**Diabetes Database**

**Description:**

The given problem statement which is Diabetes Database analysis is fitted using logistic regression and Multilayer Perceptron and both models are compared.

**Dataset:**

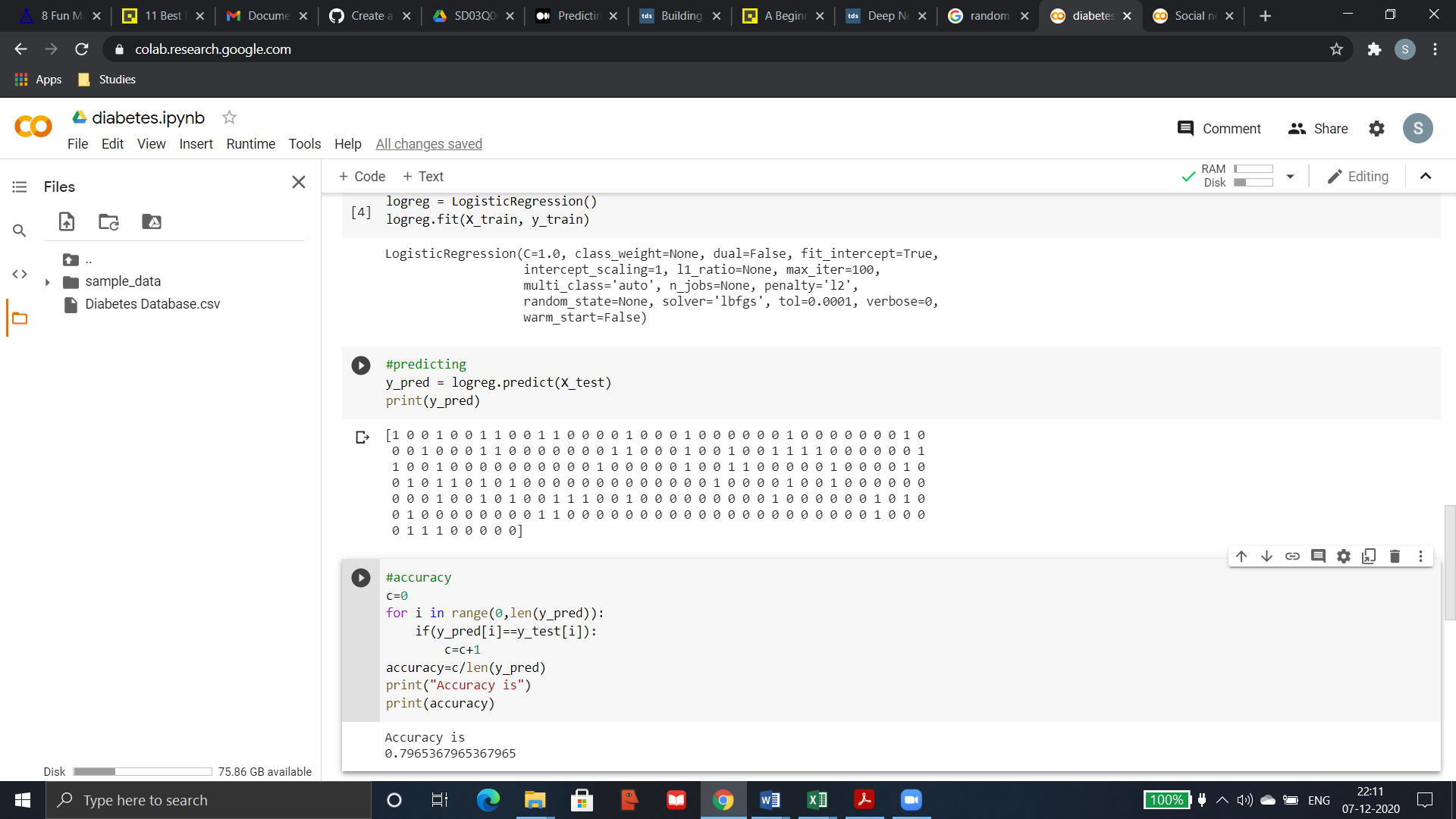
The Diabetes Database.csv data consist of Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome as features. Out of which we have chosen first 7 features as independent variable X and outcome as dependent variable Y. As the dataset was explored it was found that the outcome which is the dependant variable was highly imbalanced with around 80% of the outcome was non-diabetic and 20% was diabetic. To reduce the imbalance between the outcomes we oversampled the diabetic outcome using SMOTE algorithm which resulted in balanced data .We use the StandardScaler module from the sklearn.preprocessing library to scale all of our independent variables.

**Model:**

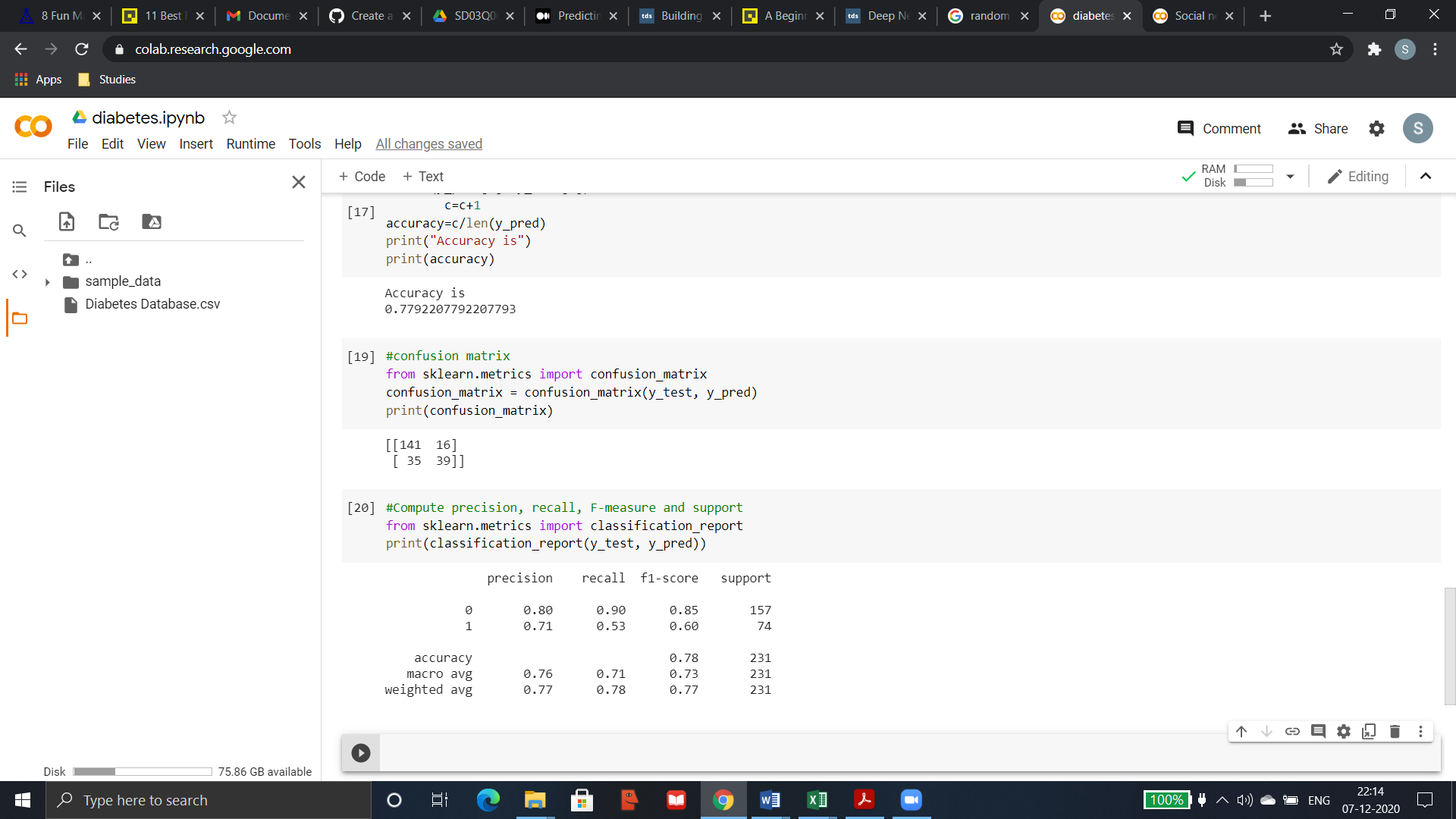
After the data is standardized, the logistic regression model and MLPclassifier is applied to the data for predicting. Then to evaluate and compare the results of both model, a confusion matrix, accuracy, precision of both is shown.

**Results (logistic regression):**

Predicting:

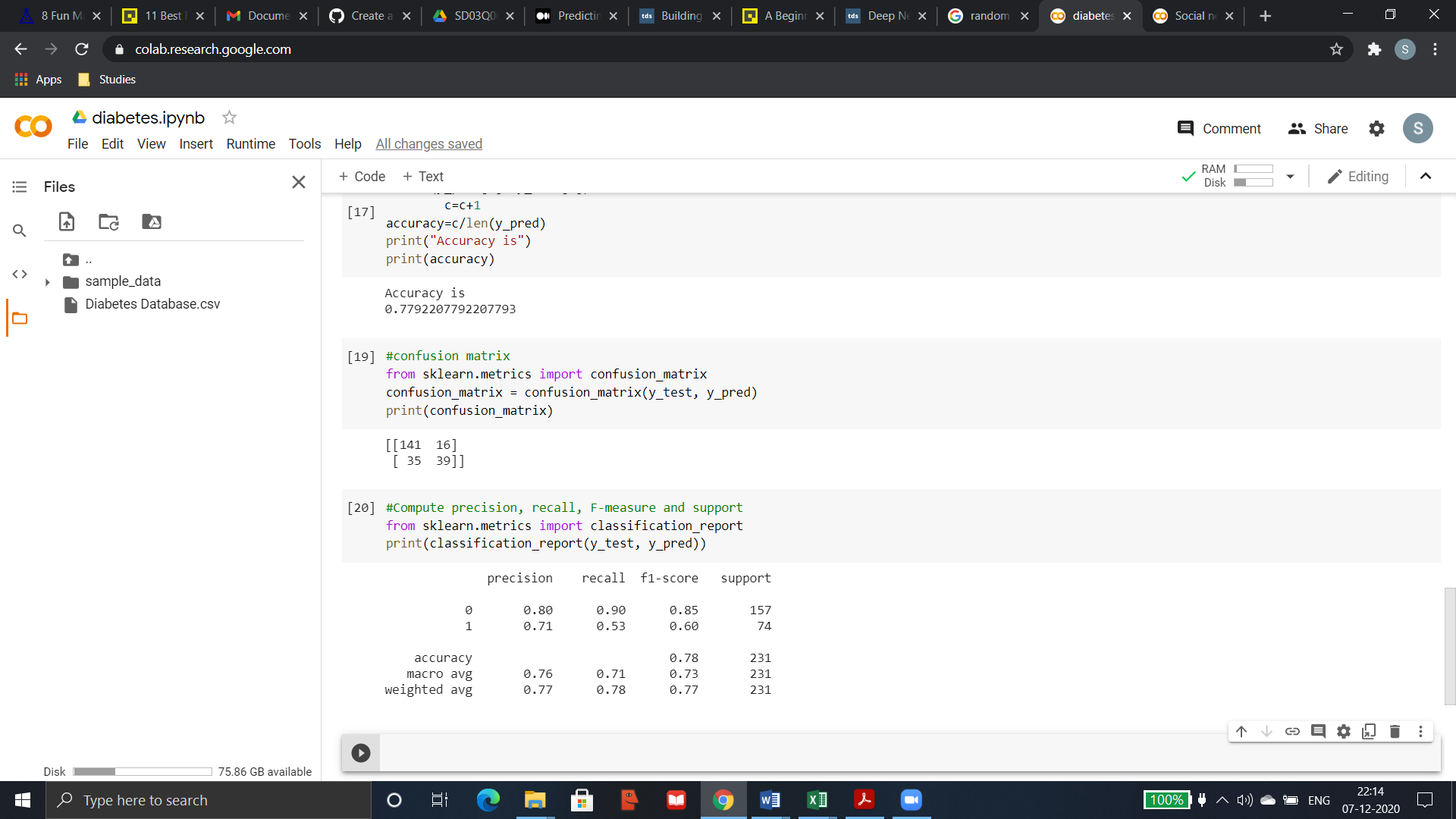


Confusion matrix:



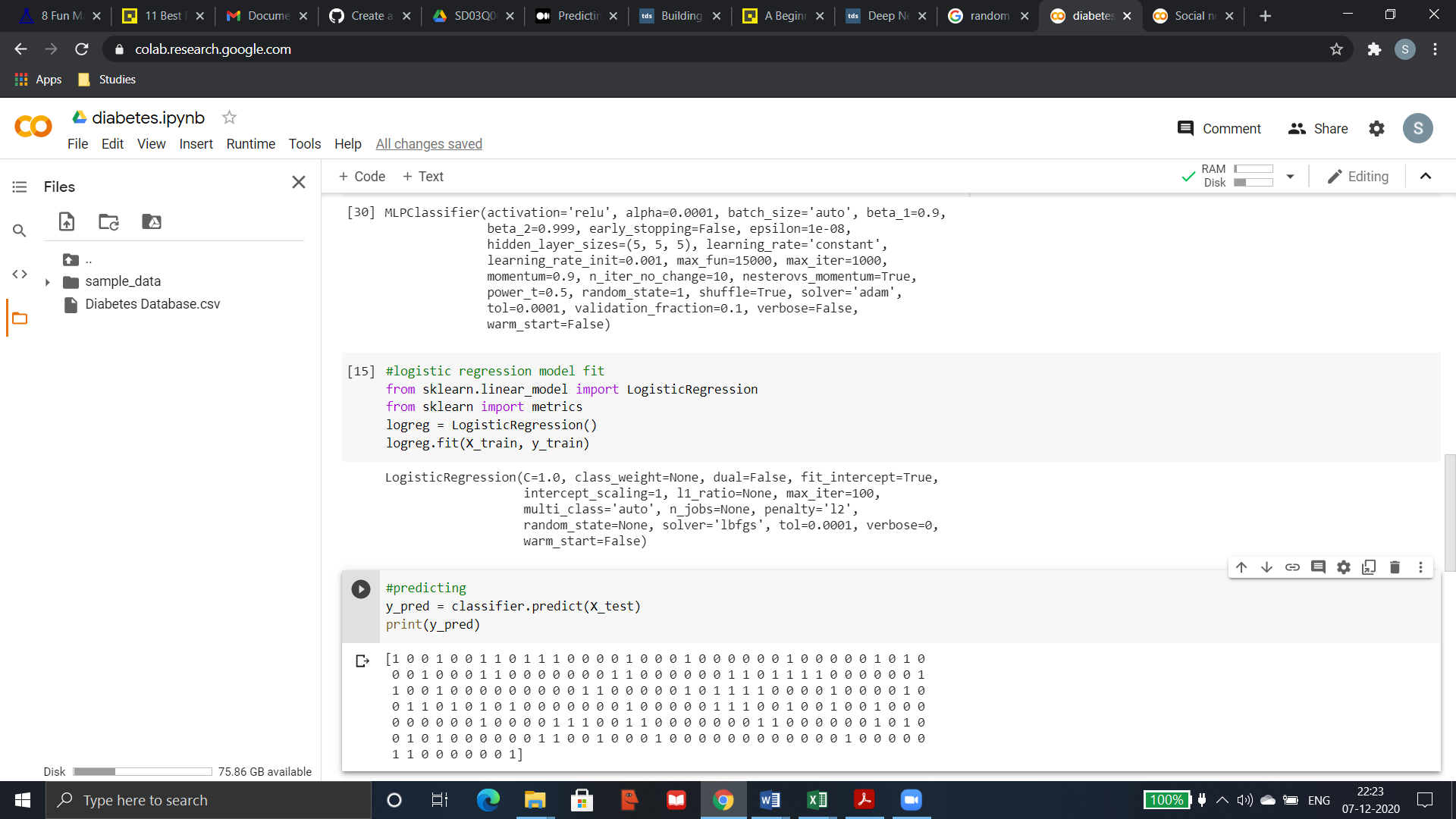
The result from confusion matrix shows that the logistic regression model has predicted 141+39 outcomes correctly. And 35+16 outcomes are wrongly predicted.

Accuracy, Precision, Recall:

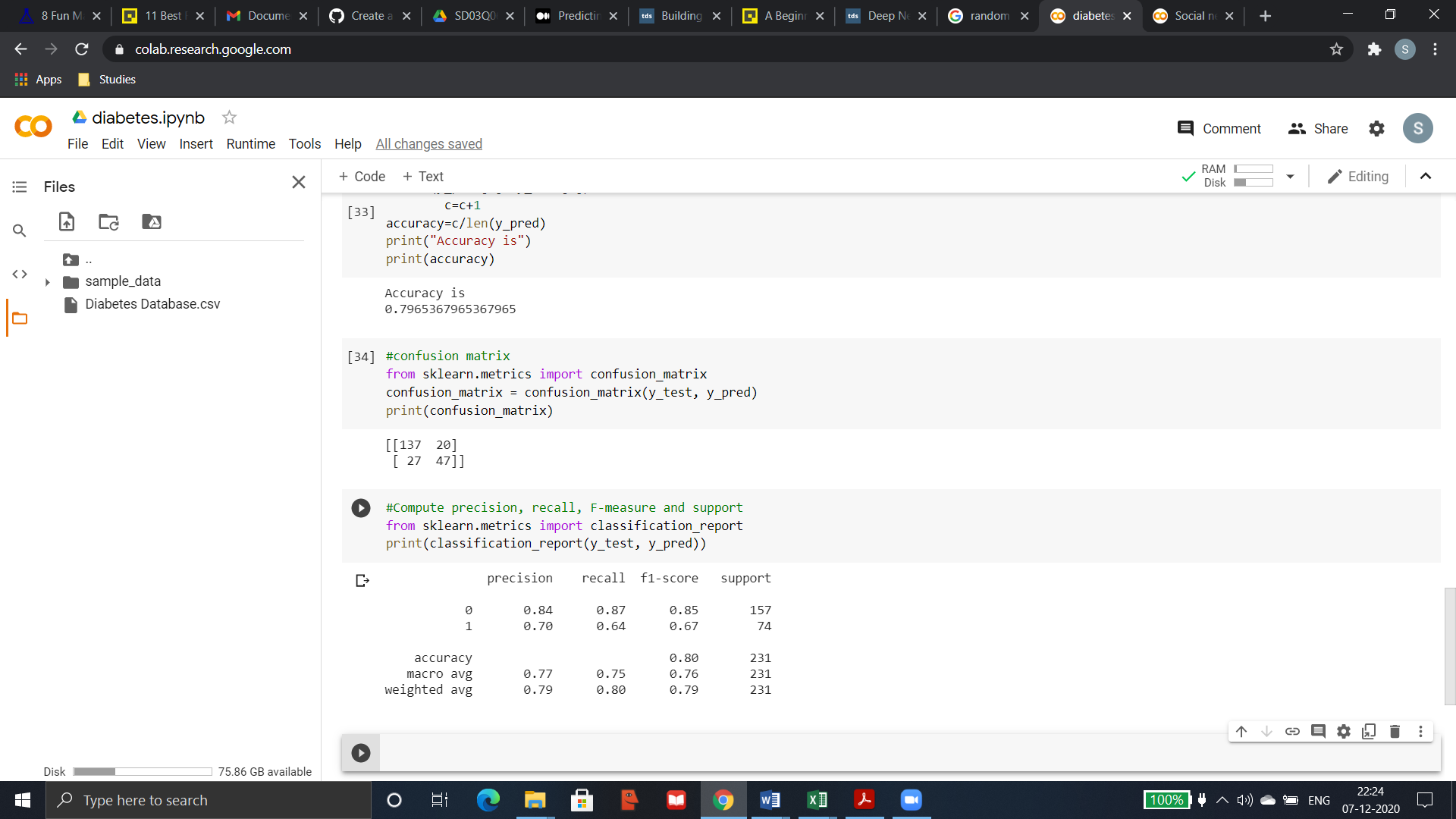


**Results(MLPclassifier):**

Prediction:

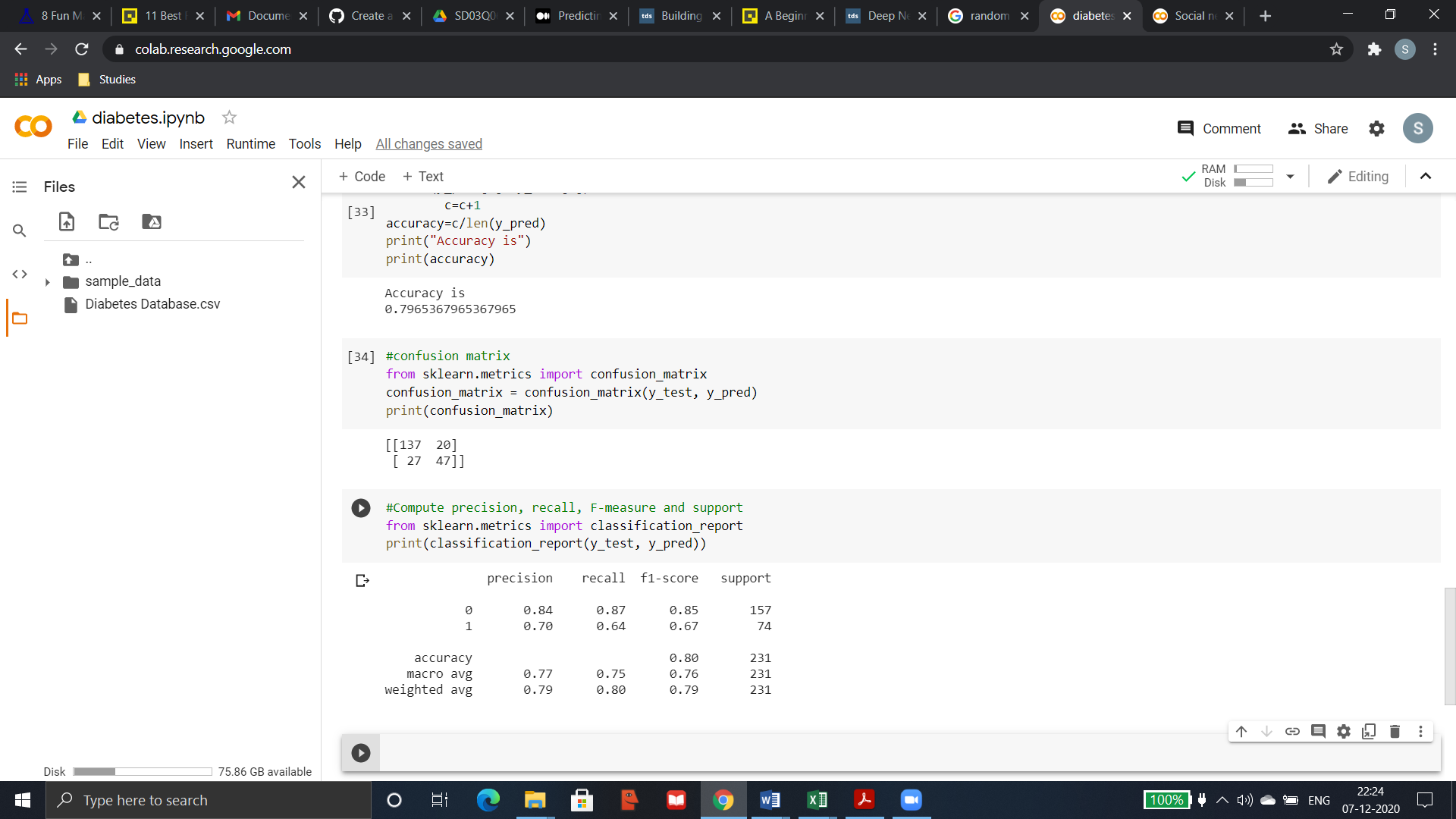


Confusion matrix:



The result from confusion matrix shows that the MLP classifier model predicts 137+47 outcomes correctly whereas it wrongly predicts 27+20 outcomes.

Accuracy, Precision, Recall:



**Conclusion:**

From the above results shown it is clearly evident that with accuracy 80% and high precision and recall values for both outcomes we conclude that compared to logistic regression MLPclassifier is the better fit for the given diabetes database. The accuracy can still be improved for MLPclassifier by increasing the iteration.