**HR Analytics Employee Attrition & Performance**

**Group Members : - Nikita Muley**

**Sharad Yadav**

**Shraddha Fulambarkar**

**1.Problem Statement and Purpose**

**Problem Statement:**

Employees are the backbone of the organization. Organization's performance is heavily based on the quality of the employees. Challenges that an organization has to face due employee attrition are:

* Expensive in terms of both money and time to train new employees.
* Loss of experienced employees
* Impact in productivity
* Impact profit

Attrition can happen for many reasons:

* Employees looking for better opportunities.
* A negative working environment.
* Bad management
* Sickness of an employee (or even death)
* Excessive working hours

**Purpose:**

The purpose of this project is to perform Classification analysis to determine target variable "Attrition". To determine whether an employee leaves the organization or not

So, what is Attrition and what determines it? It is basically the turnover rate of employees inside an organization.

## Import Libraries

The very first step is to import all the necessary libraries and Read the Dataset.

Following are the libraries used along with their versions.

| **Library** | **Version** |
| --- | --- |
| pandas | 1.5.3 |
| numpy | 1.24.1 |
| Seaborn | 0.12.2 |
| matplotlib | 3.6.3 |

**2.Data Gathering**

Reading the Data set will provide information about various features.

Dataset Consists of:

1. Independent Variables
2. Dependent Variable

Independent Variables in Dataset:

Age

BusinessTravel

DailyRate

Department

DistanceFromHome

Education

EducationField

EmployeeCount

EmployeeNumber

EnvironmentSatisfaction

Gender

HourlyRate

JobInvolvement

JobLevel

JobRole

JobSatisfaction

MaritalStatus

MonthlyIncome

MonthlyRate

NumCompaniesWorked

Over18

OverTime

PercentSalaryHike

PerformanceRating

RelationshipSatisfaction

StandardHours

StockOptionLevel

TotalWorkingYears

TrainingTimesLastYear

WorkLifeBalance

YearsAtCompany

YearsInCurrentRole

YearsSinceLastPromotion

YearsWithCurrManager

Dependent Variable in Dataset:

1. Attrition

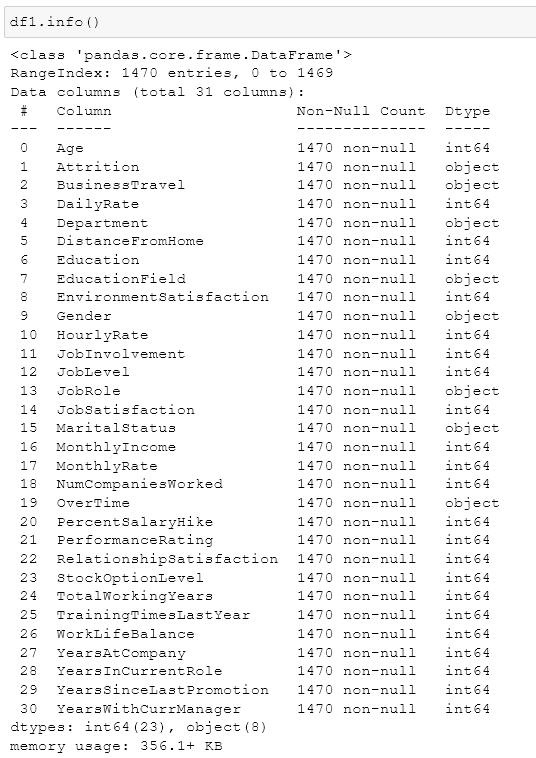
**3.EDA (Exploratory Data Analysis)**

a) Data Shape :-

Rows: 1470

Columns: 35

b) Null values :- No null values present.

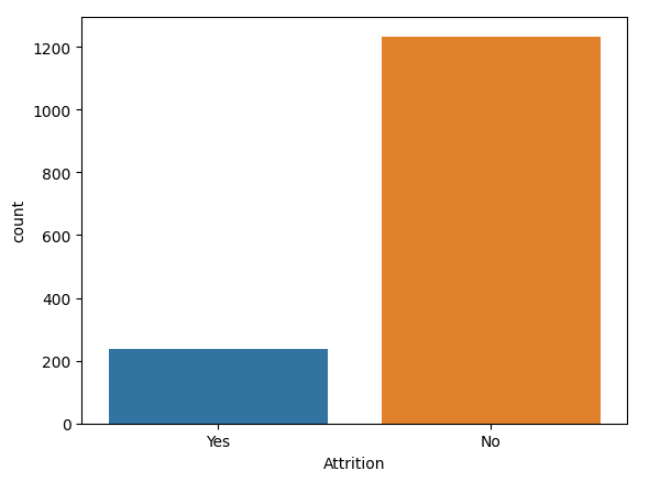
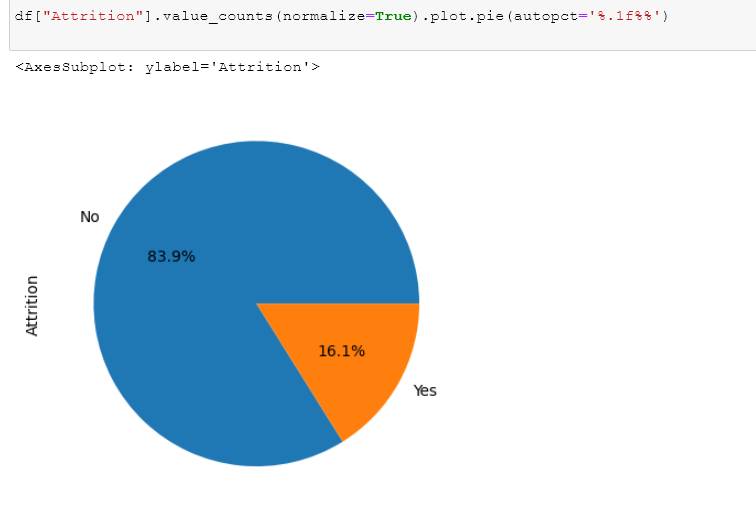
****

**Categorical Features**

* Attrition',
* 'BusinessTravel',
* 'Department',
* 'EducationField',
* 'Gender',
* 'JobRole',
* 'MaritalStatus',
* 'OverTime'

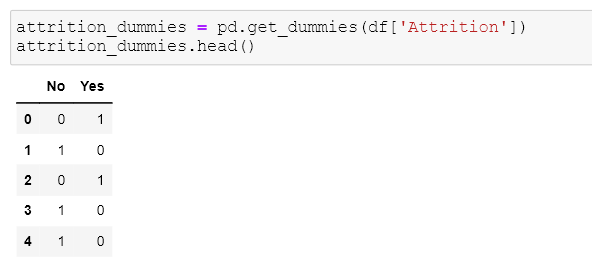
**Continuous Features**

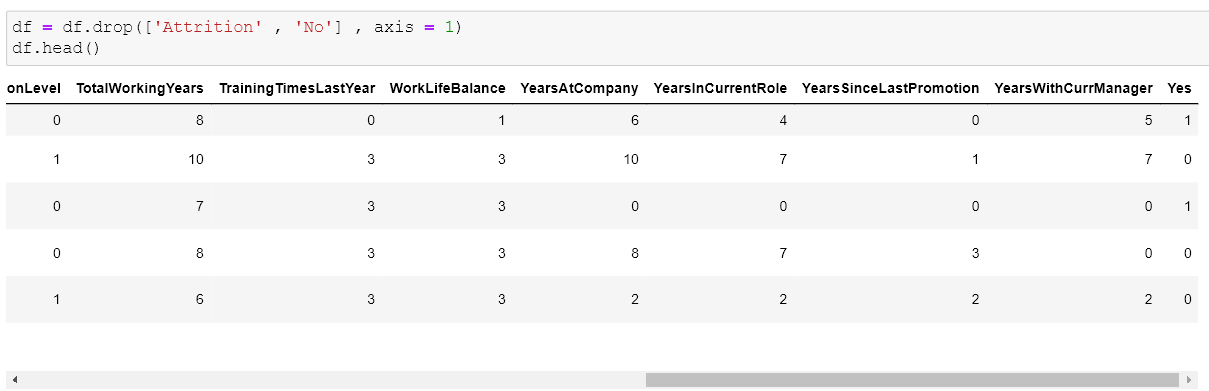
* ‘Age',
* 'DailyRate',
* 'DistanceFromHome',
* 'Education',
* 'EnvironmentSatisfaction',
* 'HourlyRate',
* 'JobInvolvement',
* 'JobLevel',
* 'JobSatisfaction',
* 'MonthlyIncome',
* 'MonthlyRate',
* 'NumCompaniesWorked',
* 'PercentSalaryHike',
* 'PerformanceRating',
* 'RelationshipSatisfaction',
* 'StockOptionLevel',
* 'TotalWorkingYears',
* 'TrainingTimesLastYear',
* 'WorkLifeBalance',
* 'YearsAtCompany',
* 'YearsInCurrentRole',
* 'YearsSinceLastPromotion',
* 'YearsWithCurrManager'

**Attrition:** 

**Here we conclude that data is highly Imbalanced**

**As we have to focus on Attrition so using get dummies Yes and No two separate feature created and then dropped the NO feature**

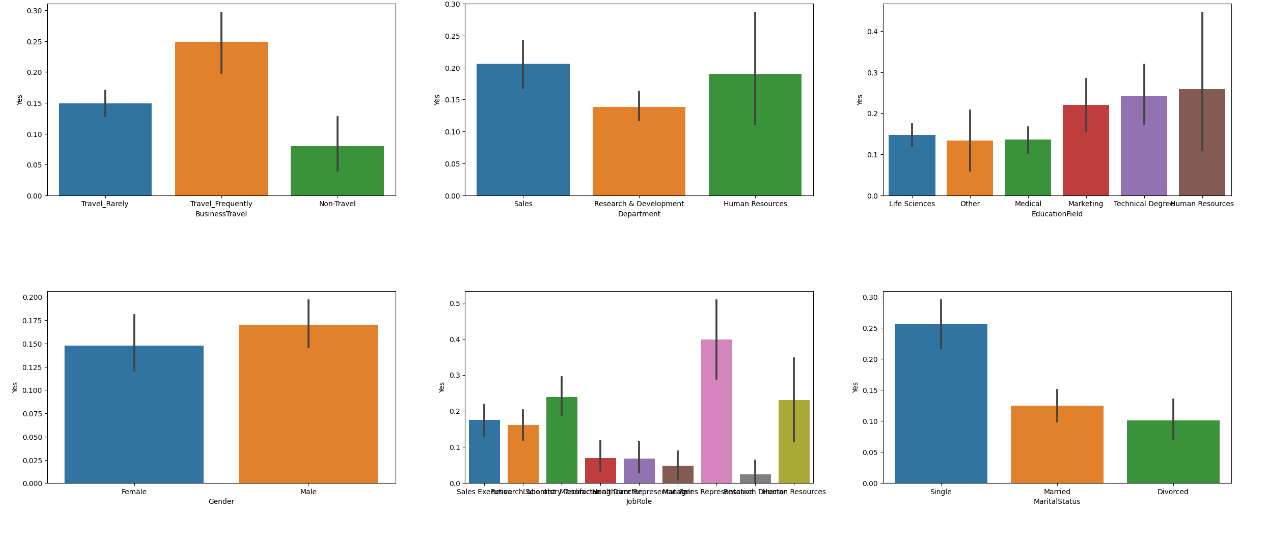


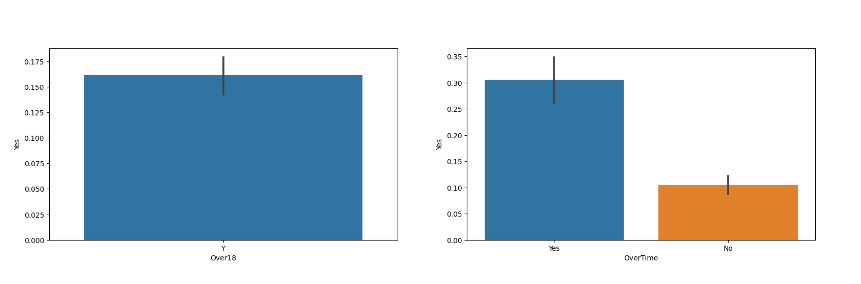


**Dropped Feature:**

EmployeeCount','EmployeeNumber','Over18', 'StandardHours']

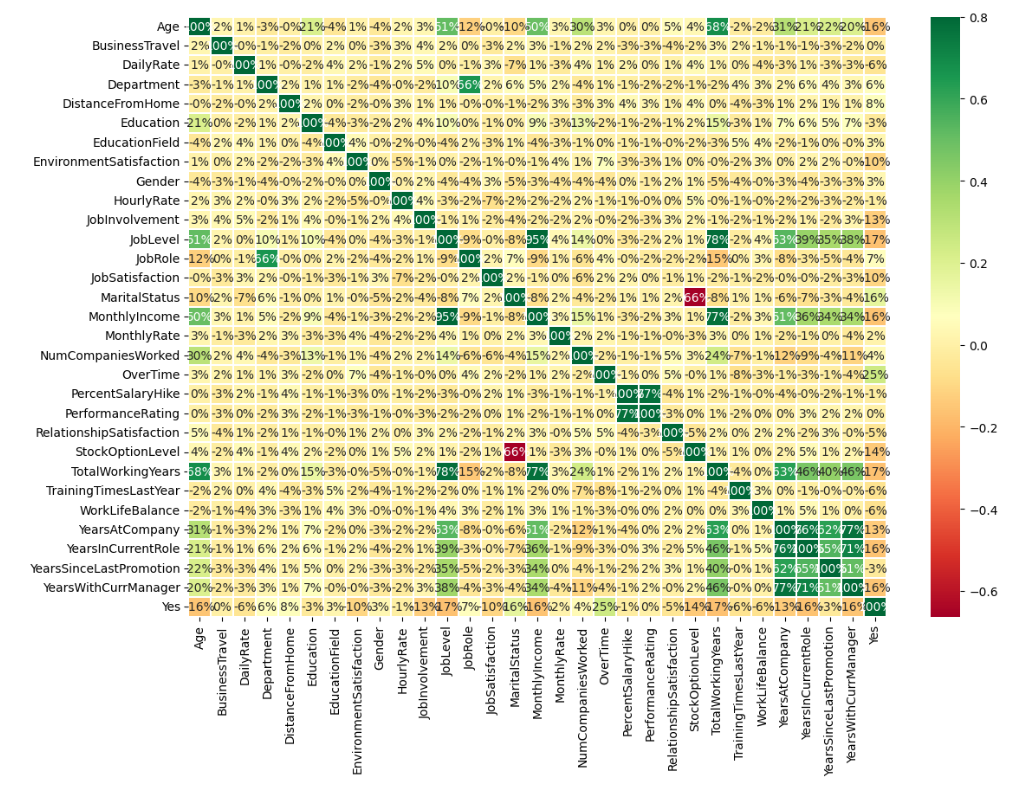
**Barplot : Categorical features VS ‘Yes’**

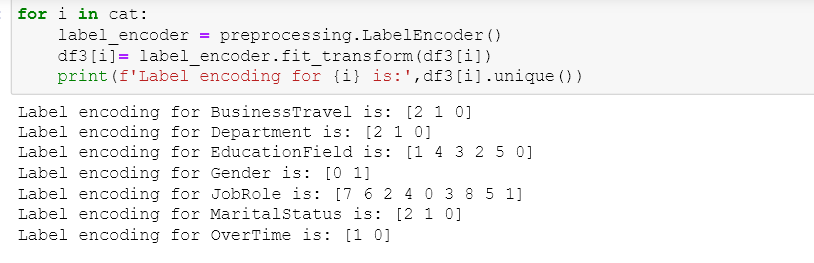




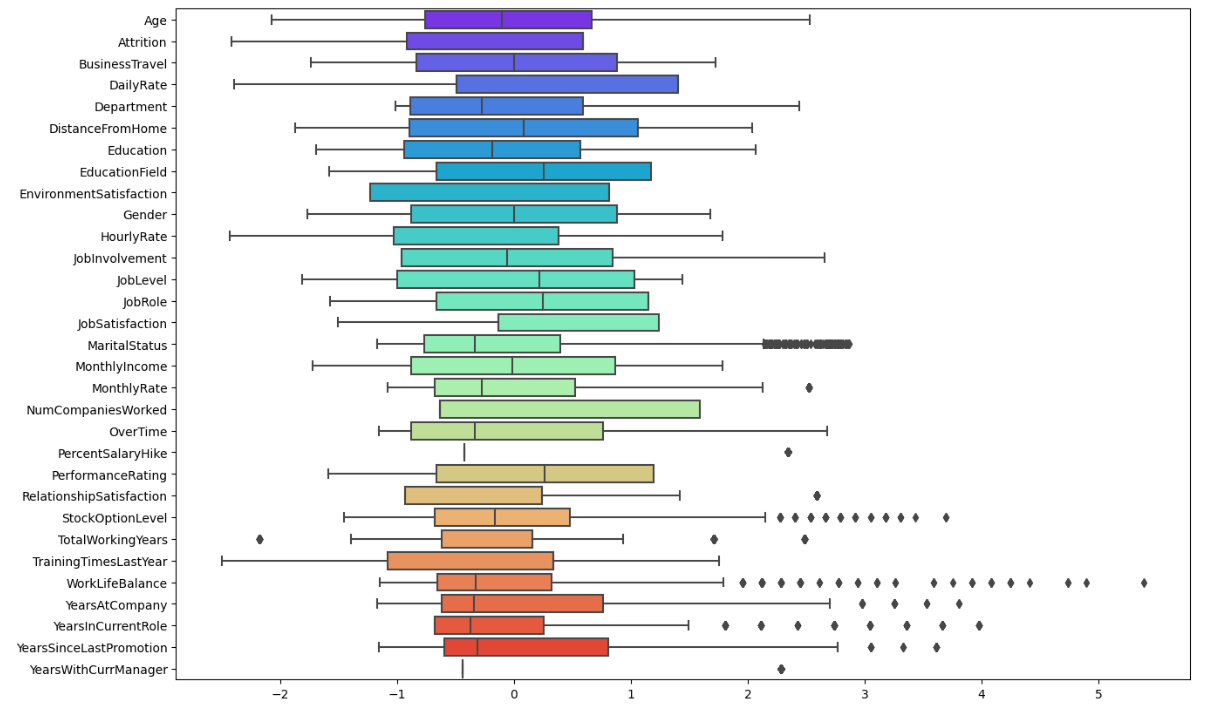
For Finding Correlation:

HeatMap



**Label Encoding for categorical feature:** 

Outliers Handling:



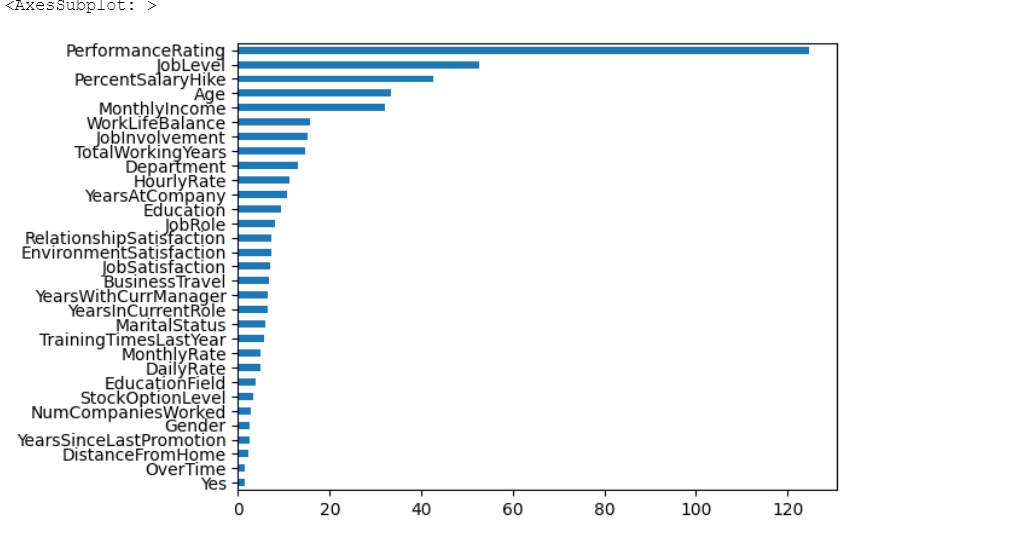
There are outliers present in following features:

1. MariatalStatus
2. MonthlyIncome
3. StockOptionLevel
4. TotalWorkingYear
5. WorkLifeBalance
6. YearAtCompany
7. YearSinceCurrentRole
8. YearSinceLastPromotion
9. YearWithCurrManager

**Outliers Handled Using IQR method**

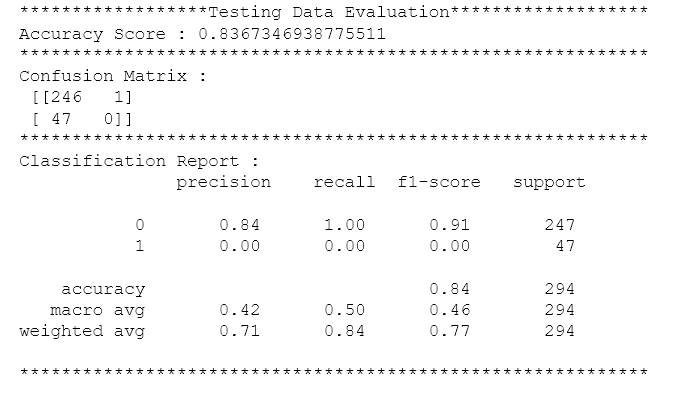
**Feature Selection**

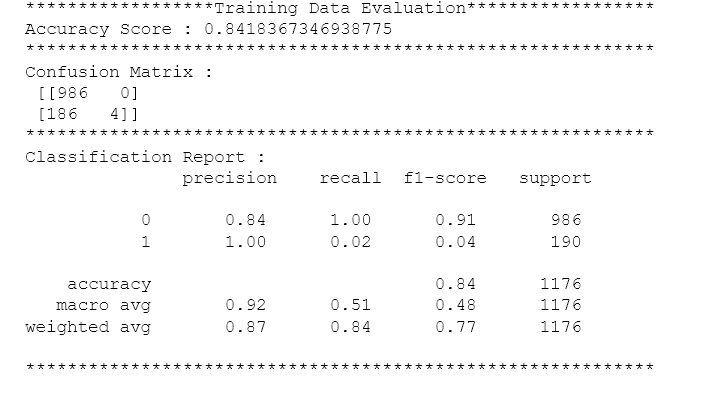
VIF:



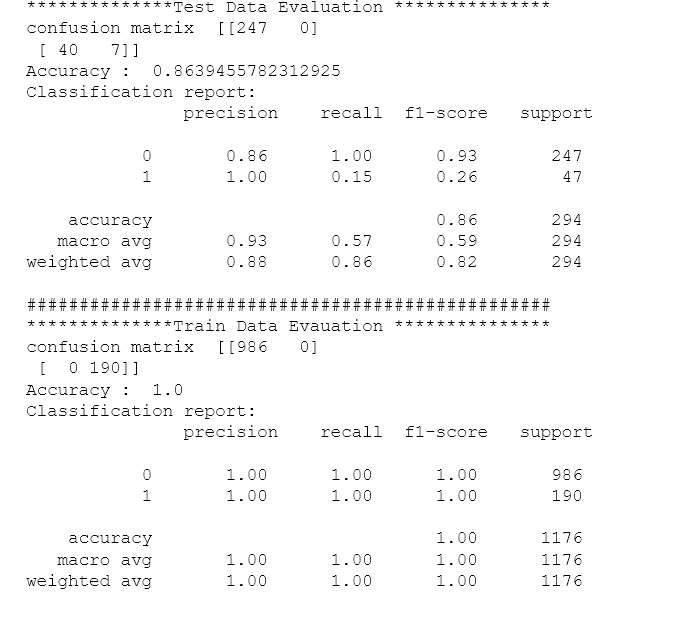
**Model Training:**

1. **Logistic Regression**

****

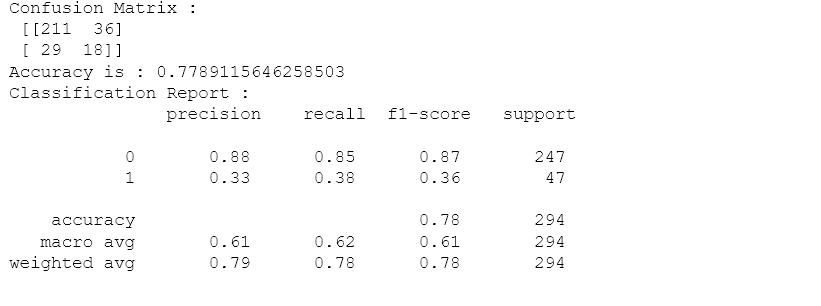
****

1. **Random Forest**

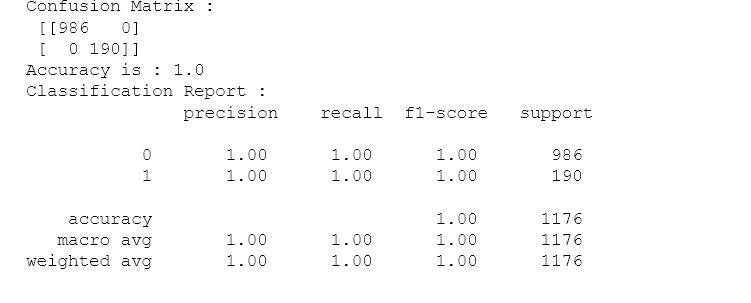
****

1. Decision Tree

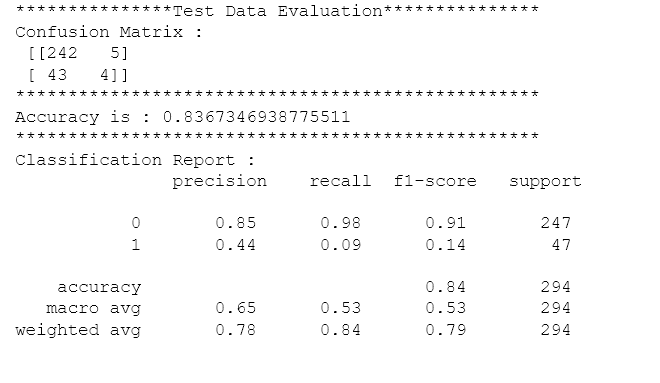
Testing Accuracy:



Training Accuracy:



1. **KNN Classifier:**



**Further Studies need to do:**

1. Need to do feature selection
2. Hyper Parameter tuning
3. Need to check accuracy before and after Oversampling