

Problem:

Within the context of human resources (HR), attrition is a reduction in the workforce caused by retirement or resignation. This is a serious problem faced by several organizations around the world as attrition is economically damaging to the organizations as the replacement employees have to be hired at a cost and trained again at a cost. High Rates of Attrition also damage the brand value of the company.

Now the Dataset belongs to a very fast-growing company. This company has witnessed several employees leaving the company in the last 3 years. The company's HR team has always been reactive to attrition but now the team wants to be proactive and wished to predict the attrition of employees using the data they have in hand.

The goal here is to predict whether an employee will leave the company based on the various variables given in the dataset.

Working with Data

Data has been split into two groups and provided in the module:

training set

test set

The training set is used to build your machine learning model. For the training set, we provide the attrition details of an employee.

The test set should be used to see how well your model performs on unseen data. For the test set, it is your job to predict the attrition value of an employee.

Metric to measure

Accuracy is the metric to measure the performance in this problem.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN)$$

Submission File Format:

You are to submit a CSV file with exactly 2630 entries plus a header row. The file should have exactly two columns

1. EmployeeID (sorted in any order)
2. Attrition

Please use the "Test_Dataset" for carrying out your predictions