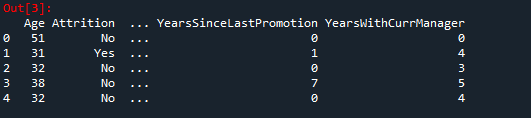
**Logistic regression for Attrition Project:**

import pandas as pd

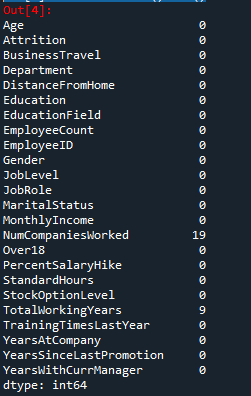
from sklearn import preprocessing

**dataset1=pd.read\_csv("general\_data.csv")**

**dataset1.head()**



**dataset1.isna().sum()**



**dataset1.dropna(inplace=True)**

**dataset=dataset1.drop(['EmployeeID','EmployeeCount','StandardHours','Over18'], axis = 1)**

**dataset.columns**

Out[10]:

Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',

'Education', 'EducationField', 'Gender', 'JobLevel', 'JobRole',

'MaritalStatus', 'MonthlyIncome', 'NumCompaniesWorked',

'PercentSalaryHike', 'StockOptionLevel', 'TotalWorkingYears',

'TrainingTimesLastYear', 'YearsAtCompany', 'YearsSinceLastPromotion',

'YearsWithCurrManager'],

dtype='object')

**//Converting text into numerical**

data=preprocessing.LabelEncoder()

dataset['Attrition']=data.fit\_transform(dataset['Attrition'])

dataset['BusinessTravel']=data.fit\_transform(dataset['BusinessTravel'])

dataset['EducationField']=data.fit\_transform(dataset['EducationField'])

dataset['Gender']=data.fit\_transform(dataset['Gender'])

dataset['Department']=data.fit\_transform(dataset['Department'])

dataset['MaritalStatus']=data.fit\_transform(dataset['MaritalStatus'])

dataset['JobRole']=data.fit\_transform(dataset['JobRole'])

**Y=dataset['Attrition']**

**X=dataset.drop(['Attrition'],axis = 1)**

**import statsmodels.api as sm**

**X1=sm.add\_constant(X)**

**Logistic=sm.Logit(Y,X1)**

**result=Logistic.fit()**

Optimization terminated successfully.

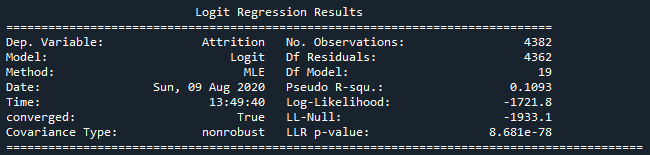
Current function value: 0.392916

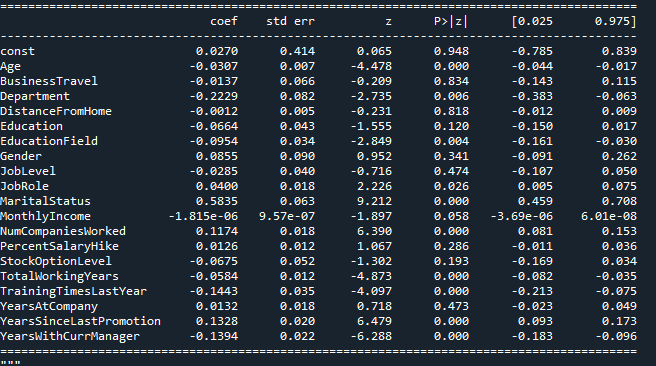
Iterations 7

**result.summary()**

Out[26]:

<class 'statsmodels.iolib.summary.Summary'>





**Inferences:**

**In attrition following factors are significantly important.**

1. Age of the employee.
2. Education field
3. Marital Status
4. Number of Companies worked
5. Total Working Years
6. Training times last year
7. Years since last promotion
8. Years with current manager
9. Department of employee