PREDICTING CUSTOMER PURCHASE BEHAVIOR

A Machine Learning Approach to
Estimating Purchase Likelihood

INTRODUCTION

Predicting customer purchase behavior is crucial for optimizing marketing strategies, improving customer retention, and increasing sales. By analyzing customer demographics, purchase history, and engagement, a predictive model can estimate the likelihood of a customer making a purchase.

DATA PREPROCESSING

- Split dataset into features (X) and target (y).
- Apply standard techniques such as label encoding and handling missing values if any.
- Split the data into training (70%) and testing sets (30%).
- Feature scaling was not necessary for the Random Forest model.

MODEL SELECTION AND TUNING

Model: Random Forest Classifier

Handles both categorical and numerical data. Can model non-linear relationships. Less prone to overfitting due to ensemble nature.

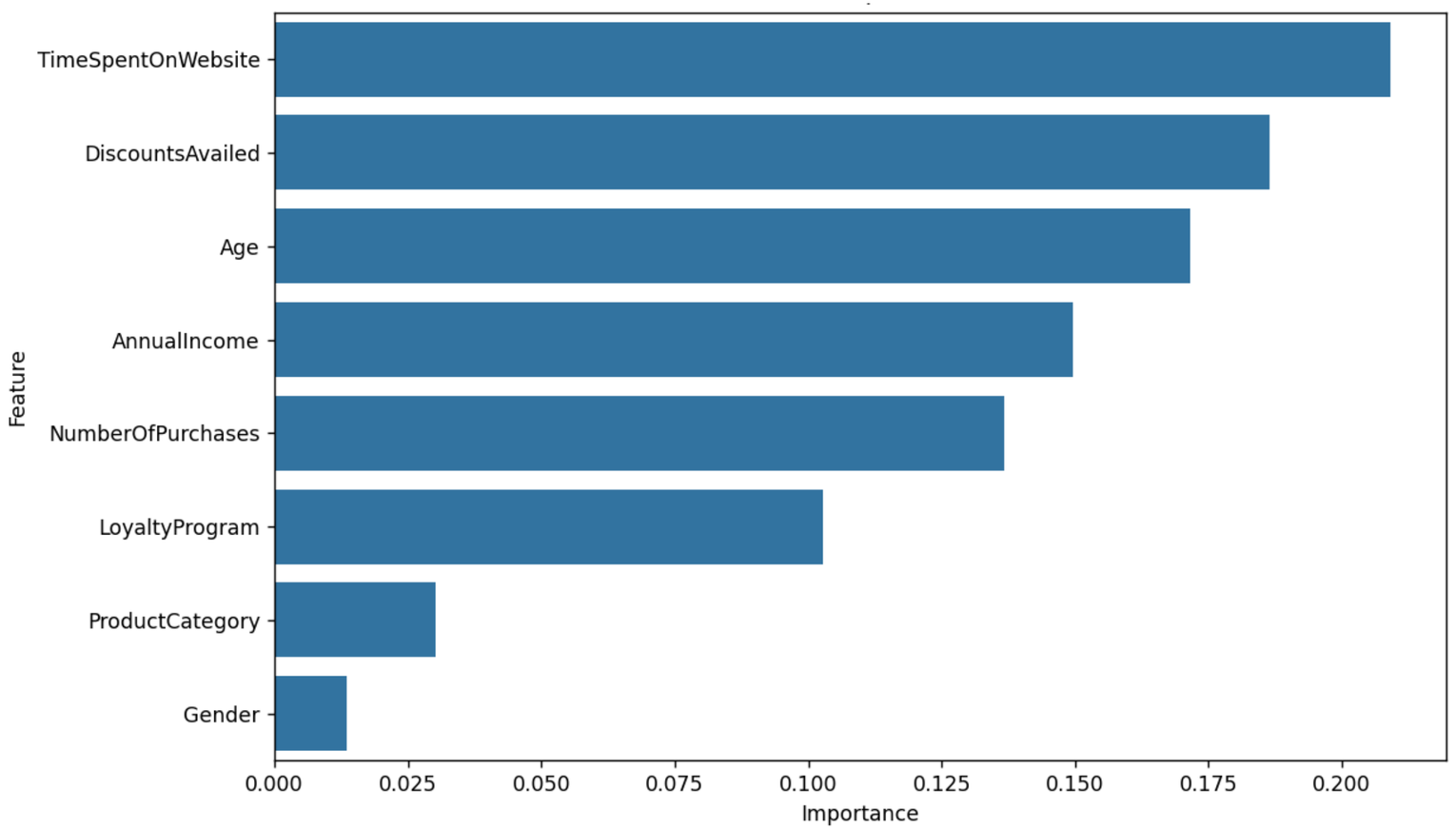
Performed GridSearchCV to optimize hyperparameters (n_estimators, max_depth, min_samples_split, min_samples_leaf).

MODEL EVALUATION METRICS

```
Cross-Validation Scores: [0.93666667 0.94666667 0.94666667 0.92333333 0.94666667]
Mean CV Score: 0.9400000000000001

Accuracy Score: 0.9422222222222222

Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.98      0.95       257
           1       0.97      0.90      0.93       193

    accuracy                           0.94       450
   macro avg       0.95      0.94      0.94       450
weighted avg       0.94      0.94      0.94       450
```

Class 0 (No Purchase) and Class 1 (Purchase) metrics show balanced performance.

FEATURE IMPORTANCE ANALYSIS



MODEL DEPLOYMENT

The best model was saved using joblib as customer_purchase_model.pkl.
This model can be deployed in production to predict customer purchase behavior based on new data.

CONCLUSION

The Random Forest model achieved a high accuracy of 94.22% and provided valuable insights into the most important features influencing customer purchases. This model can be used to predict purchase behavior, thereby assisting businesses in making data-driven decisions.

Future Work includes further tuning of the model with more advanced techniques like boosting algorithms and developing a user interface to visualize predictions.

THANK YOU

Vismaya R