Data Science Certification Training

Certification Project

edureka!

Industry: Human Resource management

Challenge –

A company has been in industry since a long time. Their business had been increasing quite well over past, however in recent years, there has been a slowdown in terms of growth because their best and most experienced employees leaving prematurely. The VP of the firm is not very happy with the company's best and most experienced employees leaving prematurely. The VP of the firm has employed you to find out insights in the company employee data and find out an answer as to know why best and most experienced employees are leaving prematurely.

Solution –

As a first step the VP planned to know the useful insights out of the employee data available. Using R, he also wanted his team members to give him a forecast model to predict which employees could be leaving the company, as well as answer to why their best and most experienced employees are leaving prematurely. This will help him plan his next steps to avoid the churn out. He wanted a script that would contain the following:

- A visualization and distribution (of all the employee relative fields)
- Forecast using different machine learning models
- Comparison among different machine learning models and cross validating through them test and train set
- Find out why best and most experienced employees are leaving prematurely.
- Give a final prediction model to forecast

Actions to be Performed

- Set the Directory and load the dataset into R, verify that the data is loaded correctly
- For finding the insights out of our data several techniques can be used
    - Find the correlation values of the attributes of our data
    - Visualize the characteristics of the whole data and only the people who left, use plots and histograms
    - Evaluate the values of each attributes for both left and non-left employees
    - Analyse the department wise turnouts and find out the percentage of employees leaving from each department
- Build a classification model to forecast what are the attributes of people who leave the company
    - Build models using Decision Tree, Random Forest, Naïve Bayes and SVM techniques and find out the most accurate one

Data to be used: The dataset being used is HR_data. It consists of dimensions like:

- Employee satisfaction level
- Last evaluation
- Number of projects
- Average monthly hours
- Time spent at the company
- Whether they have had a work accident
- Whether they have had a promotion in the last 5 years
- Department
- Salary
- Whether the employee has left

Tools required: RStudio.