****

Human Resources Analysis

Predict Attrition

**Human Resources Analysis**

Quit Job Probability

The Challenge

You are a Data Scientist and you have a dataset with information from all employees that have been working at a specific company during a time window.

During these years, the Human Resources team collected data from employees and you have been asked to predict the attrition of an employee, i.e. the probability of an employee with certain characteristics stay or quit his/her current job in the company.

Predicting the employee attrition can be a demanding task due the complexness or even hidden information that may be behind the employee’s decision. Regardless the lack of some crucial information, you must build a model to use by this company’s Human Resource team warning them about employees that potentially will quit. With something like that they can try to use some preventions methods such as giving better salaries; promote new and interesting projects; offer the opportunity to work remotely if the problem is related with job distance from home, let them suggest new tools, etc… to avoid that.

This number of variables that may be involved but doing it may bring an enormous value to the employee and to the company itself.

Another interesting mark of this use case is that many data sources can be used, giving the students the opportunity to build a good training set for their models and a good understanding of the feature extraction process (one of the most time consuming and challenging procedures in Machine Learning).

The customer we are working with is a multi-national company that works in consultancy area. They want to correlate the distance from home with the attrition rate. They also want to understand how important the variables are, and what makes an employee keep working on the company.

The main goals are:

* **Descriptive Analytics:** Find correlations between the different variables, possibility of clustering analysis, build nice visualizations that may help to get better insights for the analysis, check feature’s cardinality and analyze possible highly correlation features that may be removed and still achieve good results.
* **Predictive Analytics:** Build classification models to predict the *Attrition* probability and the final classification given by the model.

Data

Data is in csv format and can be easily imported to the platform you desire to use. It’s name is HR\_DS.csv.

Technologies

No limitations. Can be R, Python for data understanding, data preparation, feature engineering and modelling.

No limitations. Can be Tableau, PowerBI, Qlikview for data exploration, visualization, presentation of pertinent results.

Optional: Shiny APP for results and relevant insights presentation (works well for a small number of records, for bigger number a license would be required).