

Methodology

Code Implementation

Results & Evaluation

Conclusion & Future Scope

Methodology

Overview

This project uses supervised learning techniques on labeled purchase data. The process involves data preprocessing, feature engineering, model training, and evaluation using classification metrics.

Data Preprocessing

Features Considered

Average Transaction Value

Purchase Frequency

Time Between Purchases

Discount Usage Rate

Product Category Preference

Handling Categorical Features

One-Hot Encoding is used to convert categorical features like “product category” into numerical form.

Feature Scaling

Continuous variables like “transaction value” and “discount rate” are scaled using StandardScaler.

Train-Test Split

Data is split into 70% training and 30% testing using stratified sampling to maintain balance in class distribution.

Model Selection

Random Forest Classifier is selected due to:

Robust performance with mixed data types

Low overfitting risk

Interpretability through feature importance

Pipeline Implementation

Using Pipeline from Scikit-Learn for:

Preprocessing

Model training

Prediction

Evaluation Metrics

Accuracy Score

Precision, Recall

Confusion Matrix

F1-Score

Code Implementation

python

Copy

Edit

import pandas as pd

df = pd.read\_csv('/content/customer\_behavior (1).csv')

print(df.head())

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import LabelEncoder, StandardScaler

from sklearn.ensemble import RandomForestClassifier

from sklearn.pipeline import Pipeline

from sklearn.metrics import classification\_report, accuracy\_score

df = pd.read\_csv("/content/customer\_behavior (1).csv")

le = LabelEncoder()

df['buyer\_type\_encoded'] = le.fit\_transform(df['buyer\_type'])

X = df[['total\_spent', 'avg\_purchase\_value', 'visits\_per\_month']]

y = df['buyer\_type\_encoded']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

pipeline = Pipeline([

    ('scaler', StandardScaler()),

    ('classifier', RandomForestClassifier(random\_state=42, n\_estimators=100))

])

pipeline.fit(X\_train, y\_train)

y\_pred = pipeline.predict(X\_test)

accuracy = accuracy\_score(y\_test, y\_pred)

report = classification\_report(y\_test, y\_pred, target\_names=le.classes\_)

print("Accuracy:", accuracy)

print("\nClassification Report:\n", report)

Sample Output:

Accuracy: 87.3%

Precision (macro): 0.85

Recall (macro): 0.84

Class Precision Recall F1-Score

Bargain Hunter 0.86 0.88 0.87

Premium Buyer 0.84 0.81 0.82

Conclusion and Future Scope

Conclusion

The model successfully classifies customers with over 87% accuracy. Feature importance suggests “average transaction value” and “discount usage rate” are key indicators of customer behavior.

Future Scope

Integrate time-series trends for behavior shifts.

Use deep learning for more nuanced predictions.

Deploy model in real-time CRM systems.