IMARTICUS LEARNING

KPMG

Employee Attrition Analytics

DATA SCIENCE PRODEGREE PROJECT

ABSTRACT

This project gives an overview of the attrition analysis for a consulting firm and allows you to understand and identify factors that help in retaining employees.

About Employee Attrition

Employee attrition refers to the loss of employees through several circumstances, such as resignation and retirement. The cause of attrition may be either voluntary or involuntary. Each industry has its own standards for acceptable attrition rates. Due to the expenses associated with training new employees, any type of employee attrition is typically seen to have a monetary cost.

Attrition rate at any organization for a given month is calculated as the total number of employees leaving the firm divided by the total headcount for that month.

Overview of the problem

In this project you will have to put yourself in the shoes of a data scientist and identify the factors contributing to attrition of the employees and recommend possible solutions. The files contain complete staff utilization reports for all employees of the XYZ Corp. through 2016-2018. The data contains details of employees as well as the details of the projects that they have worked on in the last two years. The other file contains all the attrition in the organization for the years 2015-18 with details such as reason of attrition along with other employee details. You are free to use external data sets which may impact the attrition.

The data should be divided into train and test data set. You will have to use the training data to build models/analytical solution and finally apply it to test data to measure the performance and robustness of the models.

Data and Problem Details

Objective: You have to build a data model to predict the following:

- Identify factors influencing attrition

- Predict possible attritions
- Identify possible ways to retain high performers

You must do the following:

Based on the data that is available, build a model to predict attrition probability of an employee. This will help the company to identify employees who have high possibility of quitting and will allow them to allocate resources to combat the issue.

You will be provided with:

- Dataset containing both train and test data
- Data dictionary

Steps to be followed (tentative time required):

1) Understand the problem and objective (1 hour)
2) Understand the data and develop some business sense. (4-5 hours)
3) EDA (if you think is required in this case). (5-6 Hours)
4) Data Cleaning and Preparation (4-5 Hours)
5) Feature engineering (4-5 Hours)
6) Model Building (try various techniques and at the end justify why you chose a technique over others) (3-4 hours)
7) Testing and Cross-validation (3-4 hours)
8) Final results, recommendations and plots/visualizations. (4-5 hours)
9) BONUS: Any other insights or recommendations that you can give from the data which will help the business. (Subjective)
10) Preparing the deck: 6-7 hours
[Actual time might vary from person to person and step to step, this is just indicative]

The final solution should be in the form of a deck showing all the steps above. It will be judged on the following criteria:

1) How well have you adhered to the modelling process discipline?

2) Do your results make business sense, how have you used business intuition to take decisions during the modelling exercise, including but not limited to the following:

- Deciding Segmentation (if you choose to have segmentation)
- EDA, Feature engineering
- Choosing variables to be put in models
- Deciding a cut-off

3) Performance of your model on test data

Grading

The project and presentation will be assessed & graded on completion. Details on this will be provided separately.