



ONLINE SHOPPER'S PURCHASING INTENTION

BY
JYOTSNA ELTEPU



Objective

- ▶ In 2021, nearly \$1 in every \$5 spent on retail purchases came from digital orders.
- ▶ The goal of the project is to predict the purchasing intention of an online shopper so that the business can use targeted marketing to increase the profits.



Data Source

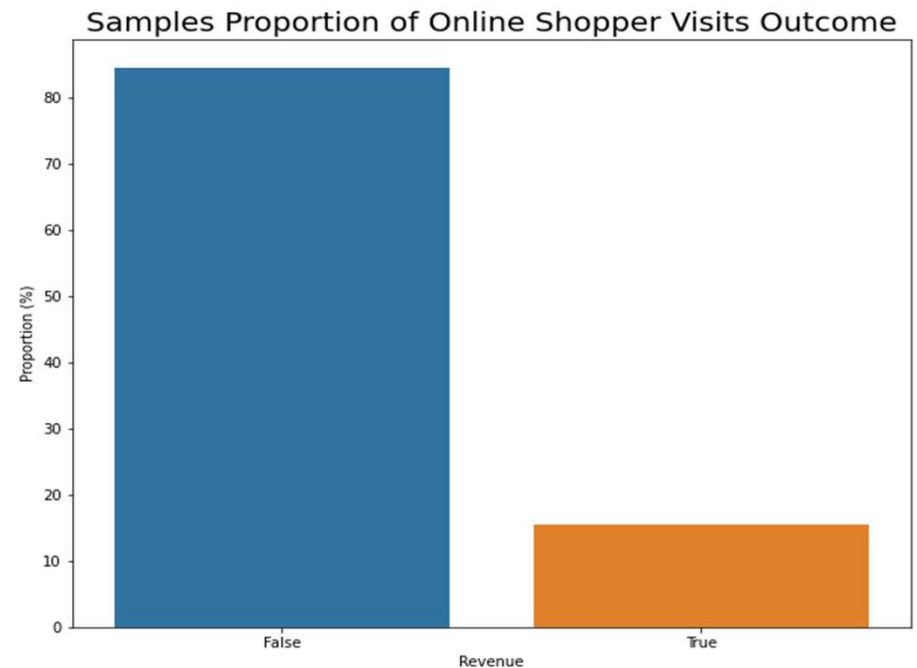
- ▶ The data used is Online Shoppers Purchasing Intention data set provided on the UC Irvine's Machine Learning Repository.

Data Wrangling

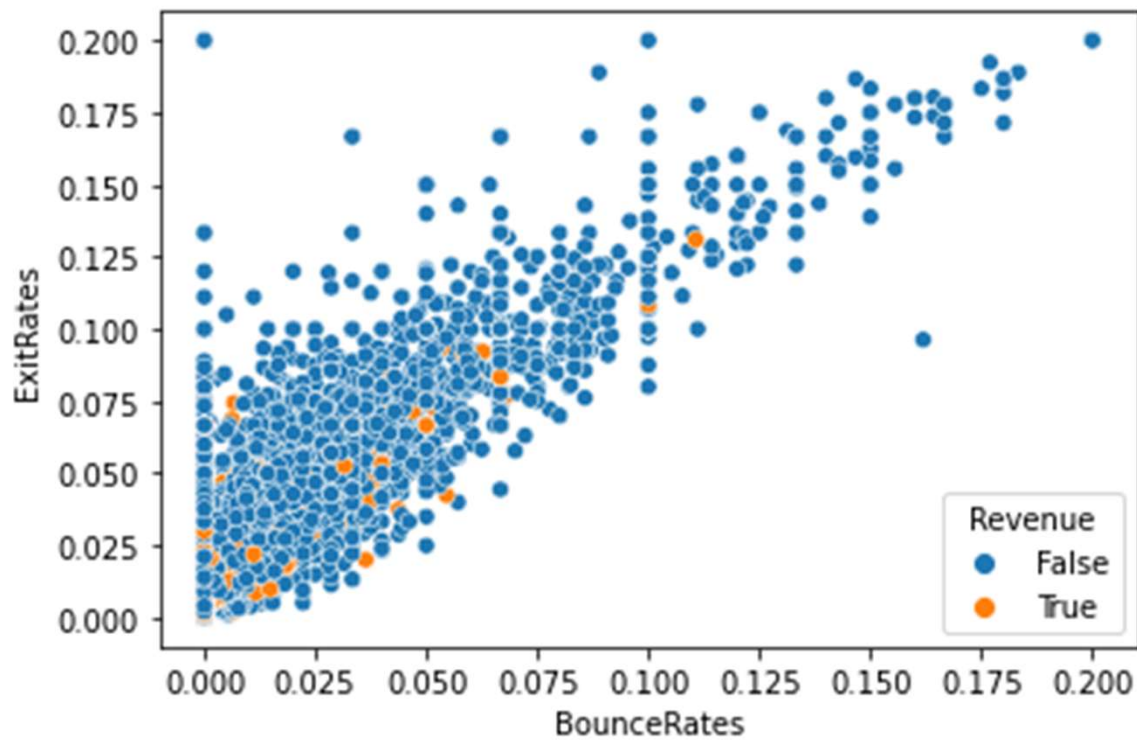
- ▶ The dataset consists of feature vectors belonging to 12,330 sessions of different users in a one-year period.
- ▶ The dataset consists of 10 numerical and 8 categorical attributes.
- ▶ There are no null values.
- ▶ Target variable is Revenue

Exploratory Data Analysis

- ▶ In this dataset, 85% of the users did not end up making any purchase and only 15% users made a final purchase.
- ▶ This is an imbalanced dataset. The figure shows the sample proportion of the revenue.
- ▶ Also, there is a high correlation between the page values and revenue.



Exploratory Data Analysis

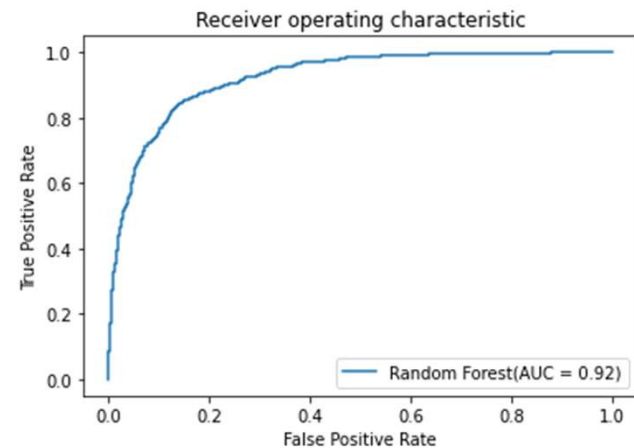


Model Selection

- ▶ Three different machine learning classification models were run:
 - ▶ Logistic Regression
 - ▶ Random Forest Classifier
 - ▶ Support Vector Classifier
- ▶ The chosen metric was F1 score as we have an imbalanced dataset.

Hyperparameter Tuning

- ▶ Applied hyperparameter tuning on the Random Forest model using the Grid Search Cross Validation.
- ▶ The accuracy of the model further increased from 88.32 % to 89.18%.

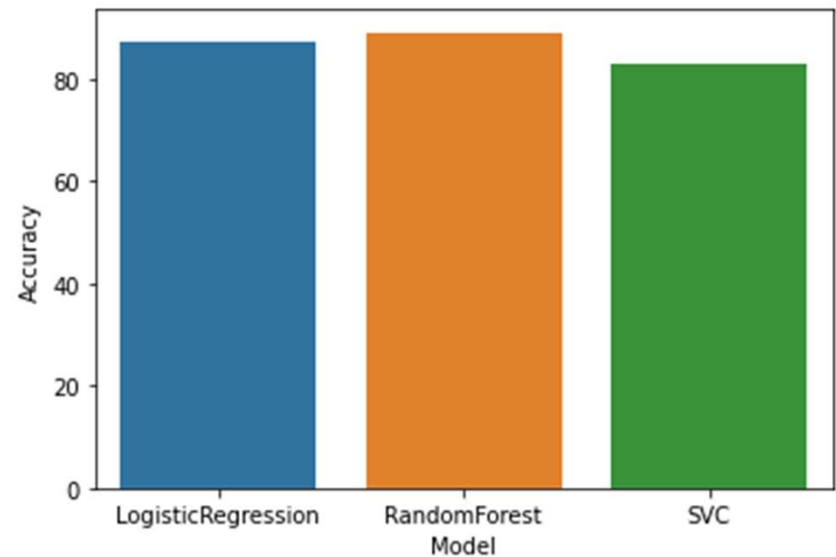


Model Comparison

► Comparison of the three models

Model Accuracy

- Logistic Regression 87.40
- Random Forest 89.13
- SVC 83.18



Conclusion

- ▶ The Random Forest model is the best model for predicting the online shopper's purchasing intention.
- ▶ The page values feature is the most important feature followed by Exit rates, product related page values and the bounce rates.
- ▶ We can further try running different models, like Naïve Byes, Neural Networks etc and see if it further enhances the model performance.



Thank You!