Predicting Online Customer's Purchase Intentions using Machine Learning

Introduction

The aim of this project was to predict whether online customers have an intention to make a purchase based on various attributes and interactions during their website sessions. Through classification and clustering techniques, the project aimed to improve understanding of customer behaviour and enhance conversion rates.

Dataset Description

The dataset used in this project is called "Online-Shoppers-Purchasing-Intention." It contains a mixture of 10 numerical and 8 categorical attributes, where the 'Revenue' attribute serves as the class label. The dataset was designed to capture user behaviour over a one-year period, avoiding biases due to specific campaigns, special days, or user profiles.

Data Preprocessing

Data preprocessing steps included handling missing values, encoding categorical variables, and preparing the dataset for modelling. Attributes such as "Administrative," "Informational," "Product Related," "Bounce Rate," "Exit Rate," and more were used to represent user behaviour and engagement metrics.

Exploratory Data Analysis (EDA)

EDA involved visualising relationships between features and the target variable ("Revenue"). Different types of pages visited, session durations, bounce rates, and other metrics were analysed to identify patterns and trends.

Model Building and Evaluation

Several machine learning models were employed to predict purchasing intention:
- Random Forest
- Decision Tree
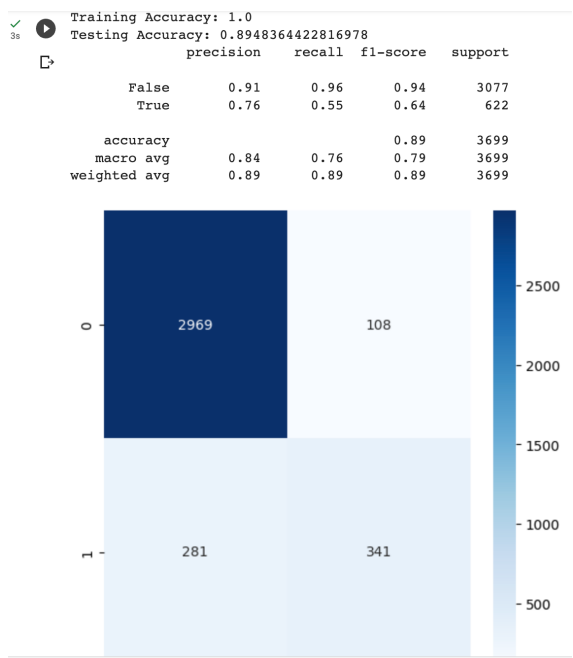- AdaBoost
- Support Vector Machine (SVM)

Each model was trained, evaluated, and tuned for hyper parameters to achieve the highest accuracy. Model evaluation included accuracy scores, confusion matrices, and classification reports.

Feature Importance Analysis

SHAP (SHapley Additive exPlanations) values were used to determine feature importance and explain the impact of various attributes on the prediction of purchasing intention. This helped in identifying the most influential factors.

Model Comparison and Selection

Random Forest consistently emerged as the most effective model for predicting purchasing intention, followed by Logistic Regression and SVM.

```
Training Accuracy: 1.0
Testing Accuracy: 0.8948364422816978
              precision    recall  f1-score   support

       False       0.91      0.96      0.94      3077
        True       0.76      0.55      0.64       622

    accuracy                           0.89      3699
   macro avg       0.84      0.76      0.79      3699
weighted avg       0.89      0.89      0.89      3699
```

Conclusion

The project successfully predicted customer's purchasing intention based on their online behaviour. The Random Forest model outperformed others, delivering the most accurate predictions. SHAP values enhanced model interpretability by revealing influential factors.

Future Recommendations

- Further feature engineering to capture more nuanced user interactions.

- Exploration of ensemble techniques and deep learning algorithms for potential performance improvement.

- Regular model updates to adapt to changing customer behaviour and website dynamics.

Project Significance

This project is significant for e-commerce platforms seeking to optimise their conversion rates. By understanding the attributes that influence purchasing intention, businesses can tailor their strategies to engage customers effectively, resulting in increased sales and revenue.

Divyanshu Yadav
divyanshuydv0002@gmail.com
Drive Folder