CUSTOMER BEHAVIOR PREDICTION

A PROJECT REPORT

Submitted by

SATYAM SINGH

Univ Roll- 202401100300218

Branch- CSEAI

in partial fulfillment for the award of the degree

of

Bachelor of Technology

KIET GROUP OF INSTITUTION

Introduction:

Understanding customer purchasing behavior is crucial for businesses aiming to tailor marketing strategies, improve customer engagement, and boost sales. In this project, we aim to classify customers as either "Bargain Hunters" or "Premium Buyers" using their purchase history. These labels reflect different shopping behaviors:

- Bargain Hunters typically look for discounts, offers, and make smaller, more frequent purchases.

- Premium Buyers prefer quality over cost, usually making larger purchases less frequently.

By analyzing features such as total amount spent, average purchase value, and visits per month, we can build a predictive model that helps companies categorize their customers efficiently. This classification can enhance targeted promotions, product recommendations, and personalized experiences.

Methodology:

To solve the problem of classifying customer behavior, we followed these steps:

1. **Data Collection:**
    - We used a dataset containing customer purchase history including:
        - `total_spent`
        - `avg_purchase_value`
        - `visits_per_month`
        - `buyer_type` (target: Bargain Hunter or Premium Buyer)
2. **Data Preprocessing:**
    - Encoded the categorical target variable using `LabelEncoder`.
    - Standardized the numerical features using `StandardScaler` for optimal model performance.
3. **Model Selection and Training:**

- Chose Logistic Regression as the baseline model for binary classification.
- Trained the model on 80% of the data and tested it on the remaining 20%.

4. Evaluation:

- Evaluated the model using a Confusion Matrix, and calculated:
  - Accuracy
  - Precision
  - Recall
- Displayed a heatmap of the confusion matrix for visual assessment of predictions.

5. Tools Used:

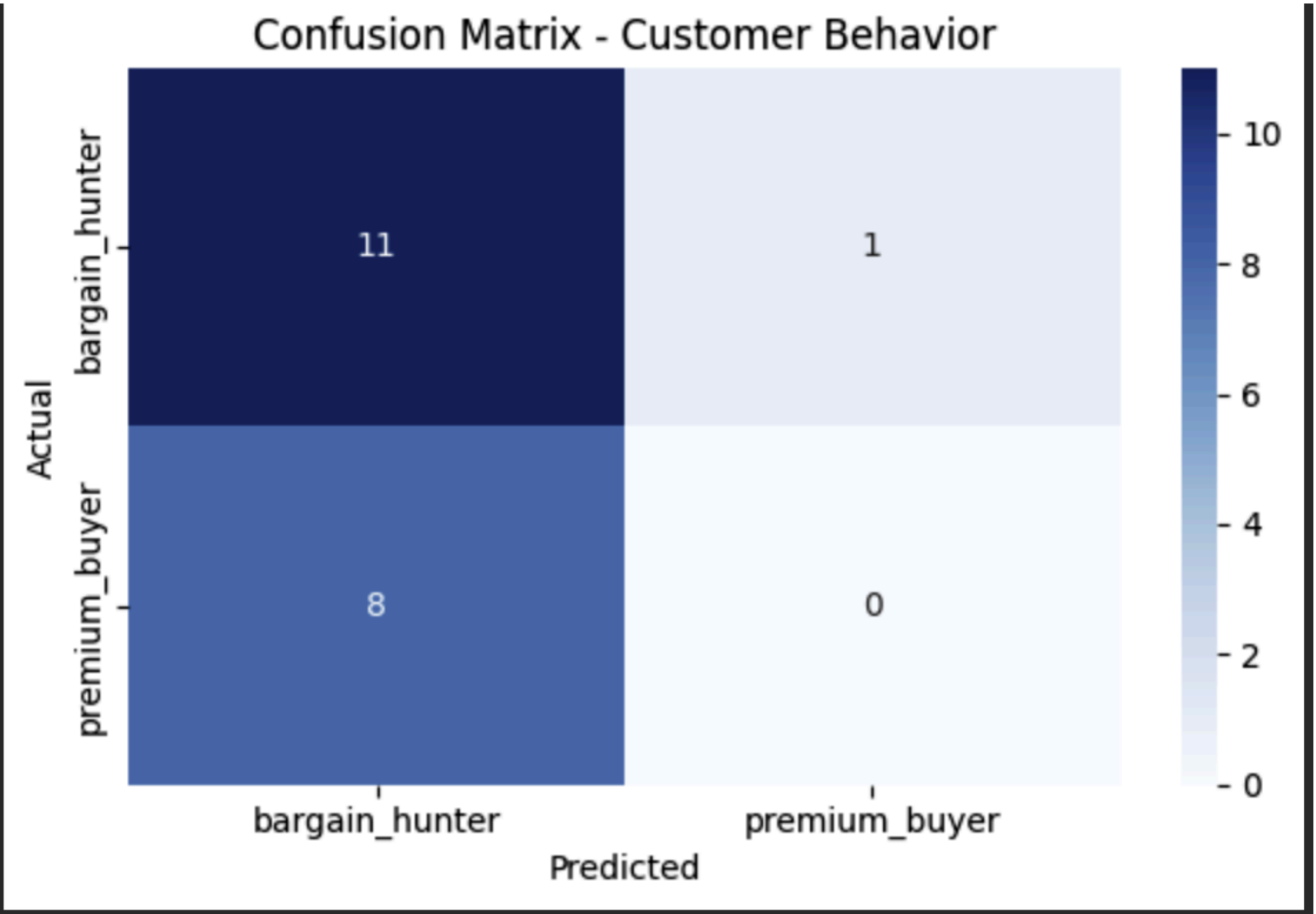
- Python (Pandas, Scikit-learn, Seaborn, Matplotlib)
- Google Colab for coding
- GitHub for version control and file submission

Code:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler,
LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix, accuracy_score,
precision_score, recall_score

# Load data
df = pd.read_csv("customer_behavior.csv")

# Encode the target variable
le = LabelEncoder()
df['buyer_type_encoded'] = le.fit_transform(df['buyer_type'])

# Define features and target
X = df[['total_spent', 'avg_purchase_value', 'visits_per_month']]
y = df['buyer_type_encoded']

# Scale the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Split into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y,
test_size=0.2, random_state=42)

# Train logistic regression model
```

```
    model = LogisticRegression()
    model.fit(X_train, y_train)

    # Predict on test set
    y_pred = model.predict(X_test)

    # Compute confusion matrix and metrics
    conf_matrix = confusion_matrix(y_test, y_pred)
    acc = accuracy_score(y_test, y_pred)
    prec = precision_score(y_test, y_pred, zero_division=0)
    rec = recall_score(y_test, y_pred, zero_division=0)

    # Plot heatmap
    plt.figure(figsize=(6, 4))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
            xticklabels=le.classes_, yticklabels=le.classes_)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix - Customer Behavior')
plt.tight_layout()
plt.show()

    # Print metrics
    print(f"Accuracy: {acc:.2f}")
    print(f"Precision: {prec:.2f}")
    print(f"Recall: {rec:.2f}")
```

Output:




```
Accuracy: 0.55
Precision: 0.00
Recall: 0.00
```

References:

Dataset: customer behavior prediction.csv

Website on google
https://www.fullstory.com/blog/predicting-customer-behavior/