Report On Machine Learning Major Project

**Create a classification model to predict salary of a person**

Name : Akash L M

Email Id: [akashlmnmamit@gmail.com](mailto:akashlmnmamit@gmail.com)

Batch: February

Project Name: Machine Learning February Major Project

Topic : Machine Learning

Submitted on : 10 April 2022

**Project Description:**

**Problem statement:** Create a classification model to predict whether a person makes over $50k a year

**Context:** This data was extracted from the 1994 Census bureau database by Ronny Kohavi and Barry Becker (Data Mining and Visualization, Silicon Graphics).

**Details of features:**

The columns are described as follows:

1) Age

2) Workclass

3) Fnlwgt

4) Education

5) education\_num

6) marital\_status

7) occupation

8) relationship

9) race

10)sex

11)capital\_gain

12)capital\_loss

13)hours\_per\_week

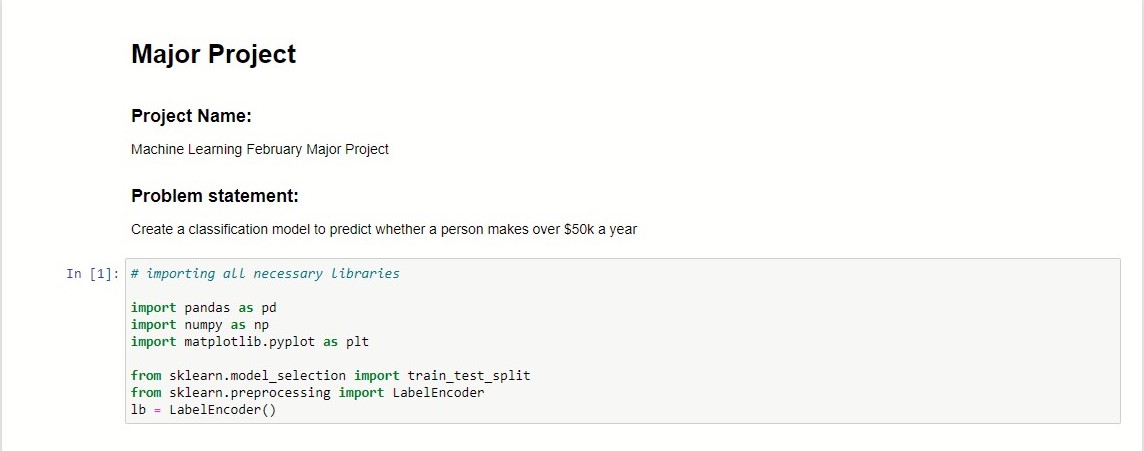
14)native\_country

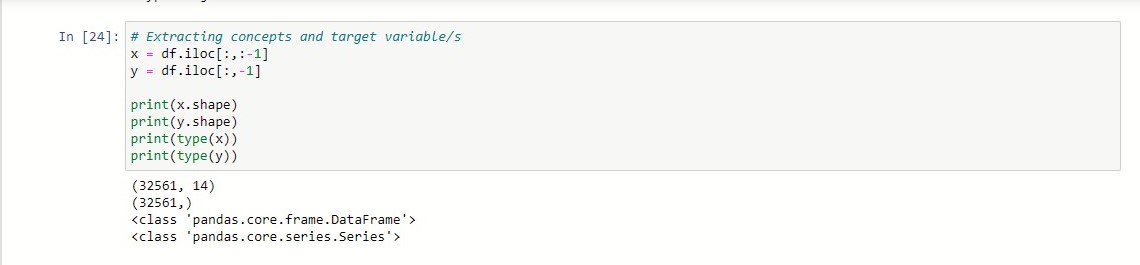
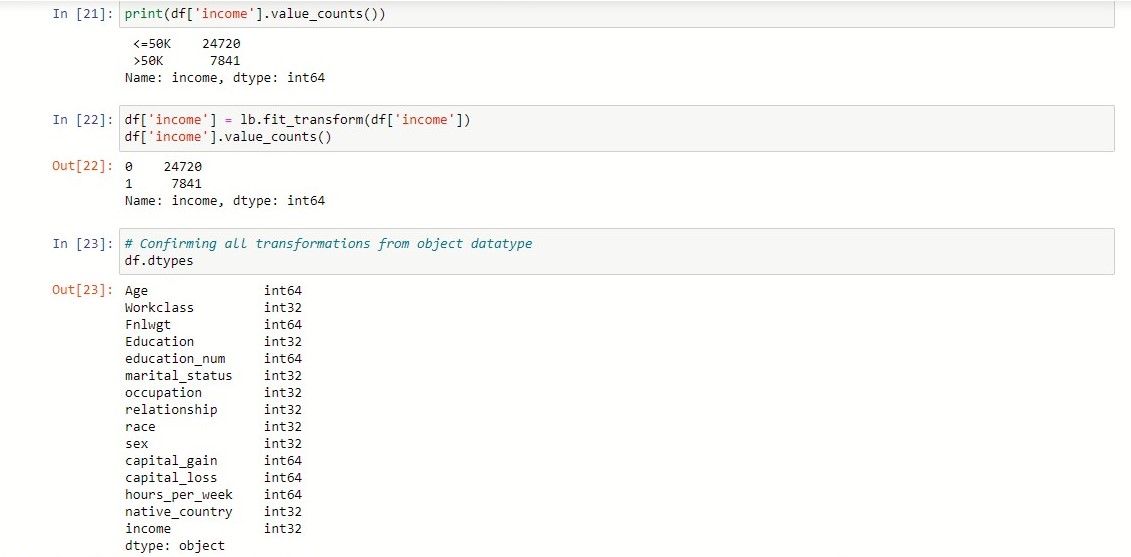
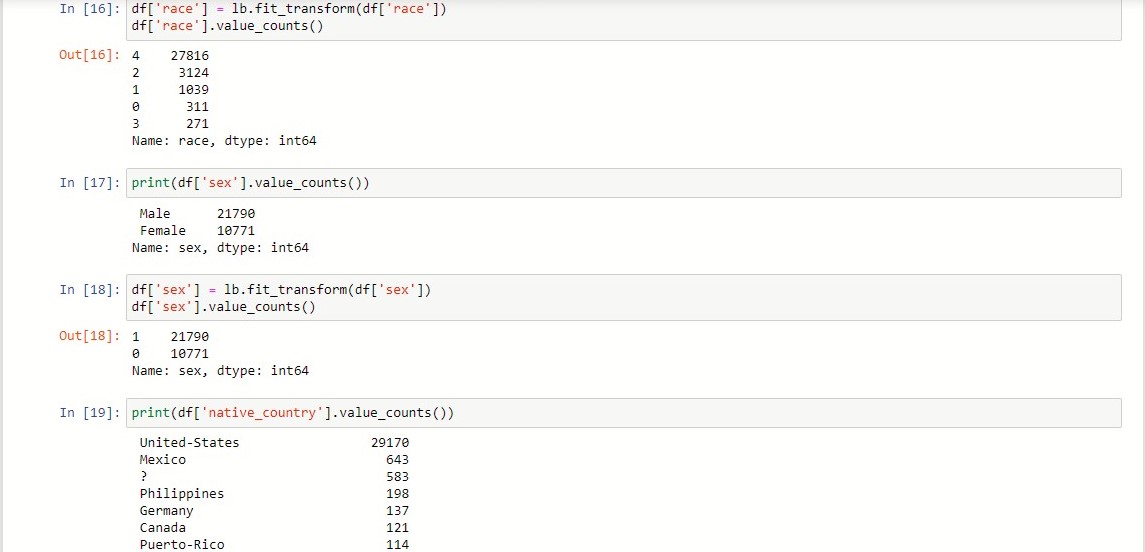
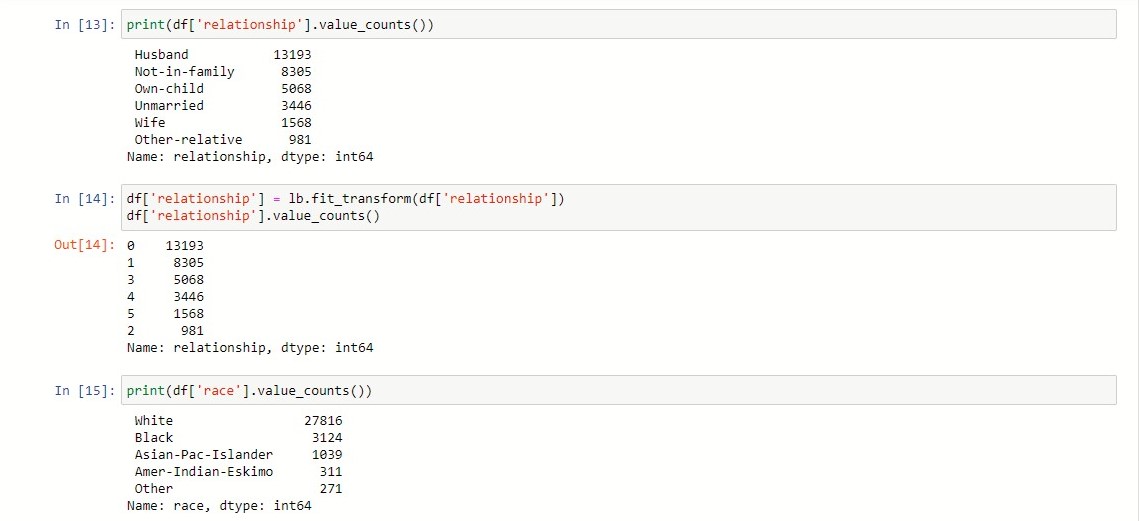
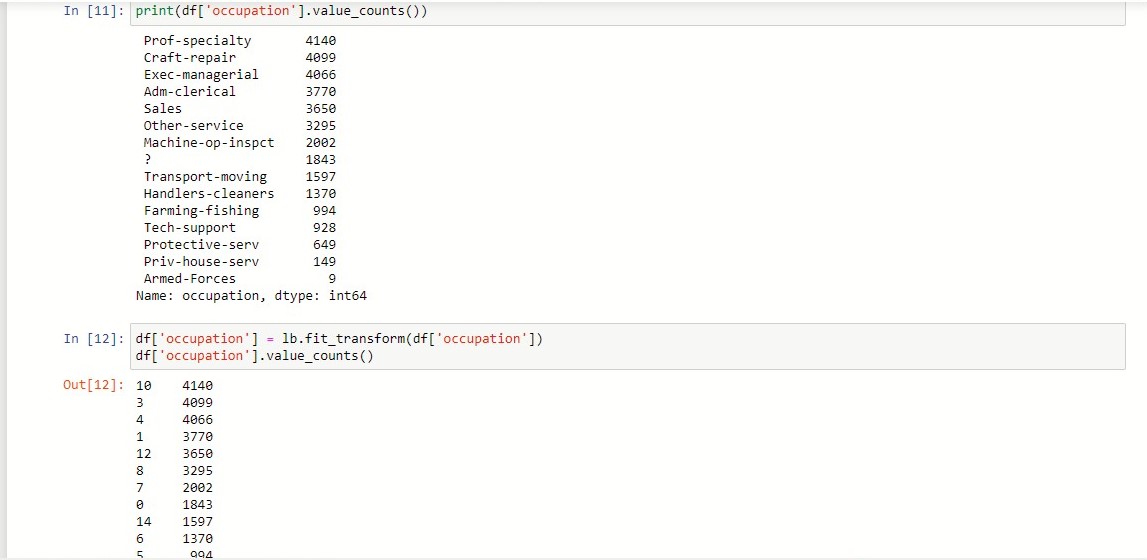
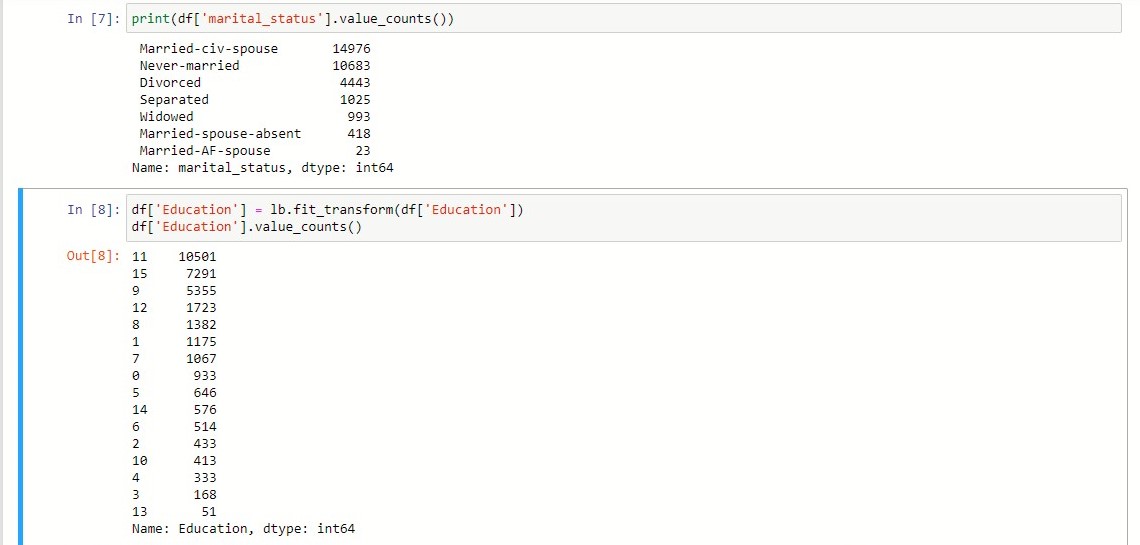
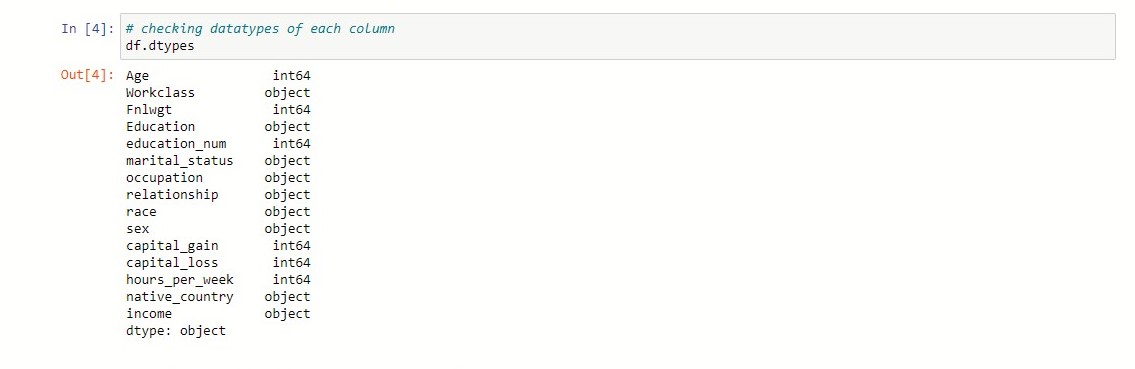
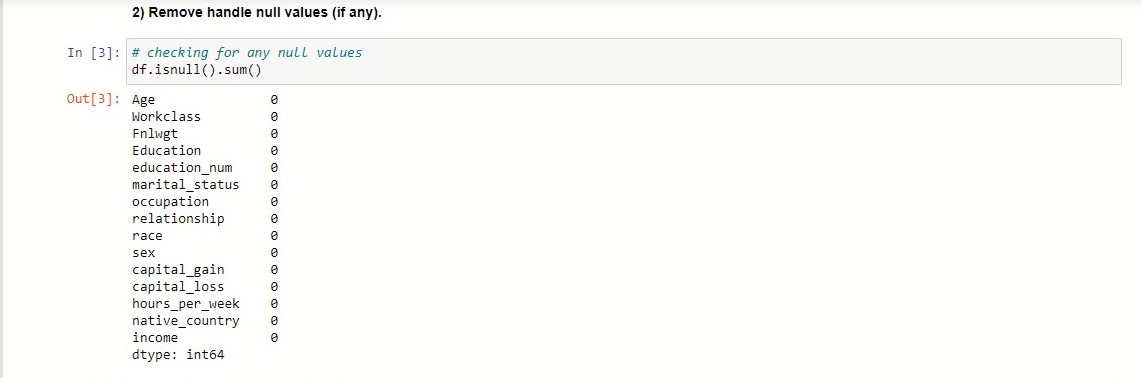
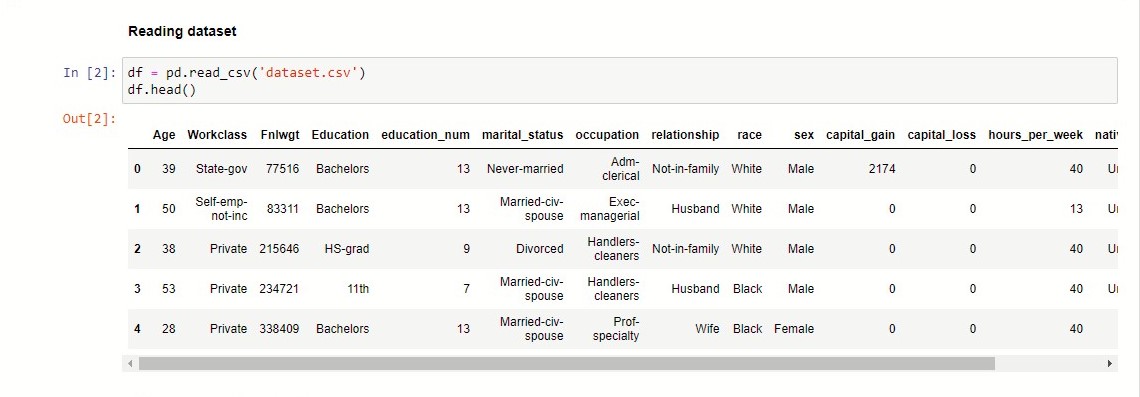
15)income

**Dataset:**

<https://drive.google.com/file/d/18rtxYghoy1ExiUOUJb6ywPsD0hQm6LMy/view?usp=sharing>

**Images :**

**1)Rename the columns**



2)Remove handle null values (if any). **Steps 2:**

3)Split data into training and test data



4)Apply the following models on the training dataset and generate the predicted value for the test dataset

a. Decision Tree

b. Random Forest Classifier

c. Logistic Regression

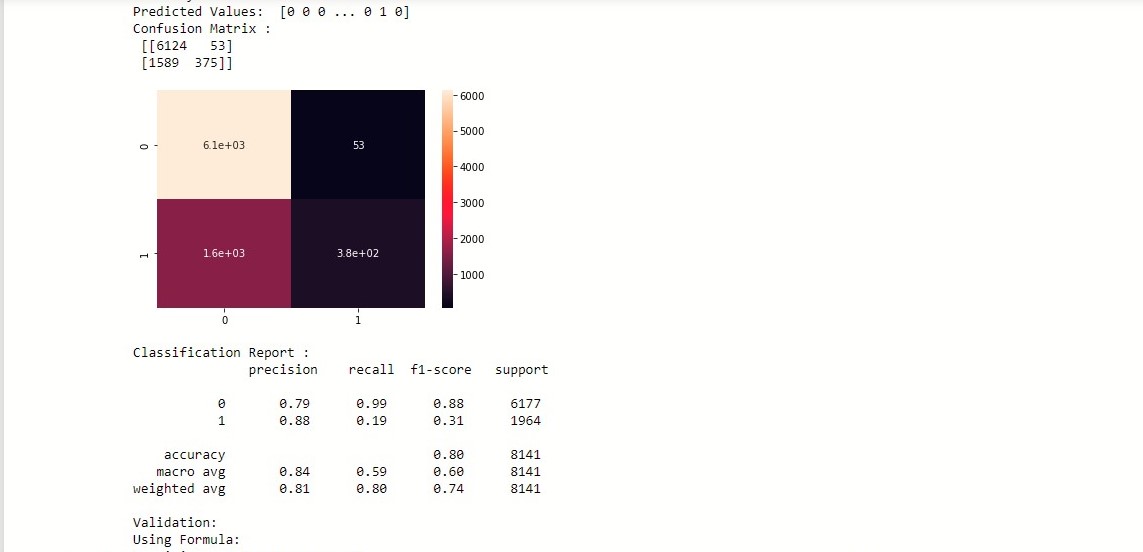
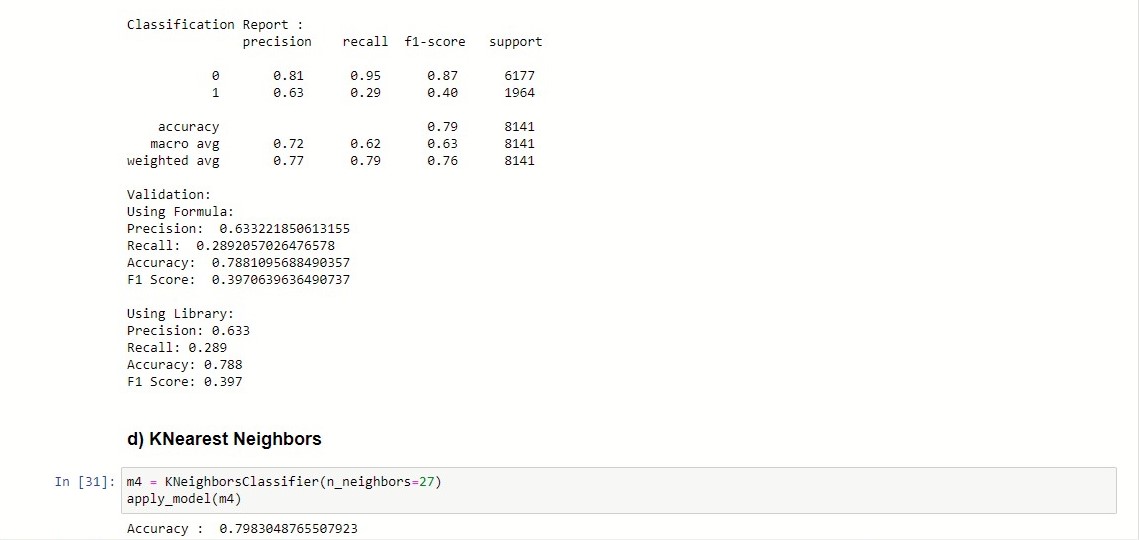
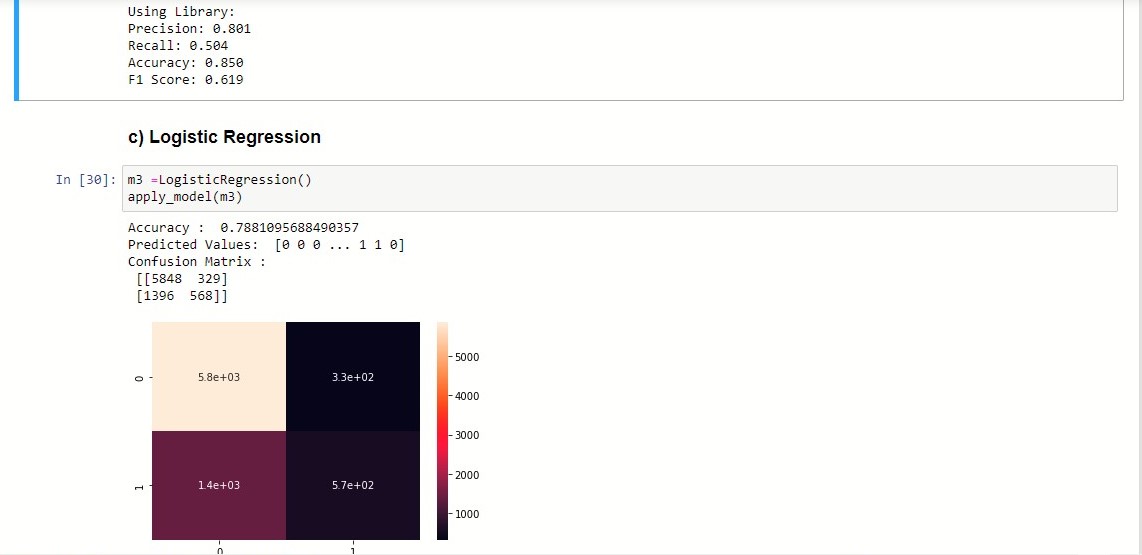
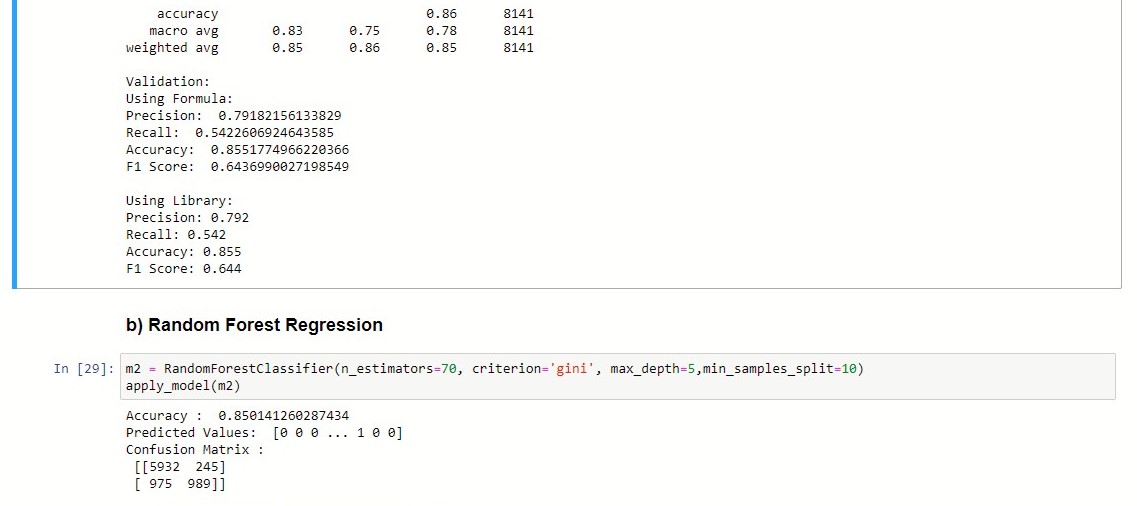
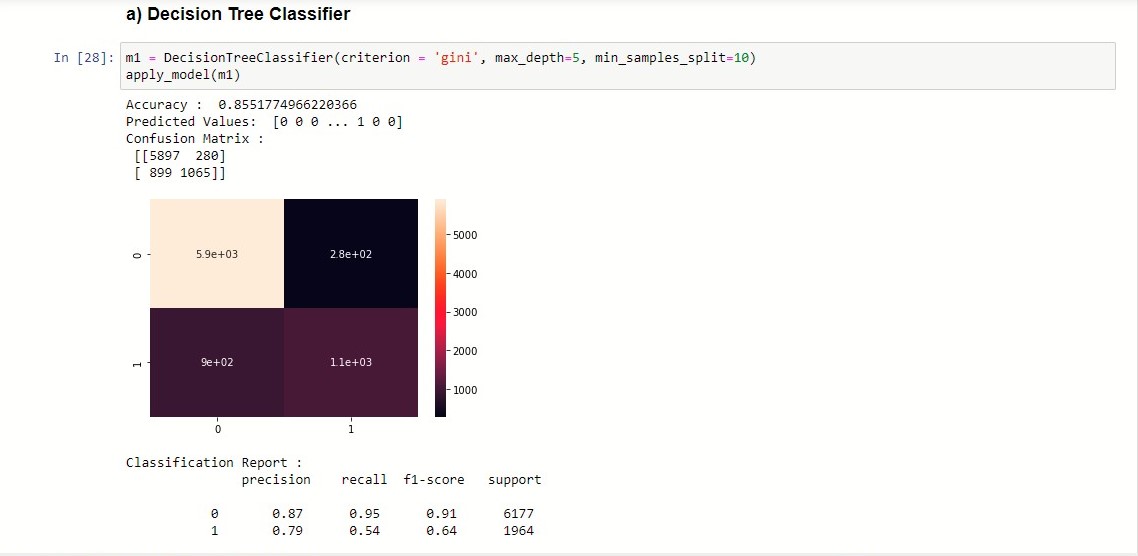
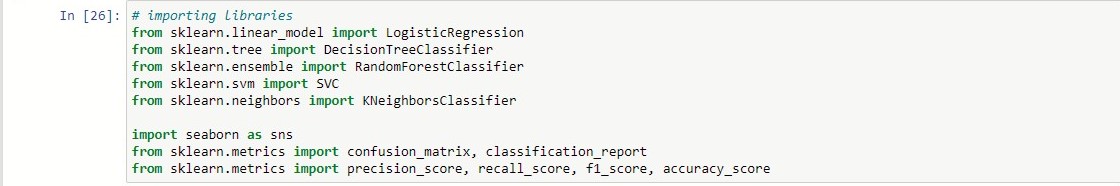
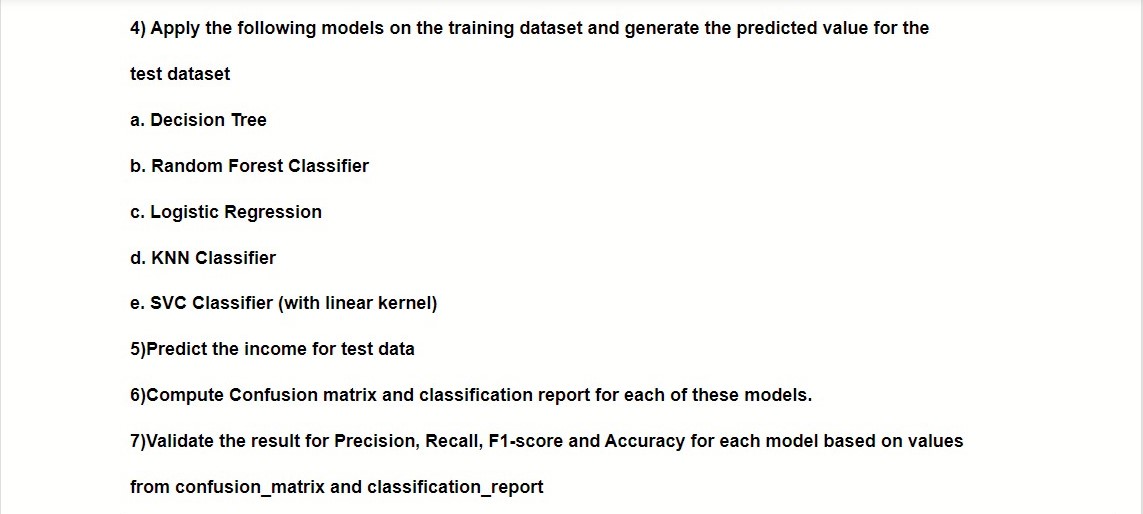
d. KNN Classifier

e. SVC Classifier (with linear kernel)

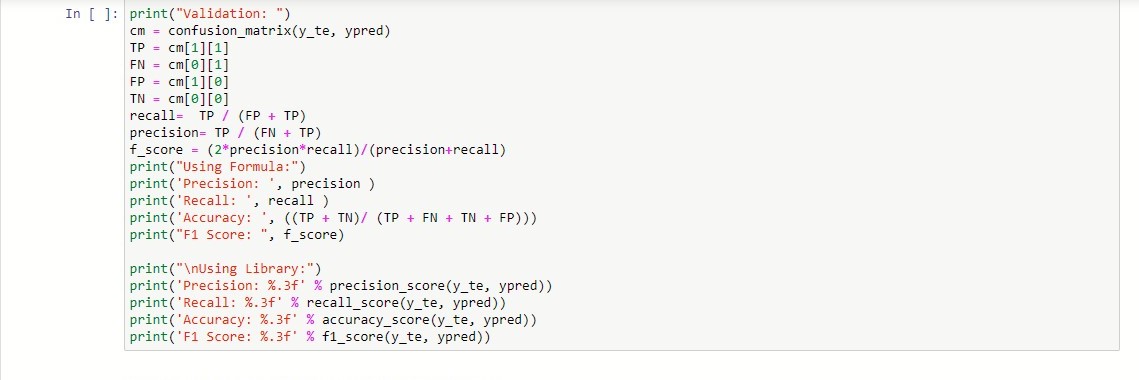
5)Predict the income for test data

6)Compute Confusion matrix and classification report for each of these models.

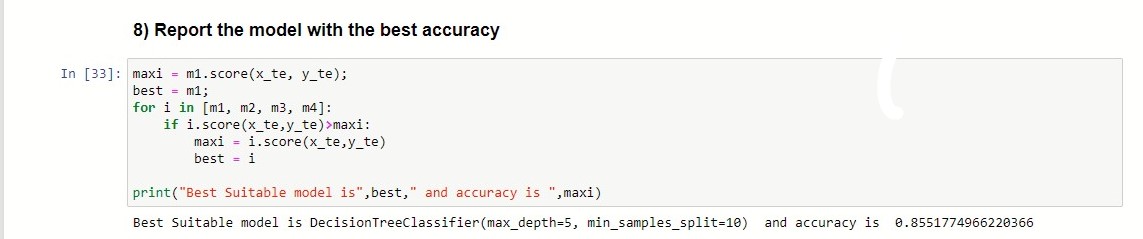
7)Validate the result for Precision, Recall, F1-score and Accuracy for each model based on values from confusion\_matrix and classification\_report



****



8)Report the model with the best accuracy



Best Suitable model is **DecisionTreeClassifier** and accuracy is **0.855177.**