

**PROJECT TITLE:-**

**GENERATING EMPLOYEE ATTRITION**

Submitted By:

K. Gagan Sai

Y. Sampath

Y. Dinesh Kumar

V. Adithya

***CERTIFICATE***

This is to certify that this project based Class report entitled **“EMPLOYEE ATTRITION”** is a bonafide work done by K. Gagansai, Y. Sampath, Y. Dineshkumar,V. Adithya, in partial fulfillment of the requirement for the award of **internship conducted by smart bridge** during the summer 2019.

***DECLARATION***

We hereby declare that this project based lab report entitle **EMPLOYEE ATTRITION”** has been prepared by us in partial fulfillment of the requirement for the award of **internship conducted by smart bridge** during the summer 2019.

We also declare that this project based class report is of our own effort and it has not been submitted to any other university for the award of any degree.

**Date: 01-06-2019**

**Place: Hyderabad**

Submitted By:

K. Gagansai

Y. Sampath

Y. Dinesh kumar

V.Adithya

**TABLE OF CONTENTS**

1. Abstract…………………………………..

2. Introduction………………………………

3. Objectives of Research

4. Review of Literature

5. Data Collection

6. Methadology

Exploratory Data Analysis

Figure:-6.1 Data Cleaning

Figure:-6.2 Datatype

Figure:- 6.3 Mean imputation for numerical data

Data Description

Figure:-6.4 Data Description

Data Correlation

Figure:- 6.5 Data Correlation

Heat Map

Figure:- 6.6

Plots and Model

Figure:- 6.7 Displot

Figure:- 6.8 Box Plot

Figure :- 6.9 Pairplot

Figure:- 7.0 Scatter Plot

7. Algorithms………………………………..

7.1 KNN:

7.2 Logistic Regression:

8. Result…………………………………….

9. Conclusion……………………………….

ABSTRACT

This project aims through machine learning techniques at creating a model, employee attrition classification. Data includes Education, job level, job involvement, Daily rate, Monthly Income, Experience, No. of Companies worked, No of companies changed ,Total work experience, Years at company, Years at current role, Years since last promotion, Years with current manager, Training times last year, Performance rate, Attrition, By following above given data we finding the complete job profile of the person. By this data and by satisfying the above given data we know the employees data ad their experience in company and the level of the person in the company. And by the graph and the plots we can prove it graphically. And we can the no of employees working in the company. An employer is the authority which employs and pays employees for their labor. It may be an individual person or it may be a company representing many people. Within the relationship between employers and their employees, the employer is the party which will typically define the terms of employment and write the contract. They are then obligated to provide the agreed-upon compensation to workers for any labor they perform that is contained within the terms of their contract. An employer is also the party legally liable for work conditions, maintaining labor laws and handling any legal action an employee may pursue. Celebrating organizations seeing the best return on an investment in people, this Award recognizes the employee strategy that best attracts, retains and develops talent – and how being an employer of choice contributes to sustainable growth.

Introduction

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data. In this article, we’ll see basics of Machine Learning, and implementation of a simple machine learning algorithm using python.

Machine learning involves computer to get trained using a given data set, and use this training to predict the properties of a given new data.

Process of training and prediction involves use of specialized algorithms. We feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. One such algorithm is [K-Nearest-Neighbor](https://www.geeksforgeeks.org/k-nearest-neighbours/) classification (KNN classification). It takes a test data, and finds k nearest data values to this data from test data set. Then it selects the neighbor of maximum frequency and gives its properties as the prediction result

Objectives of Research

This project aims through machine learning techniques at creating a model, employee attrition classification. Data includes Education, job level, job involvement, Daily rate, Monthly Income, Experience, No. of Companies worked ,No of companies changed ,Total work experience, Years at company, Years at current role, Years since last promotion, Years with current manager, Training times last year, Performance rate, Attrition, By following above given data we finding the complete job profile of the person. By this data and by satisfying the above given data we know the sustainable growth. Judges will look for evidence of how employees are engaged in strategic goals employees data ad their experience in company and the level of the person in the company. And by the graph and the plots we can prove it graphically. And we can the no of employees working in the company. Celebrating organizations seeing the best return on an investment in people, this Award recognizes the employee strategy that best attracts, retains and develops talent – and how being an employer of choice contributes to and company values, and how human and supporting resources have been optimized to achieve and sustain commercial and competitive success. They will also look at investment in people, initiatives to create a more collaborative culture, and the impact this has had on commercial performance.

Review of literature

**Employee turnover (attrition) is a major cost to an organization, and predicting turnover is at the forefront of needs of Human Resources (HR) in many organizations.** Until now the mainstream approach has been to use logistic regression to model employee attrition.

The factors on which the Employee Attrition depends upon are:

1. *Age of the Employee*
2. *Monthly Income*
3. *Overtime*
4. *Monthly Rate*
5. *Distance from Home*
6. *Years at Company*

And so on…

A possible solution to solve this problem is by applying Machine Learning i.e., by imparting Machine Intelligence which involves development of a Predictive Model by training it, using the data available and validating it for Model Performance Analysis by using logistic regression

Data Collection

Attrition is a reduction in the workforce caused by retirement or resignation, without plans to fill or replace that vacant job position. Outside the HR context, the term attritionrefers to a reduction in strength or a weakening – which is likely the reason why the term has a negative connotation, even when there can be positive outcomes resulting from a reduction in staff.

<https://smallbusiness.chron.com/meaning-attrition-used-hr-61183.html>

<https://www.businessscience.io/business/2017/09/18/hr_employee_attrition.html>

<https://towardsdatascience.com/solving-staff-attrition-with-data-3f09af2694cd>

<https://hackernoon.com/a-machine-learning-approach-to-ibm-employee-attrition-and-performance-b5d87c5e2415>

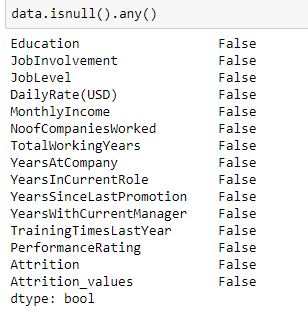
<https://www.wikihow.com/Calculate-Attrition-Rate>

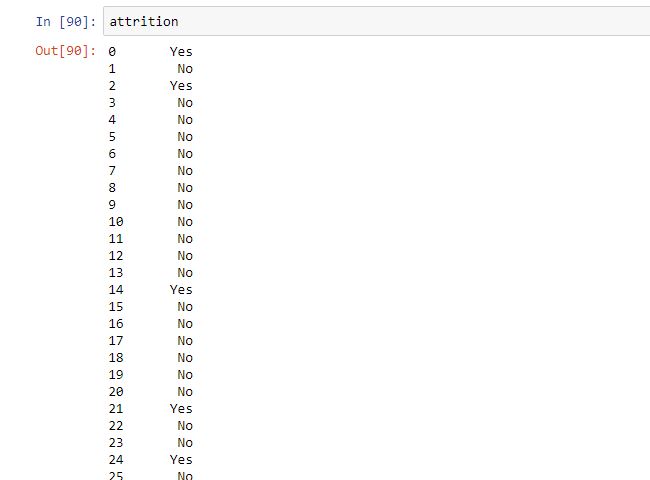
6 Methodology-

- **Exploratory Data Analysis**

**Data Cleaning:-**

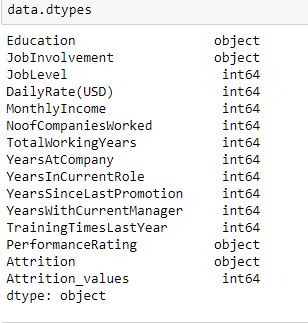
**Figure 6.1:-**

****

****

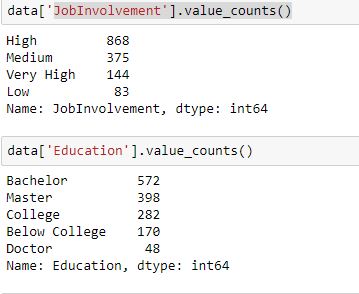
**Datatype:-**

**Figure 6.2:-**



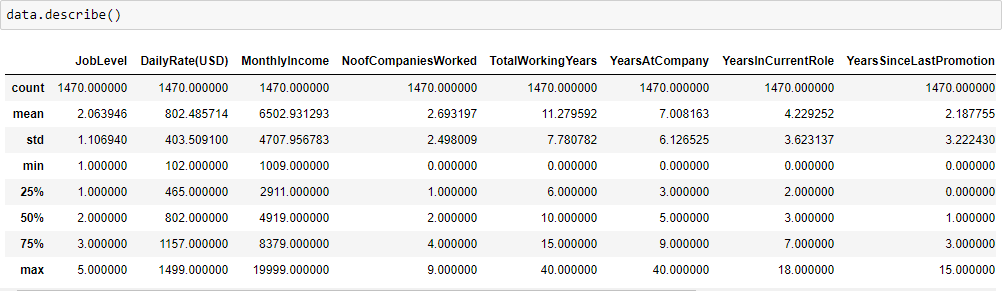
**Mode imputation for categorical cleaning:-**

**Figure 6.3 :-**



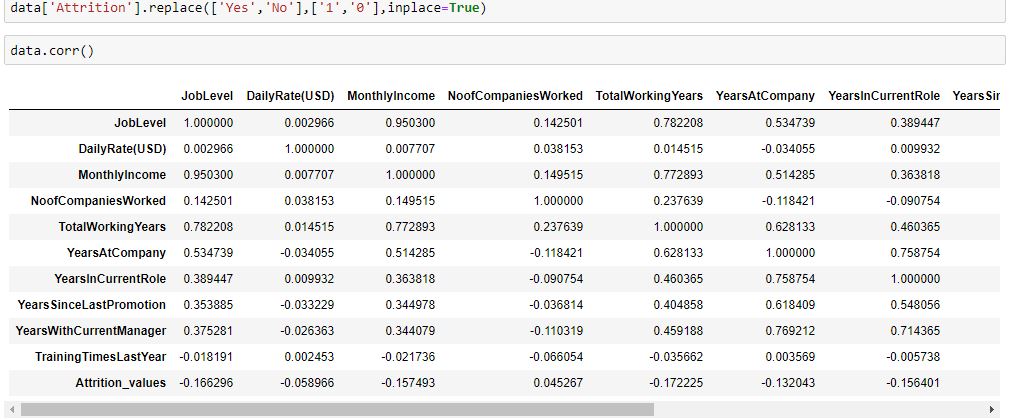
**Data Description:**

**Figure 6.4:-**



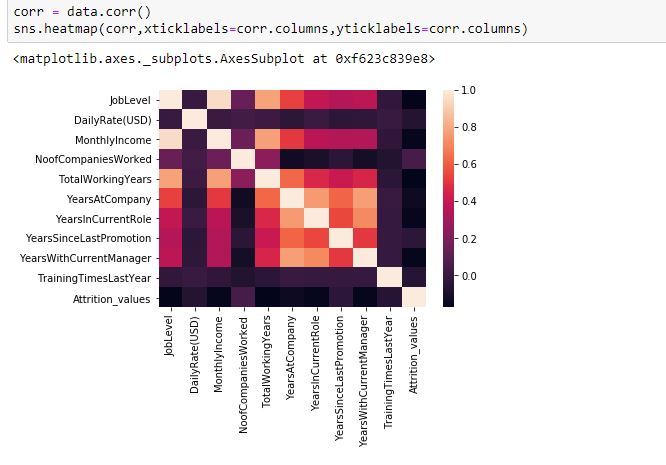
**Data Correlation:-**

**Figure 6.5:-**



**Heat Map**

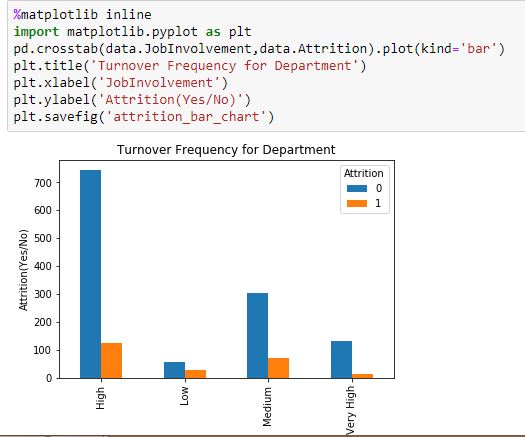
**Figure 6.6:-**

****

**Plots and Models**

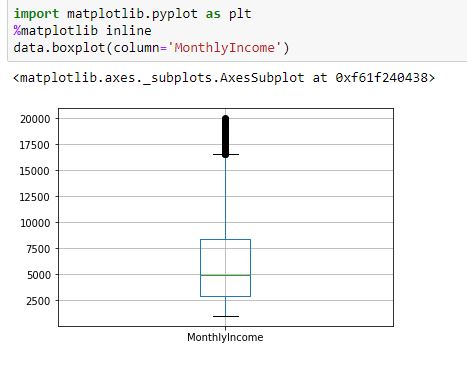
**Displots:**

**Figure 6.7:-**



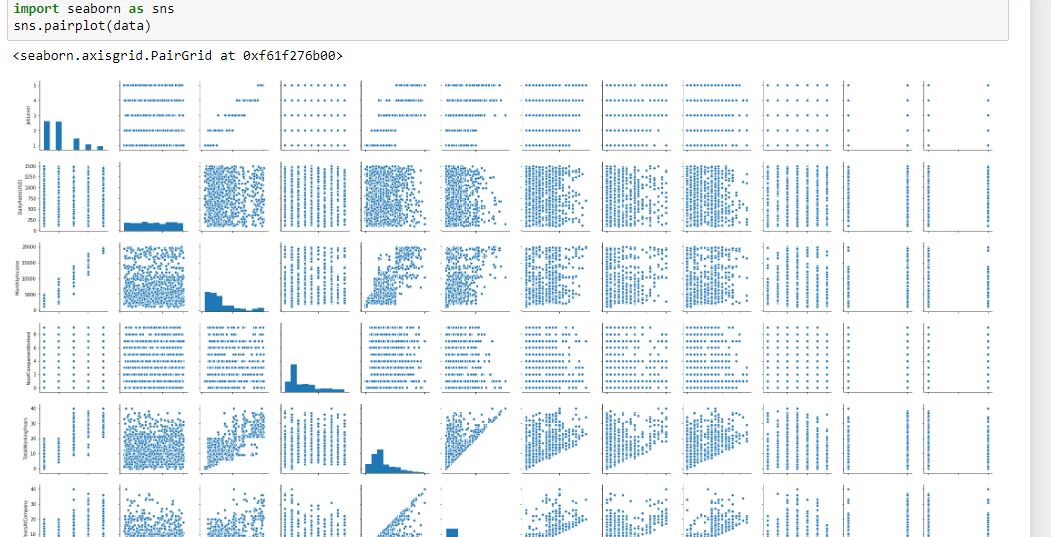
**Figure 6.8:-**

**BOX PLOT:**

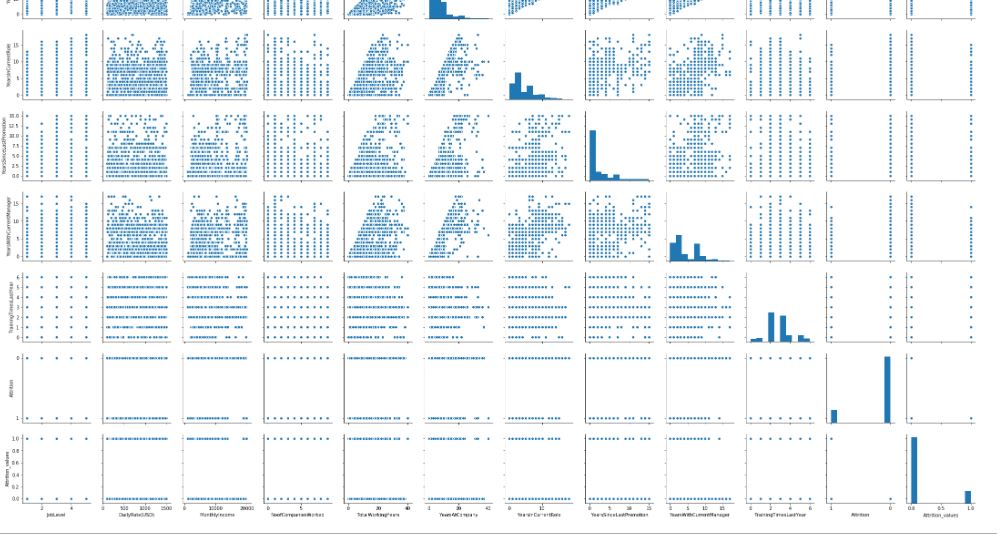


**PAIR PLOT:**

**Figure 6.9(A):-**

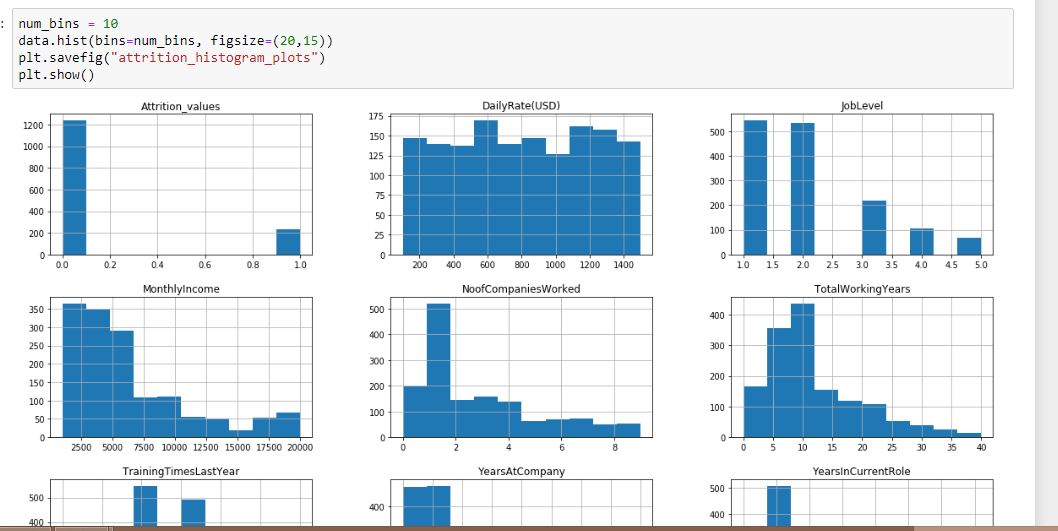
****

**Figure 6.9(B):-**

****

**SCATTER PLOT:**

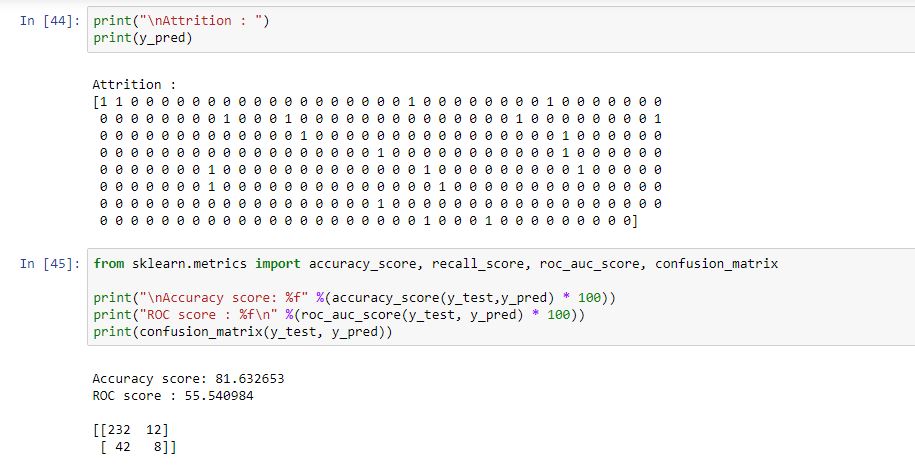
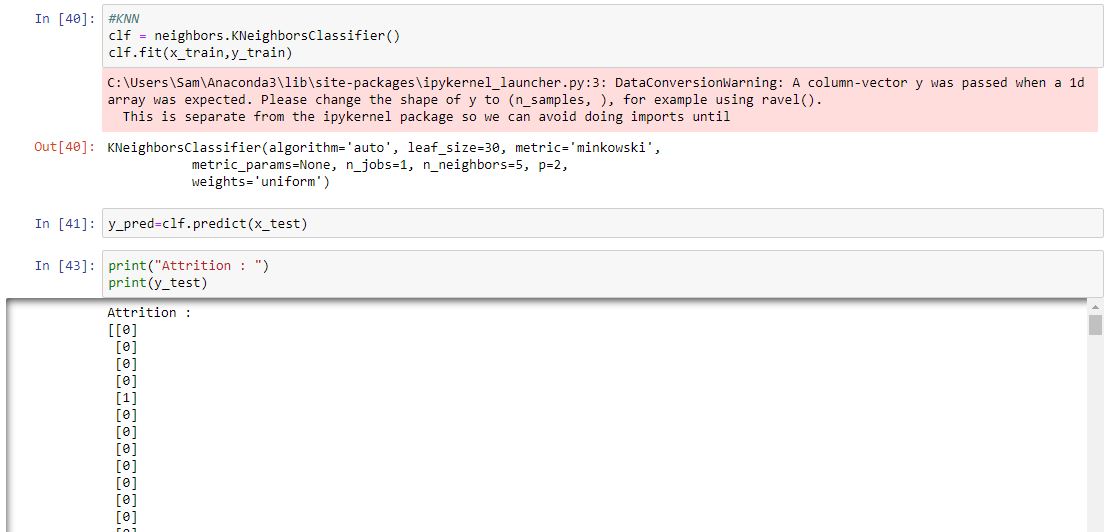
**Figure 6.9:-**

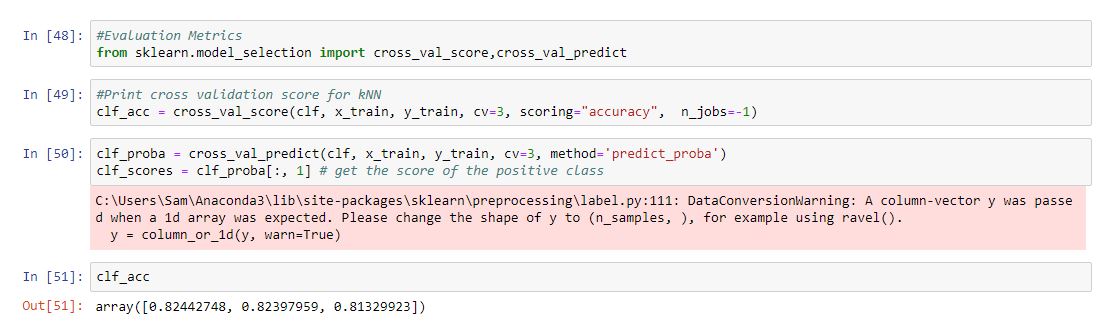
****

**ALGORITHMS**

**KNN:-**

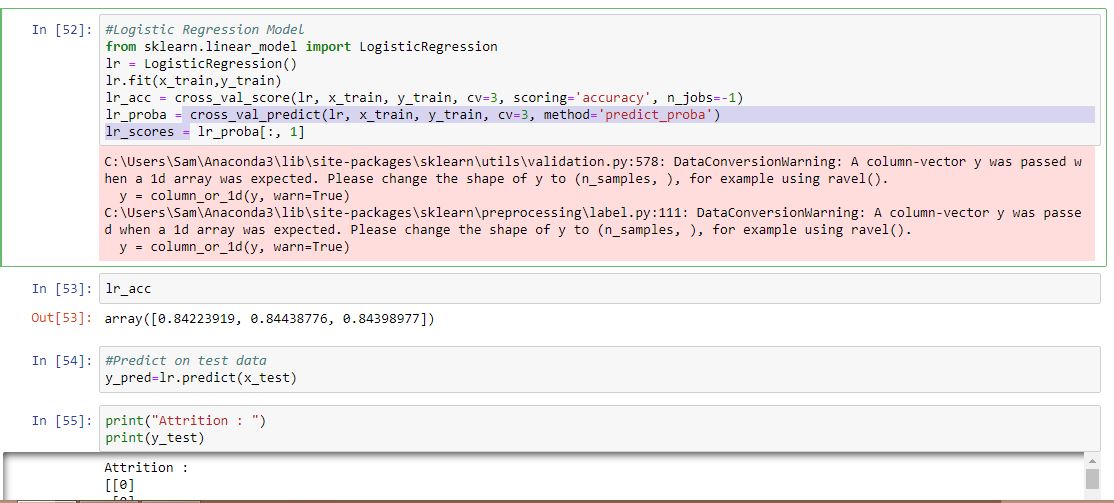
**Figure:7.1:-**

****

****

**LOGISTIC REGRESSION:**

**Figure-7.2:-**

****

****

**RESULT**

These are the following observations from the above models:

* The accuracy of KNN classifier on test set is:81.63%
* The accuracy of Logistic regression on test set is:82.99%

**CONCLUSION**

So, from the result we can say that Logistic regression, yielding the highest accuracy of 82.99% on Employee Attrition data set whereas KNN algorithm yields an accuracy of 81.63 %.