[Project Code: EALG]

Employee Attrition Prediction using Logistic Regression

Project Duration: 25-Feb-2024 – 16-Mar-2024
Submission Information: (via) CSE-Moodle

Objective:

The primary goal is to predict employee attrition based on key factors such as education level, environment satisfaction, job involvement, job satisfaction, performance rating, relationship satisfaction, and work-life balance.The goal is to provide actionable insights for HR analytics, helping organizations understand and address workforce attrition effectively. Build a logistic regression to verify whether these factors can predict attrition(stay/leave).

Tasks to be done:

1. Randomly divide the dataset into 80% training set and the rest as test set. Choose the important features from the dataset.Choose a mini-batch size to divide the dataset into batches.
2. Build the Logistic Regression model by your own using only numpy and pandas.
3. Give the results of the Hyper-parameters if you have any.
4. Classification Report
    a. Create a classification report for comparing the performance of your algorithm in terms of accuracy.
    b. You need to calculate precision, recall, f1-score and accuracy of the model.

Note: *The program can be written in C / C++ / Java / Python programming language from scratch. No machine learning /data science /statistics package / library should be used for model creation.*

Relevant information:

Dataset Filename: HR-Employee-Attrition`.csv`
Data Description:
The IBM HR Analytics dataset is designed for predicting employee attrition and understanding factors influencing employee performance. This dataset typically contains a range of features related to employees, including demographic information, job satisfaction metrics, performance ratings, and other relevant attributes.

Features:
The dataset commonly includes features like:

- Age
- Education level
- Environment satisfaction

- Job involvement
- Job satisfaction
- Performance rating
- Relationship satisfaction
- Work-life balance
- and more.

Target Variable:
The target variable is typically binary, indicating whether an employee has left the company (attrition) or is still employed.

Submission Details: (to be submitted in CSE-Moodle, **by one representative of the group**)

1. ZIPPED folder containing code (with comments) and the dataset files
2. Report (in pdf format)

Submission Guidelines:

1. You may use one of the following languages: C / C++ / Java / Python.
2. Your program should run on a Linux Environment.
3. Your program should be standalone and should not use any special purpose library for Machine Learning. (Apart from numpy, pandas and the one's mentioned in tasks). And, you can use libraries for other purposes, such as formatting and visualization of data.
4. You should submit the program file and a README file with instructions to run the code.
5. You should name your file as <GroupNo_ProjectCode.extension>.
   E.g., *Group99_EALG.zip* for code-distribution and *Group99_EALG.pdf* for report)
6. The submitted program file *should* have the following header comments:
   ```
   # Group Number
   # Roll Numbers : Names of members (listed line wise)
   # Project Number
   # Project Title
   ```
7. Submit through CSE-MOODLE only.
   Link to course page: https://moodlecse.iitkgp.ac.in/moodle/course/view.php?id=561

You should not use any code available on the Web. Submissions found to be plagiarized or having used ML libraries (except for parts where specifically allowed) will be awarded zero marks.

For any questions about the assignment, contact the following TA:
Akash Kundu (Email: kunduakash91@gmail.com)