**Machine Learning Model**

**RA1811030010023**

**Sayan Pramanick**

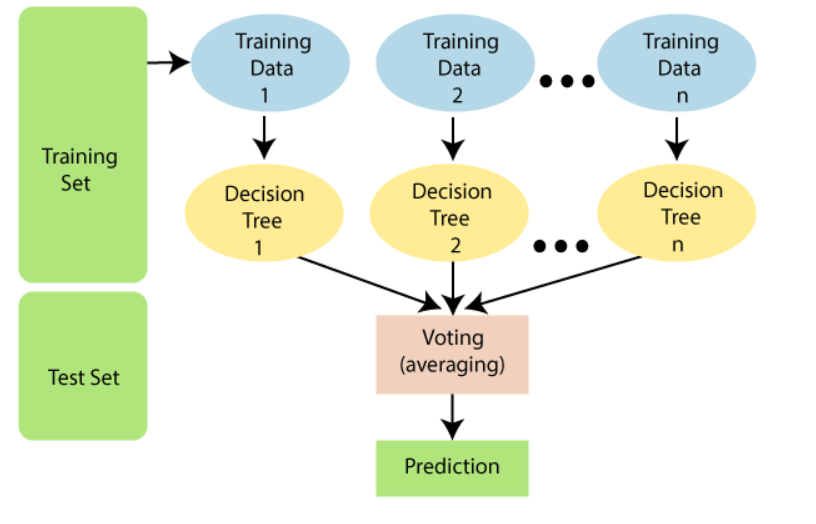
**Random Forests**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, **"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

**The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.**

The below diagram explains the working of the Random Forest algorithm:

****

**Code:**

import pandas

from sklearn import model\_selection

from sklearn.ensemble import RandomForestClassifier

url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.csv"

names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']

dataframe = pandas.read\_csv(url, names=names)

array = dataframe.values

X = array[:,0:8]

Y = array[:,8]

seed = 7

num\_trees = 100

max\_features = 3

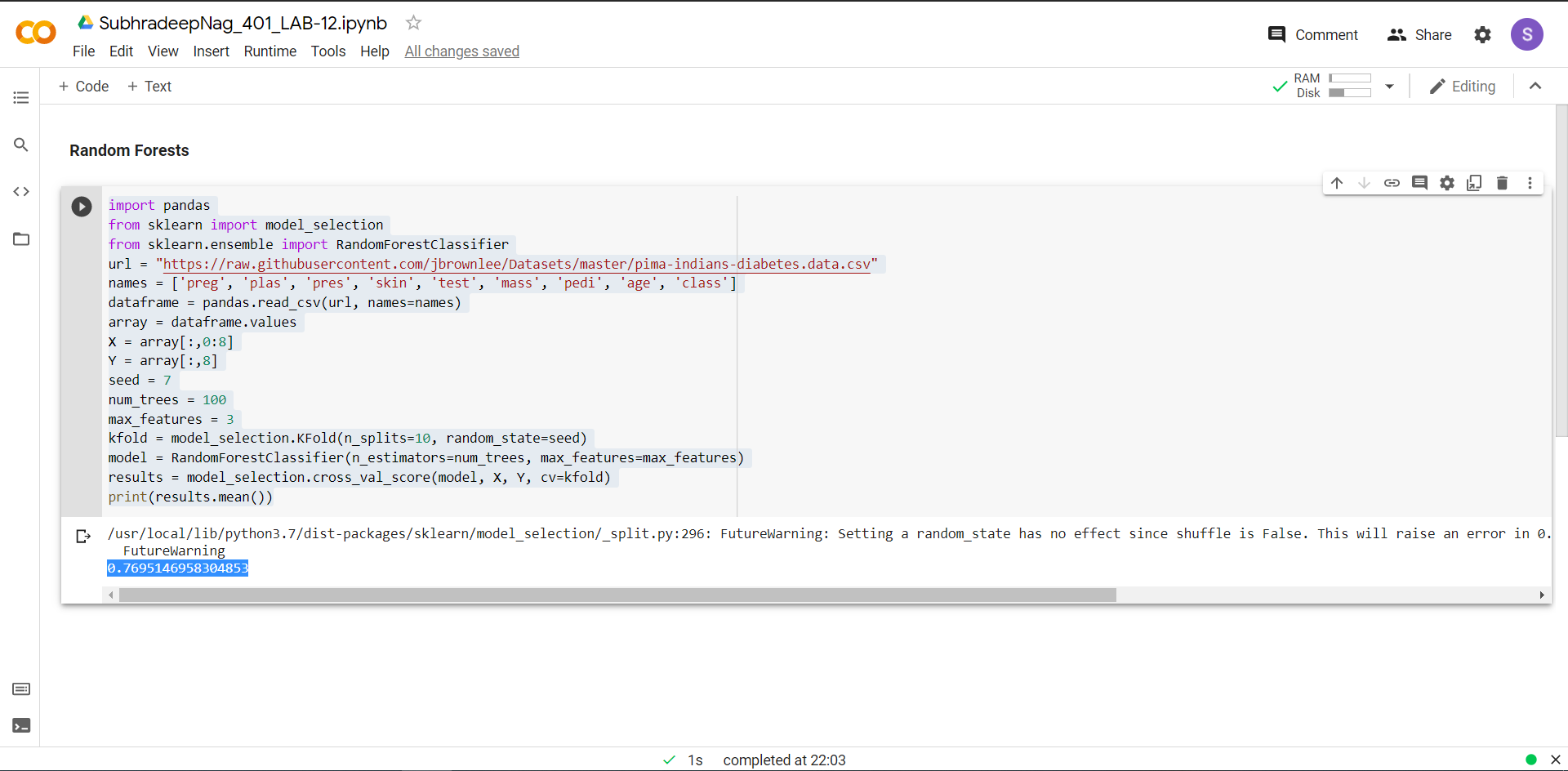
kfold = model\_selection.KFold(n\_splits=10, random\_state=seed)

model = RandomForestClassifier(n\_estimators=num\_trees, max\_features=max\_features)

results = model\_selection.cross\_val\_score(model, X, Y, cv=kfold)

print(results.mean())

**Output:**



Ans - 0.7695146958304853