

# PROJECT SYNOPSIS

## Predicting Employee Attrition

*Submitted towards the partial fulfillment of the criteria for an award of Genpact  
Data Science Prodegree by Imarticus*

*Submitted By:*

MANISHA ANAND

*Course and Batch: PRODEGREE DATA SCIENCE\_DSP31*



## **Scope & Objective**

### **Introduction:**

Employee Attrition is when an employee leaves a company due to normal means, (loss of customers, retirement, and resignation), and there is not someone to fill the vacancy. Can a company identify employees that are likely to leave a company?

Lack of growth and advancement opportunities is a reason for attrition. When companies do not offer advancement and developmental opportunities for capable and ambitious workers, they end up losing talented employees and retaining poor performers. A company with a high employee attrition rate is a sign of underlying problems and can affect a company in a very negative way. One such way is the cost related to finding and training a replacement, as well as the possible strain it can put on other workers that in the meantime have to cover.

### **SCOPE OF THE STUDY:**

- To determine the relationship between various variables with employee attrition.
- Determination of predictors.
- To fit a model that best predicts Employee Attrition.
- To test the model for accuracy.
- To present the findings and conclusion to the management.

### **OBJECTIVES OF THE STUDY:**

- This study will help the company to understand more about the attrition rate in the company.
- The study aims at identifying the causes of attrition for employees in the organization (feature selection).
- To build a predictive classification model for finding whether an employee will leave the organization in the future or not.

### **Business Problem Statement**

Employees are the most important part of an organization. Successful employees meet deadlines, make sales, and build the brand through positive customer interactions.

Employee attrition is a major cost to an organization and predicting such attritions is the most important requirement of the Human Resources department in many organizations.

**Problem Statement:** To determine the factors that most significantly influence employees' decision to leave the organization and build a classification model for predicting any future attrition.

## **Data Sources**

The csv file 'Attrition.csv' contains the data related to 1470 employees of IBM HR Analytics Employee Attrition & Performance. The data has 35 variables. Data contains different attributes of an employee and the target variable *Attrition*. We use this dataset to predict employee churn.

## **Analytics Tools**

The project is complete using R software and various analytics tools like Exploratory data analysis using ggplot, Hypothesis testing (t test, chi square and Boruta test used for feature selection), Fitting of classification Models like Logistics Regression, Decision Trees, Random Forest, kNN and SVM models, comparison and selection of Models based on ROC Curve, Accuracy, Sensitivity & Specificity values.

## **Analytics Approach**

- Feature Selection
- Fitting various models
- Model Comparison
- Model Selection
- Prediction from best Model

## **KPI's**

Evaluation will be based on:

- Feature Selection (10%)
- Feature Engineering (25%)
- Model Comparison (30%)
- Model Selection (20%)
- Presentation (15%)