DATS 6103 - Project Proposal on HR Analytics

Team Members:

1. Adina Dingankar
2. Pranay Bhakthula
3. Rehapriadarsini Manikandasamy

Project Description:

The focus of this project is to predict the probability of a candidate to look for a new job or who will continue to work for the company, as well as interpreting the research question "Can the current factors like credentials, demographics, experience will affect the employee's decision of continue working in the firm or will that employee quit". It will be demonstrated by using three machine learning algorithms: Random Forest Classifier, Decision Tree and Support Vector Machine respectively and develop a GUI based application to display the end-to-end modelling.

The dataset used is sourced from Kaggle which has educational and professional records of various candidates who have completed training in a company. The dataset has 19158 observations and 14 features, most features are categorical (Nominal, Ordinal, Binary) and some with high cardinality. The dataset is imbalanced and 8 amongst the 14 features has missing values. Hyperparameter tuning will be implemented on features and the model will be built by using NumPy, pandas (EDA), Sklearn (Feature Engineering), matplotlib & seaborn (Visualization), PyQt5(GUI). We will be referring to documentation of different packages which will be used for the application development. The performance of the models will be calculated using metrics like Accuracy score, Confusion matrix, recall score, f1 score.

Project Schedule:

DEADLINES	TASKS
04/08/2021	Proposal
04/12/2021	Data cleaning and Preprocessing
04/14/2021	Data Visualization and Random Forest
04/19/2021	Decision Tree and SVM
04/24/2021	GUI Enhancement
04/29/2021	Project Documentation
05/1/2021	Demo Recording

Dataset source:

Source: Kaggle
Name: HR Analytics: Job change of Data Scientists
Source Link: https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists

Group GitHub:

https://github.com/adingankar/FINAL_PROJECT_