Identifying the employee churn

For this activity you will be evaluating a dataset composed of human resources data.

A Company X has a goal to reach less than 1 % attrition rate by the end of the year by identifying employees who need one to one intervention before they drop papers of resignation. As a data scientist, your task is to model the factors that " predict how likely an employee is to churn the company" , using supervised learning techniques. You will need to analyse the dataset on employee churn.

Your task is to build a logistic regression model (or K-NN model) that predict whether or not an employee is likely to leave his/her job based on characteristics in the dataset.

- id - Anonymous ID number for each employee
- satisfaction - Employee satisfaction level
- evaluation - Last evaluation score
- number_of_project - Number of projects assigned to
- average_monthly_hours - Average monthly hours worked
- time_spend_company - Time spent at the company
- work_accident - (1= Yes, 0 = No) - Whether they have had a work accident
- churn - Whether or not employee left company (1 = Yes, 0 = No)
- promotion - (1 = Yes, 0 = No) - Whether they have had a promotion in the last 5 years
- department - Department name (not just sales)
- salary - Salary category

Steps Involved:

1. analyse the data
2. Do exploratory data analysis
3. Do data pre-processing
4. Build models and evaluate the performance
5. Identify the cross validated score