BITS Pilani Pilani | Dubai | Goa | Hyderabad Work Integrated Learning Programmes	Introduction to Data Science M.Tech Data Science and Engineering	Start Date : 21-06-2021 End Date: 21-08-2021

Data Scientists Job churn analysis

Introduction

Job change from one organization to other is a never ending problem in the private sector. Particularly there seems to be a trend in job churns in IT companies. Data Scientist Job is the Golden job of the 21st century and they are in demand. The following data is collected by IBM HR department for churn analysis before enrollment of employees to a training session.

Objective

Follow the Data Science Methodology that we studied in Module 3. Analyze the data set and identify the churn trend of Data scientists in IBM. Answer the questions like

- Whether male or female employees tend to change jobs often?
- What is the influence of experience on job change?
- What is the role of education in job change?
- Represent the trend in different cities and job change.

Dataset

The dataset, aug_train.csv, is taken from Kaggle. It has been uploaded to canvas.

Tasks in this assignment

1) Write a Data Science Proposal for achieving the objective mentioned.
2) Perform exploratory analysis on the data and describe your understanding of the data.
3) Perform data wrangling / pre-processing.
    a. E.g., missing data, normalization, discretization, etc.,.
4) Apply any two feature selection engineering techniques.
5) Compare the two selected feature engineering techniques.
6) Plot top 5, 6, and 8 features.
7) Provide a high-level description of Machine Learning models – Logistic regression and Decision tree to predict the employee job change trend and answer the listed objectives.
8) Compare the performance of the two classifiers – Logistic regression and Decision tree to predict.
9) Present the conclusions/results in a neat format.
10) Any suggestion on the identification of the performance parameters to be improved, for the given classification problem.

Expected Submissions

Two files are expected as the assignment submission.

BITS Pilani Pilani | Dubai | Goa | Hyderabad Work Integrated Learning Programmes	Introduction to Data Science M.Tech Data Science and Engineering	Start Date : 21-06-2021 End Date: 21-08-2021

1. The summary of the work in the template provided. (you may fill only the boxes relevant to this problem statement)
2. The executed ipynb file with clear subdivision of the codes and brief description of the purpose of respective code. All the executed tables or graphs and results should be present in the ipynb file. The ipynb file maybe submitted as a single .pdf file.