Group Number: 72

Names of Students and Index Numbers:

a) EG/2020/3795 – Aafrith U.M - aafrith_um_e22@engug.ruh.ac.lk
b) EG/2020/3807 – Ahamed M.F.R - ahamed_mfr_e22@engug.ruh.ac.lk

Project Title: Employee Attrition Prediction in HR Analytics

Introduction:

- In HR Analytics, predicting employee attrition is crucial for talent retention and workforce stability. Our project develops an Employee Attrition Prediction system using machine learning. It analyzes factors like job satisfaction, work-life balance, and salary to forecast employee likelihood of leaving, providing actionable insights for HR professionals.

Defining the ML Problem (T: Task, E: Experience, P: Performance Measure):

- **Task:** Predict employee attrition based on historical HR data.
- **Experience:** Employ machine learning to discern patterns from past employee data.
- **Performance Measure:** Evaluate model using accuracy, precision, recall, F1 score, and ROC-AUC for discrimination between employees who leave and those who stay.

Data:

1. Data Source : https://www.kaggle.com/code/carmelgafa/ibm-hr-dataset-analysis-and-prediction.

2. Features: The features include variables are Age, Gender, Number of Companies Worked, Distance from Home, Marital Status, Years at Company, Total Working Years, Years in Current Role, Years Since Last Promotion, Job Satisfaction, Work Life Balance, Environment Satisfaction, Monthly Income, Percent Salary Hike, Performance Rating. The number of features will be determined during the exploratory data analysis.

3. Dataset Size: We have 34 features that describe each employee in our dataset.

4. Handling Missing Data: Using mean or median imputation for numerical features, mode imputation for categorical features, or KNN imputation. Consider domain-specific imputation for job roles and weigh options such as listwise deletion or predictive models based on dataset characteristics. Document the chosen method for transparency.

Algorithms:

a. We do the project from the following algorithms.
    1. Logistic Regression
    2. Naïve Bayes

b. Availability: Both algorithms are available off the shelf in popular machine learning libraries such as scikit-learn for Python.