Diabetes predection project



Supervised By:

prof/ Hala Zayed

Student Name:

Maryam Gaber Ahmed

**Diabetes Prediction**

**Introduction:**

**Machine Learning to predict diabetes**

The objective of the dataset is to diagnostically predict if a patient has diabetes, established on definite diagnostic quantities incorporated in the dataset.

Content: The datasets consist of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

Dependencies include python libraries like:

1-sklearn

2-pandas

3-matpltlib

4-Numpy

**Algorithms used:**

1-Decition Tree

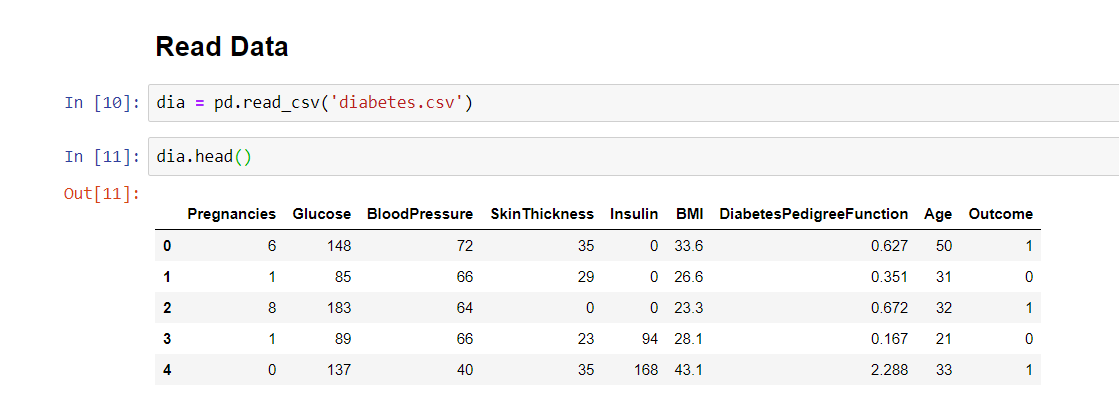
2-Random forest

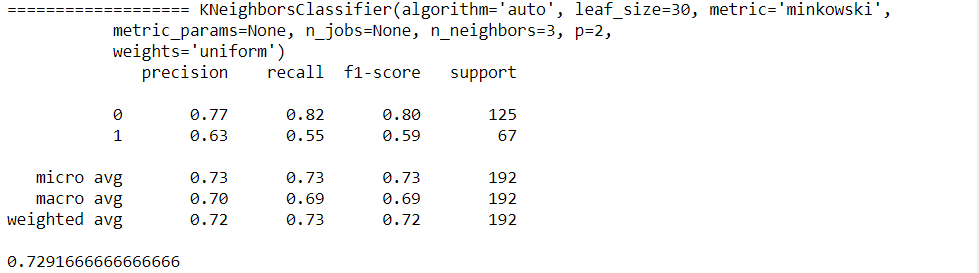
3-Logistic Regression

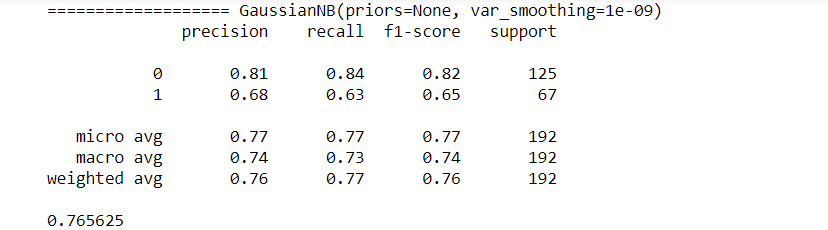
4-Niave Bayes

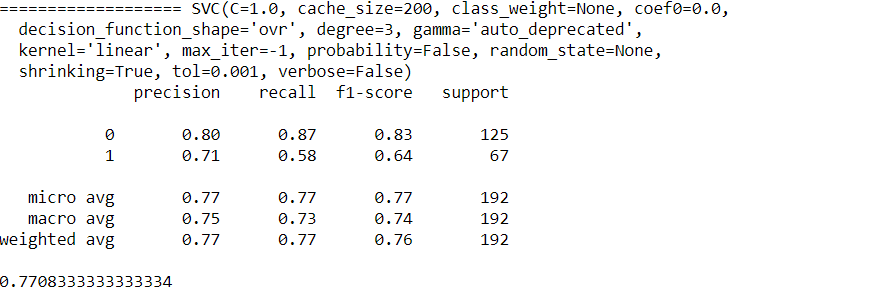
5- K Nearest

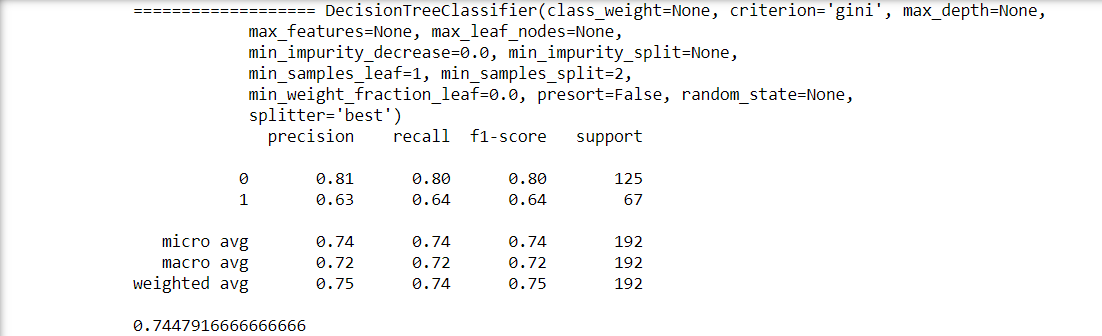
**Screen shots of the run:**

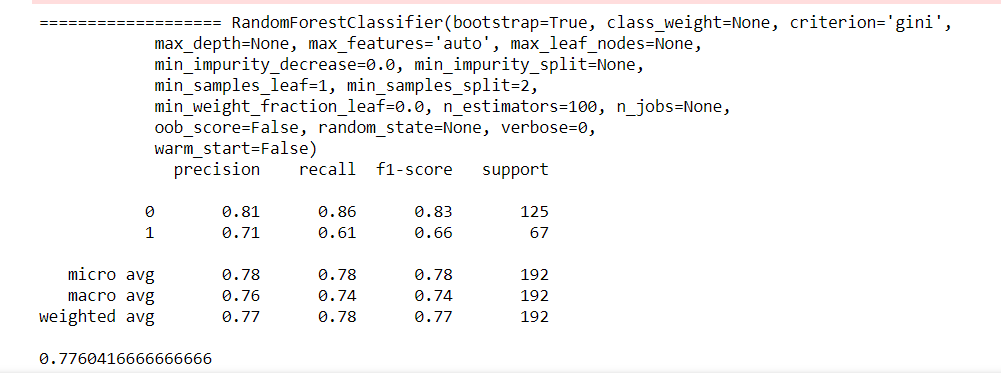
****

****

****

****

****

****

**Result:**

Decision Tree %74

Random Forest %77

SVC (kernel='linear’) %76

Logistic Regression %77

Naive Byes 76%

KNeighbors 72%

**Discussion:**

Accuracy in Decision tree Algorithm was 74%

Accuracy in Random Forest Algorithm was 77%

Accuracy in SVM Algorithm was 76%

Accuracy in Logistic Regression Algorithm was 77%

Accuracy in Naive Bayes Algorithm was 76%

Accuracy in K Nearest Algorithm was 72%

**Conclusion:**

Random forest and logistic Regression the best Algorithms in this case.

**Data set description**

1-Data source: Kaggle

2-Data Name: Pima Dataset

3-This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

4-Content:

The datasets consist of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

5-Features:

1)Pregnancies Number of times pregnant

2)Glucose Plasma glucose concentration 2 hours in an oral glucose tolerance test

3)Blood Pressure Diastolic blood pressure (mm Hg)

4)Skin Thickness Triceps skin fold thickness (mm)

5)Insulin:2-Hour serum insulin (mu U/ml)

6)BMI: mass index (weight in kg/ (height in m) ^2)

7)Diabetes Pedigree Function pedigree function

8)Age (years)

9)Outcome Class variable (0 or 1) 268 of 768 are 1, the others are 0

**References**

Git: <https://github.com/kashyap16/Healtcare-Diabetes-Prediction>

Kaggle Dataset: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>