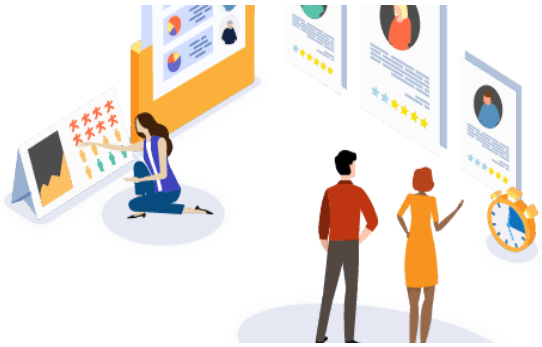
HR Analytics: Job Change of Data Scientists

Business Problem:



The situation is a company which is active in the Big Data and Data Analysis space is offer in courses to some of its employees. The company is offer paid training to their employees. However, they have been running into the situation where upon finishing a course they end up switching companies. Company wants to know which of these candidates really wants to work for the company after training or looking for a new employment.

Objective:

As a Data Scientist, I could determine if these candidates really want to work for the company after training or looking for a new employment and create human understandable insights. The project can help to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates.

Process:

Feature Selection:

- My approach in selecting Feature 'SelectKBest' and 'f_classif' techniques.

Oversampling:

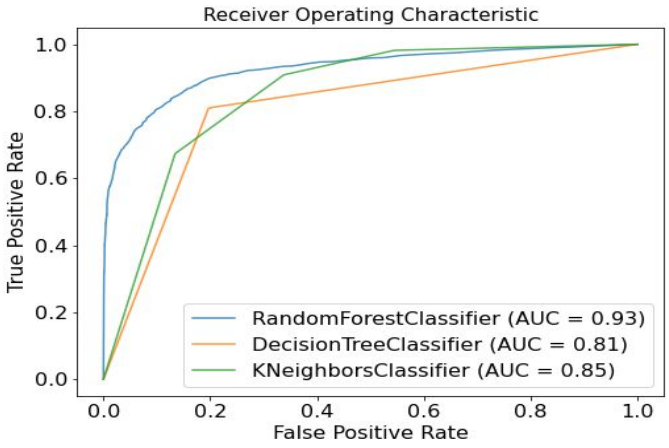
- Used a Synthetic Minority Oversampling Technique (SMOTE) .
- The accuracy score using the RandomForestClassifier before SMOTE is (0.78), after is (0.86).

Evaluation the Models

- Compare the accuracy of Baseline (0.75) to three machine learning models that applied on data.
- Confusion Matrix.
- ROC/AUC curve

Results:

I have applied three different machine learning models to the data. Random Forest Classifier seem to be the most successful out of them. Random Forest achieved 0.86 accuracy score and 0.93 AUC scores. It is perfect and I think we can call it a successful classification.According to my approach in selecting feature importance we can see 'city_development_index', 'company_size' are the most important factors in hob changes. Data scientists in cities with better development index which work in a higher size of company don't tend to change their jobs.



Project link:

[Link github repos]