Online Shoppers Intention Analysis

Abeer Mohammed Alghamdi

Abstract:

The goal of this project is to use classification models to analyze the Customer's Intentions based on the transactions, duration made online in a year , And to create a model that can predict the purchasing intentions of customers.

Design:

This project is one of the T5 Data Science BootCamp requirements. Data obtained from Kaggle website. Source

Data:

This dataset consists of 18 features belonging to 12,330 shopping sessions.The 'Revenue' attribute is the target feature. The dataset is clean and there are no missing values.

Algorithms:

Feature Engineering:

1. Converting categorical attributes to ordered factor variables and are numerically encoded.
2. Normalize numerical variables of the dataset for clustering and scale for classification methods.

Models Used:

Logistic Regression. KNeighbors Classifier Decision Tree Gradient Boosting Naive Bayes Random Forest Hyperparameters used: GridSearchCV and RandomizedSearchCV

Random Forest Classifier has 91% accuracy score with and without hyperparameter optimization which is the highest accuracy score over all other models.

- Accuracy: 0.91
- F1: 0.95 No Revenue, 0.64 Revenue
- Precision: 0.92 No Revenue, 0.77 Revenue
- Recall: 0.97 No Revenue, 0.54 Revenue

Tools:

1. Numpy and Pandas for data manipulation.
2. Scikit-learn for modeling.
3. Matplotlib and Seaborn for plotting.

Communication:

The presentation slides are provided here.