Other :- Faldu Jay

Predicting Employee Attrition in Down of Recession

1 Introduction

1.1 Background

As the COVID-19 keeps unleashing its havoc, the world continues to get pushed into the crisis of the great economic recession, more and more companies start to cut down their underperforming employees. Companies firing hundreds and thousands of Employees are a typical headline today. Cutting down employees or reducing an employee salary is a tough decision to take. It needs to be taken with utmost care as imprecision in the identification of employees whose performance is attiring may lead to sabotaging of both employees' career and the company's reputation in the market.

1.2 Aim of the project

Create a model that can predict the attrition probability of any employee based on the past data available of that particular employee.

1.3 Stakeholder

Employee work in sector which is highly affected by economic recession and also employer can show their interest in this project. The model can help them to find weather there are going to throughout from job or other hand employer can decide which employ they need to attiring.

2 Data Description

2.1 Source of Data

I collect this data from kaggle database.

Link :- data of employee history

2.2 Data fields

Our data set contain below field in this. I list all column name and the data what data contain in particular columns.

- Id - an anonymous id given to an Employee
- Age - Age of an Employee
- Attrition - Did the Employee leave the company, 0-No, 1-Yes
- BusinessTravel - Travlling frequency of an Employee
- Department - Work Department
- DistanceFromHome - Distance of office from home
- EducationField - Field of Education
- EmployeeNumber - Number of Employees in the division of a given Employee
- EnvironmentSatisfaction - Work Environment Satisfaction
- Gender - Gender of Employee
- MartialStatus - Martial Status of an employee
- MonthlyIncome - Monthly Income of Employee in USD
- NumCompaniesWorked - Number of Companies in which Employee has worked before joining this Company
- OverTime - Does The person work overtime
- PercentSalaryHike - Average annual salary hike in percentages

- StockOptionLevel - Company stocks given to an Employee
- TotalWorkingYears - Total working experience of an employee
- TrainingTimesLastYear - No. of trainings an employee went through last year
- YearsAtCompany - Number of years worked at this company
- YearsInCurrentRole - Number of years in current role
- YearsSinceLastPromotion - Number of years since last promotion
- YearsWithCurrManager - Number of years with the current manager

Some columns have categorical data. Category of those columns is listed below.

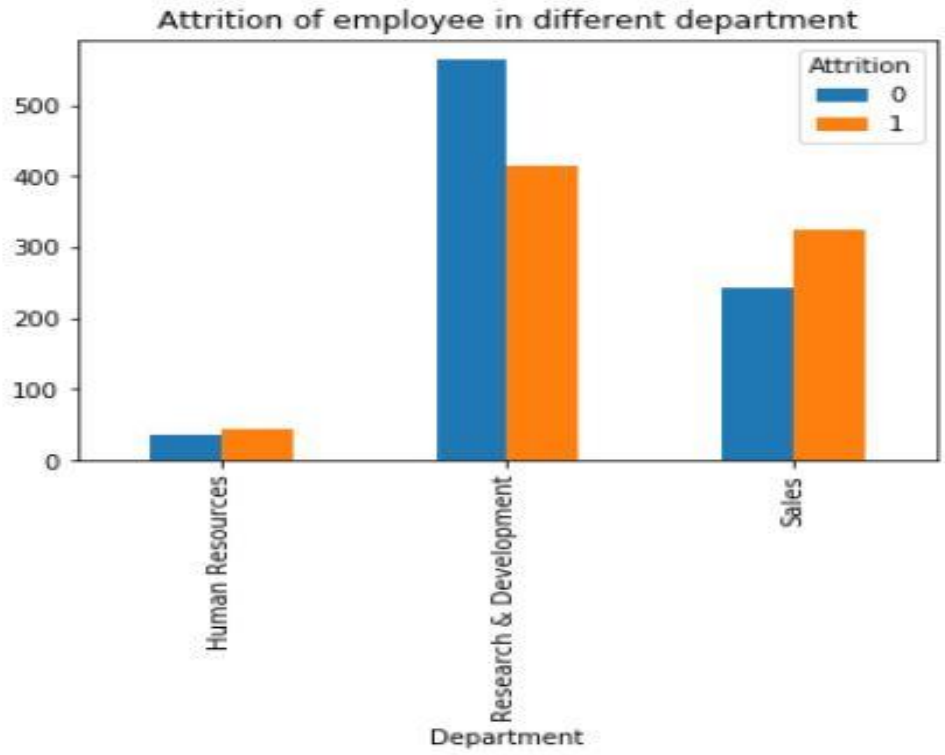
- Education
  1 'Below College'
  2 'College'
  3 'Bachelor'
  4 'Master'
  5 'Doctor'
- EnvironmentSatisfaction
  1 'Low'
  2 'Medium'
  3 'High'
  4 'Very High'
- JobInvolvement
  1 'Low'
  2 'Medium'
  3 'High'
  4 'Very High'
- JobSatisfaction
  1 'Low'
  2 'Medium'
  3 'High'
  4 'Very High’

- PerformanceRating
  1 'Low'
  2 'Good'
  3 'Excellent'
  4 'Outstanding'
- Behaviour
  1 'Good'
  2 'Bad'
  3 'Not Rated'
- CommunicationSkill
  1 'Bad'
  2 'Average'
  3 'Good'
  4 'Better'
  5 'Best'
- StockOptionLevel
  0 'No stocks'
  1 'Less Stocks'
  2 'Moderate Stocks'
  3 'A lot of Stocks'

Our Dataset has total 1628 employees history. There are total 29 column including one for Id of employee and one column name “Attrition” are target value means it contain 0 and 1 which show weather particular employee are attiring or not.

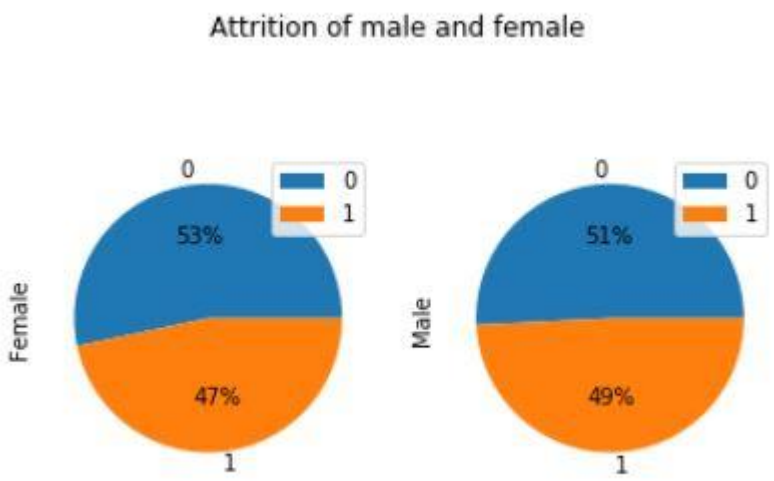
3 Exploratory Data Analysis

3.1 Attrition of employee in different department



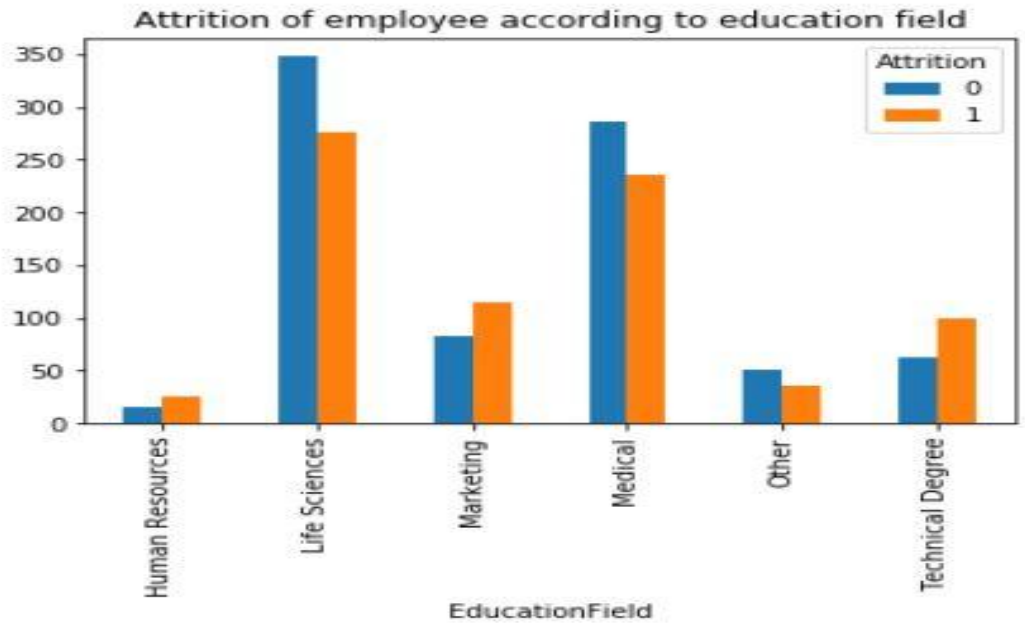
In above chart we can show that in sales and HR department attrition percentage are more then 50%. And in R&D it is less then 50%.

3.2 Attrition of male and female



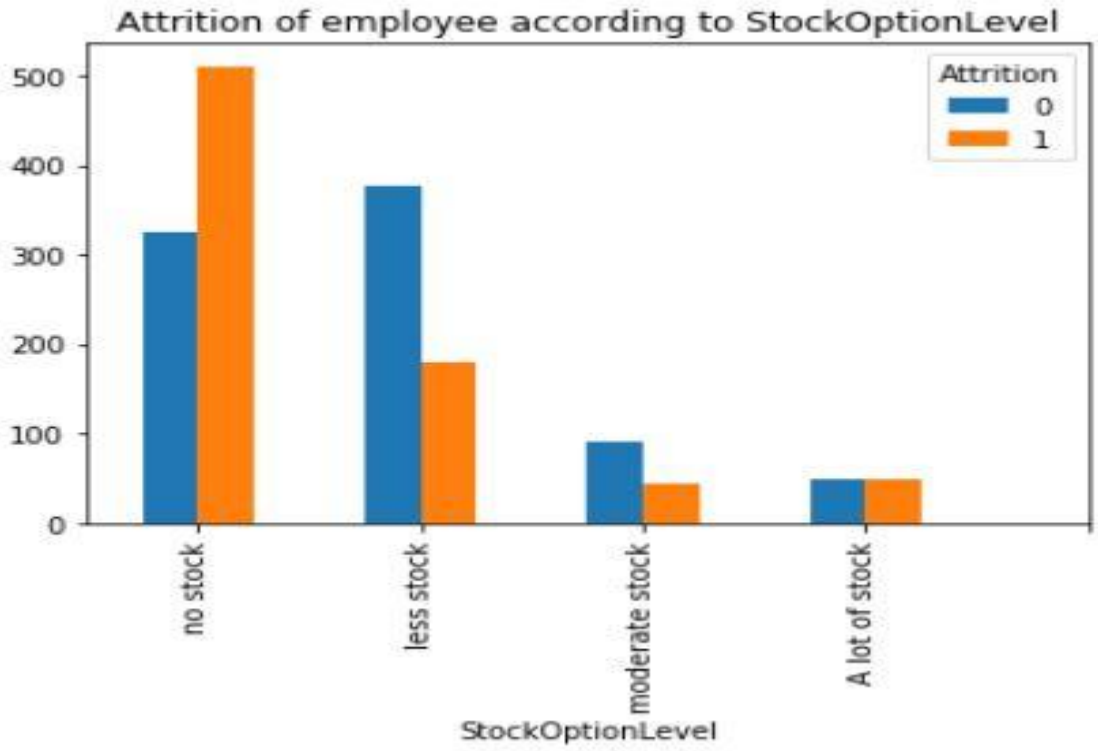
From above chart we can sey that male and female comparatively equal percentage of attrition.

3.3 Attrition of employee according to education field



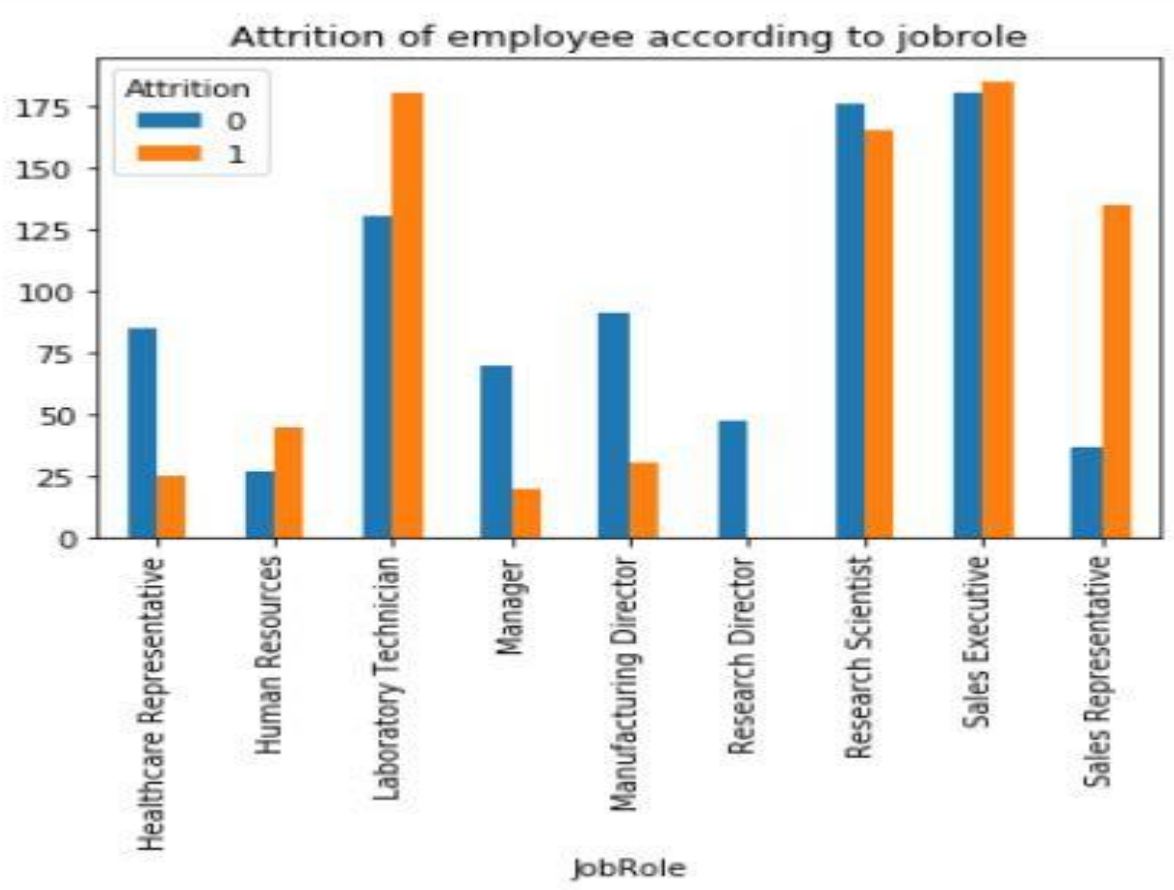
From above chart we can say tha attrition percentage are more in employee having education in field like Hr, marketing and technical degree compare to life science and medical.

3.4 Attrition of employee according to it's StockOptionLevel



From above graph we can say that employee having no stock has more likely to attiring then employee having some stock.

3.5 Attrition of employee according to it's JobRole



From above graph its clear that in managerial role like manufacturing director, research director, manager has less chance of attrition then employee in role of research scientist, sales executive, sales representative, human resources and laboratory technician has more chance of attrition.

Sales representative has highest chance of attrition. And in post of research director there was no attrition at all.

4 Data preparation

In this phase we prepare the data for ferther processing of prediction. And make them ready to apply machine learning algorithm.

In our data set there are lot of feature are having object data type so first we have to convert this data type in to int. So we label them in int data type.

Now we remove some feature from data set which has no meaning in prediction process like Id, employee number etc. We separate the target value which is in column name ‘Attrition’ and store in different data frame.

5 Machine Learning Model for Prediction

We have classification problem to solve. We need to classify employee in Attrition (1, 0) based on given feature.

There is some machine learning model which we can use for example logistic regression classifier, SVM classifier, decision tree, random forest.

Decision tree and random forest are doing well for categorical data. in our data mostly all feature are categorical. So we use any one from those two. Random forest work better then decision tree and it has less variance so we used random forest for this classification.

Accuracy:-

Training set accuracy: - 99.32%

Testing set accuracy: - 98.05%