Contents

I. Problem Description

Predict employee attrition

Employee Attrition is a huge problem across industries and generally costs the company a lot for hiring, retraining, productivity and work loss for each employee who leaves. Price and Waters, a boutique data science consulting firm, is looking to build a Machine Learning model to predict whether an Employee might quit. Using this model, they might plan human intervention to alleviate the issues faced by the employee. The firm is also interested in specific features that are highly indicative of attrition.

The company in a pilot program, recorded employee data. The company collected employee performance data for some of the months randomly for each employee to understand it in the context of attrition. The company wants you to predict whether an employee would quit in the near future, given the data and to discover features indicative of attrition.

‘Left_Company’ is the target variable and you would have to predict either ‘1’ (Left), ‘0’ (Retained) for each unique employee id in the test dataset.

II. The datasets are provided as cited below:

Target attribute: "Left_Company" (yes – 1, no – 0)

Train Data (2 CSV Files): ***train_attrition.csv &amp; train_work.csv***

Test Data (2 CSV Files): ***test_attrition.csv &amp; test_work.csv***

Attribute details:

- Left_Company (Target) : Whether the employee left the company or not (1 - Yes, 0 - No)
- EmployeeID : A unique identification key for every individual employee
- TotalWorkingHours : The total working hours logged for the employee at the location
- Billable_Hours : The number of hours that are used to charge the Client
- Hours_off_Duty : Number of hours the employee took off
- Touring_Hours : Number of hours the employee spent working at an offsite location

- NoOfProjects : Number of Projects the employee is assigned to
- ActualTimeSpent : Actual time the employee spent working according to the timesheets

Supplemental Data: employee_data.csv

Note: employee_data.csv has the details of employees present in both train and test

- Specific data regarding Employees for both Train and Test data
- EmployeeID : A unique identification key for every individual employee
- Job_History : A feature containing the previous companies where the employee was employed
- Joining_Date : The date on which the employee Joined the organisation
- Designation : The role of the employee in the company, with the following levels: EVP, Junior, MD, Senior, VP
- Sex : The gender of the employee

III. Tasks:

Model Building:

You are expected to create an analytical and modelling framework to predict whether an employee will leave the company or not based on the quantitative and qualitative features provided in the datasets. You may derive new features from the existing features and also from domain knowledge, which may help in improving the model efficiency.

Visualization Tasks:

Exploratory Data Analysis using visualizations in R Notebook or Jupyter notebook format. (all of the training data to be used for this task)

- List down the insights/patterns observed from the visualizations
- Explain the impact of most important attributes on target attribute observed from the visualizations.

Observations:

- Which employees tend to leave the firm?
- Is there any overfitting or underfitting problem? If yes, how do you address it?

IV. Evaluation Metric:

- Consider '**F1-score**' for 'Leave_Company' = 1 as the error metric for classification task to tune the model and for submissions in the tool.

V. Submission File

- You must submit a file where each EmployeeID has the corresponding prediction for 'Left_Company'

- Ensure that the EmployeeIDs are in the same order as given to you