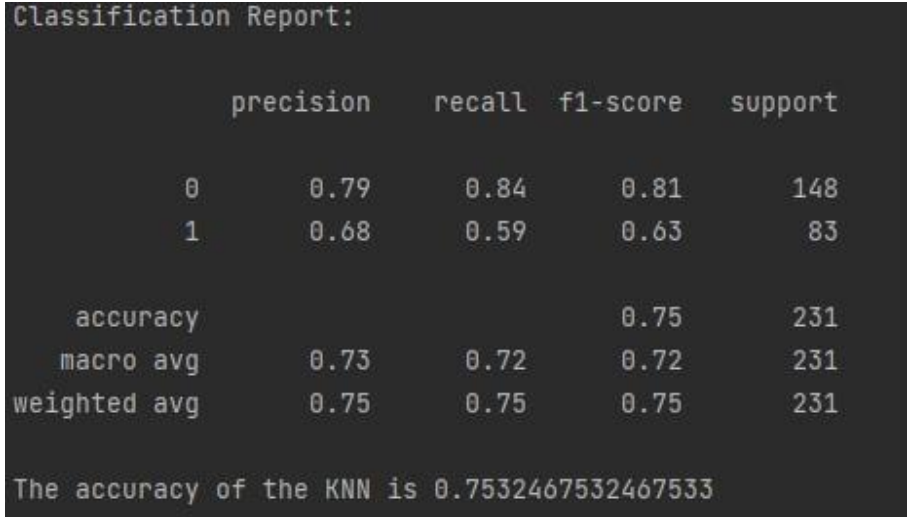


ML Major Project Report

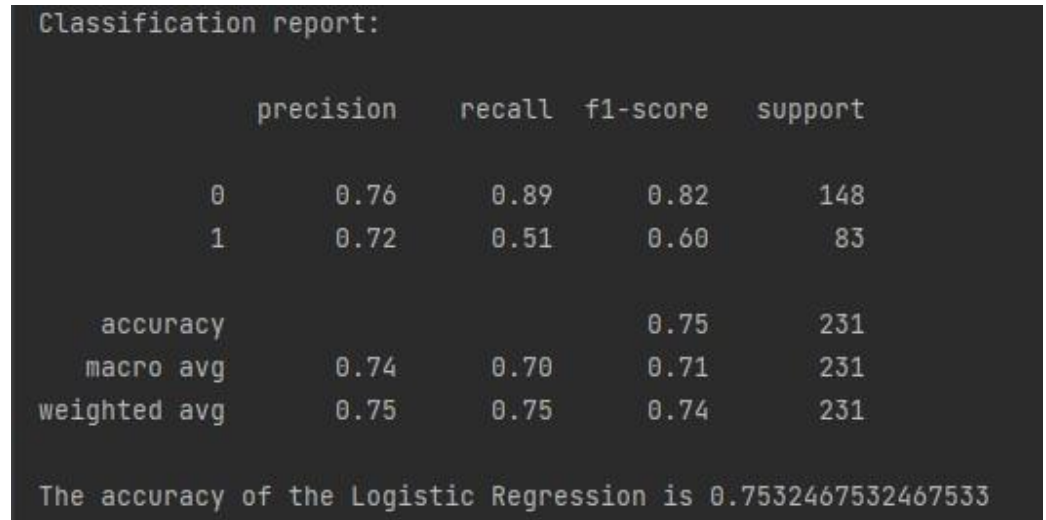
Done by – Bhandha Bharadwaj.
Date – 20/08/2021

Problem statement	<p>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.</p>
Tool used	<p>Ide – PyCharm Modules – Pandas Sklearn Matplotlib Seaborn.</p>
Algorithms used	<p>KNN Algorithm. Logistic Regression Algorithm. SVM Algorithm.</p>
Algorithm Explanation	<p>1) KNN Classification Algorithm:</p> <p>K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on the Supervised Learning technique. KNN algorithm can be used for Regression as well as for Classification but mostly it is used for Classification problems. Screenshot of the output-</p>  <pre>Classification Report: precision recall f1-score support 0 0.79 0.84 0.81 148 1 0.68 0.59 0.63 83 accuracy 0.75 0.75 0.75 231 macro avg 0.73 0.72 0.72 231 weighted avg 0.75 0.75 0.75 231 The accuracy of the KNN is 0.7532467532467533</pre>

2) Logistic Regression Algorithm

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

Screenshot of the output



```
Classification report:

              precision    recall  f1-score   support

     0       0.76       0.89       0.82       148
     1       0.72       0.51       0.60        83

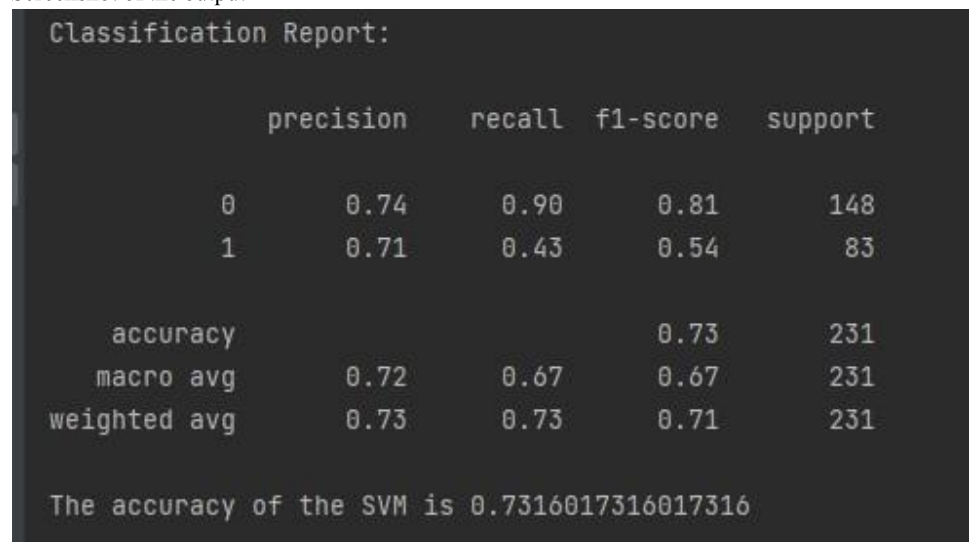
 accuracy          0.75       231
 macro avg       0.74       0.70       0.71       231
 weighted avg    0.75       0.75       0.74       231

The accuracy of the Logistic Regression is 0.7532467532467533
```

2) SVM Classification Algorithm

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

Screenshot of the output



```
Classification Report:

              precision    recall  f1-score   support

     0       0.74       0.90       0.81       148
     1       0.71       0.43       0.54        83

 accuracy          0.73       231
 macro avg       0.72       0.67       0.67       231
 weighted avg    0.73       0.73       0.71       231

The accuracy of the SVM is 0.7316017316017316
```

Conclusion

The best algorithm for the diabetes dataset is Logistic regression Algorithm and K Nearest Neighbor Algorithm because it has more accuracy than the SVM algorithm.

