Minor Project

Verzeo ML May Batch-1

Diabetes Prediction

By,

Prince P. Pipariya

[princeppipariya24@gmail.com](mailto:princeppipariya24@gmail.com)

**Problem Statement :**

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.

**Packages:**

* pandas
* numpy
* sklearn
* matplotlib

**IDE:**

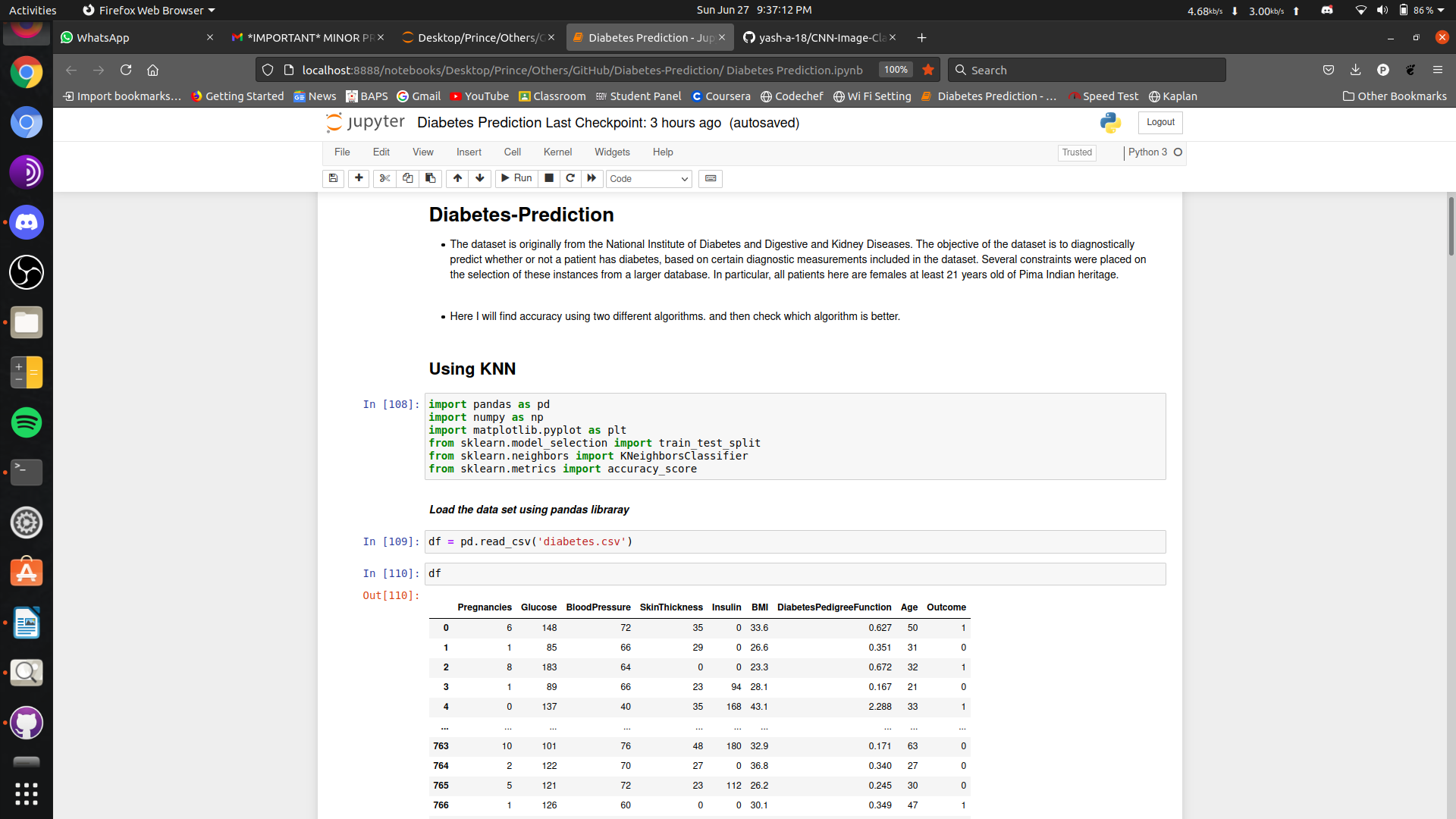
* Jupyter Notebook
* Google Colab (Just for Testing Purpose)

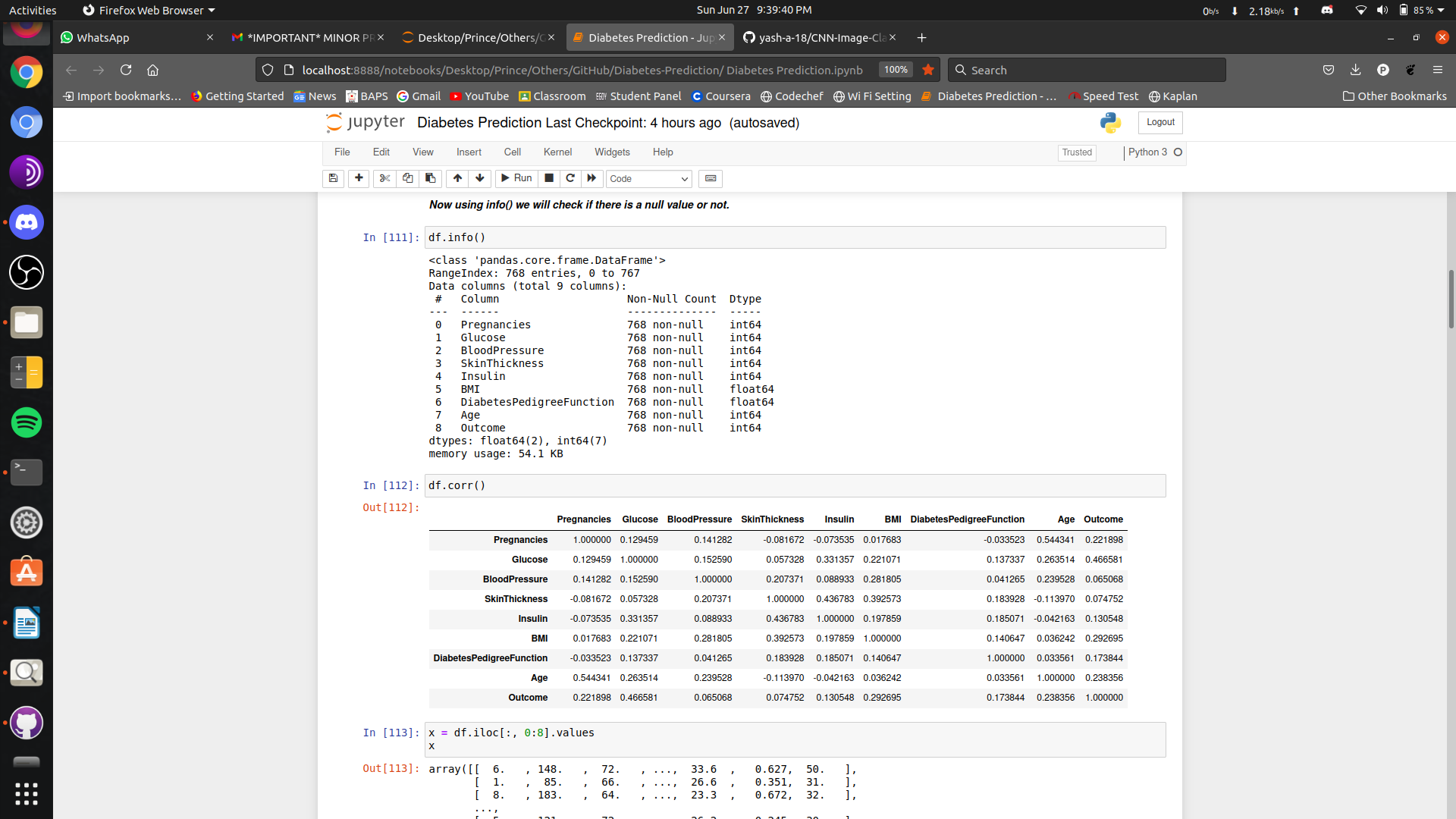
**Details:**

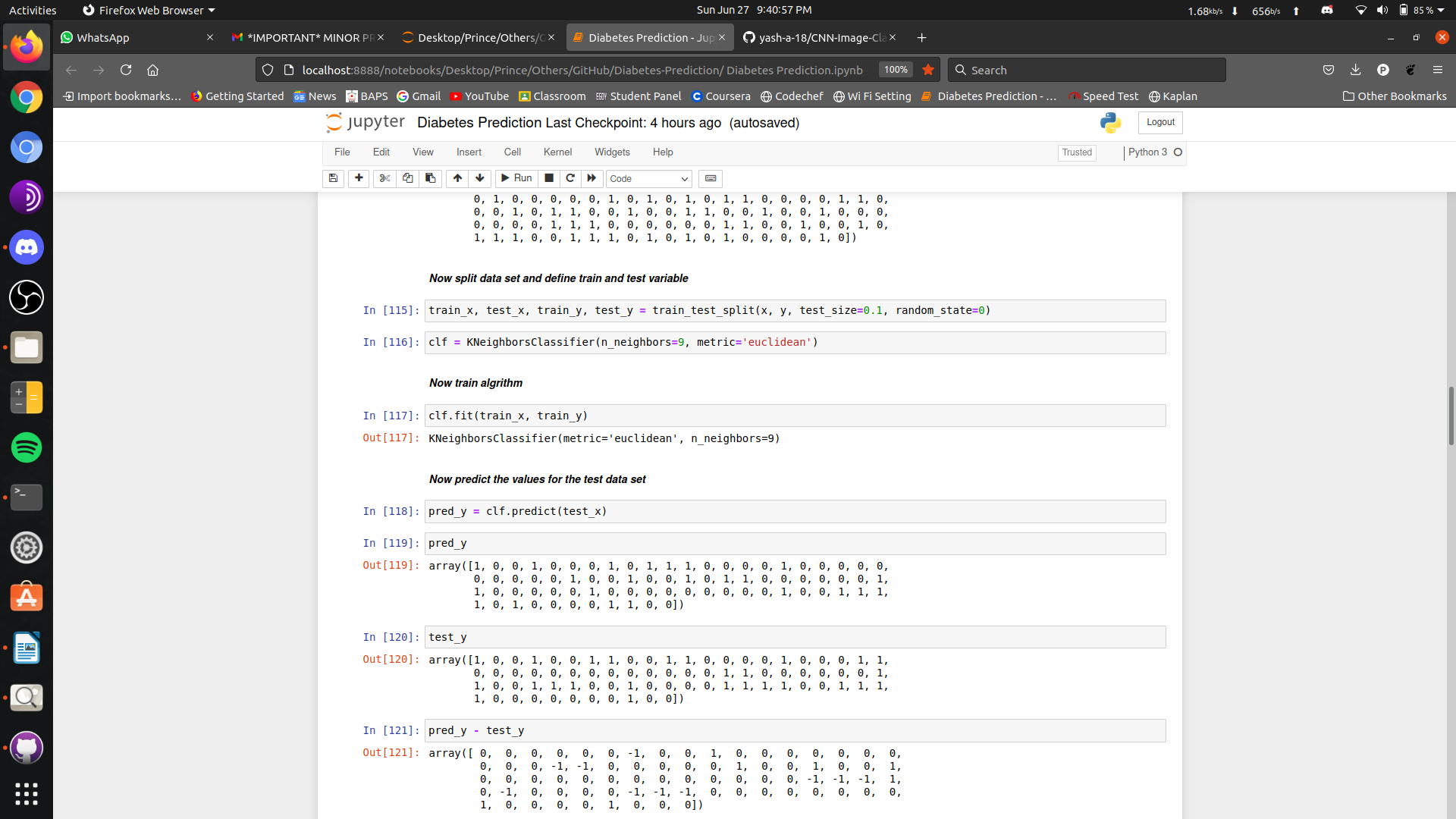
In this project, we required to find that any lady whose age is more than 21 has diabetes or not so for this prediction we must need of data set and it also provided. Here for this problem statement we have to apply classification techanique and for that i have tried two different classification like knn and logistic regression and find out which technique is better.

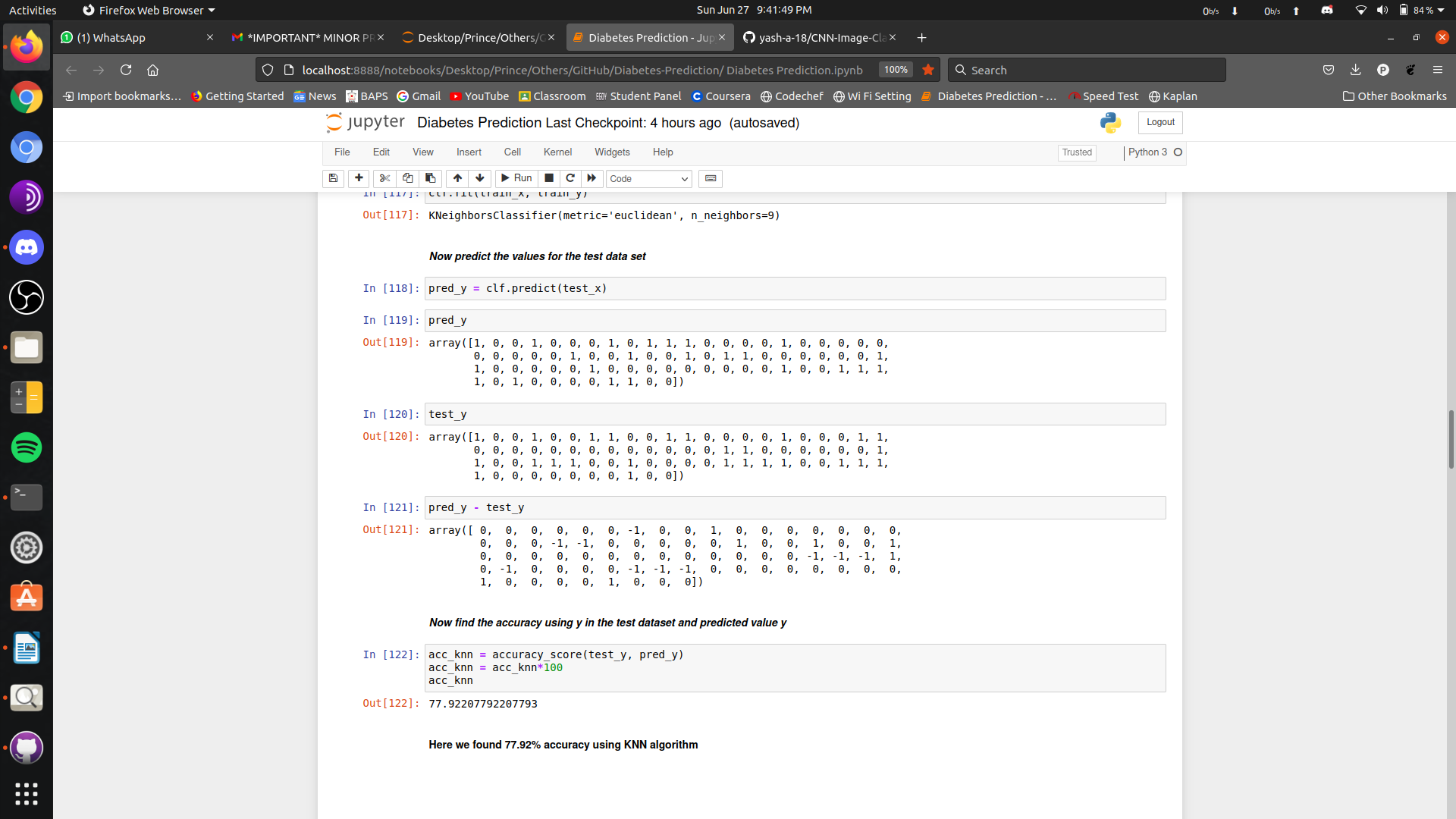
**Working & Screen shots:**

* First i applied knn algorithm.

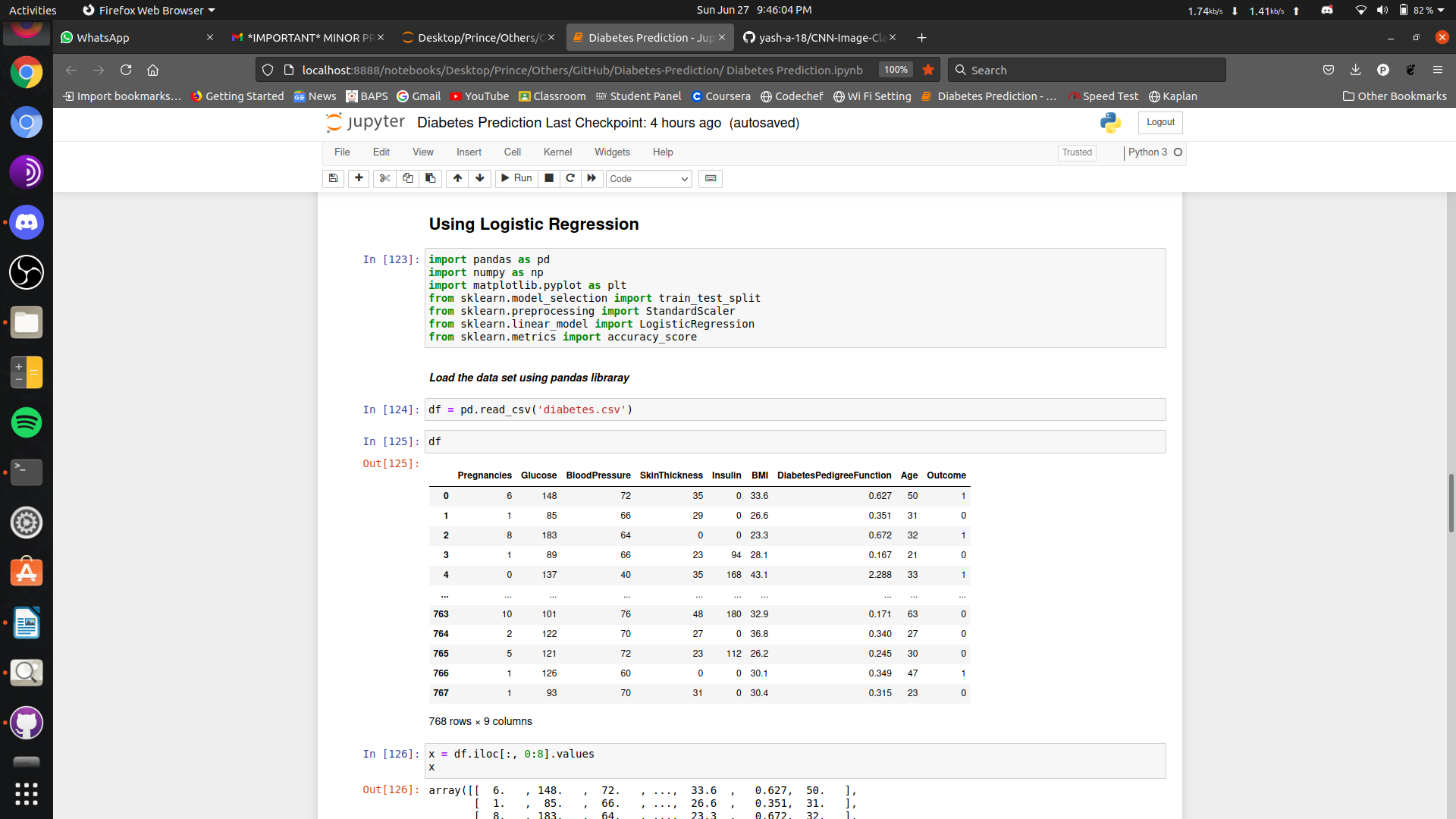


* Now i go for that there is null value or not.
* Now split data and apply knn algrithm for k=9 and find accuracy.

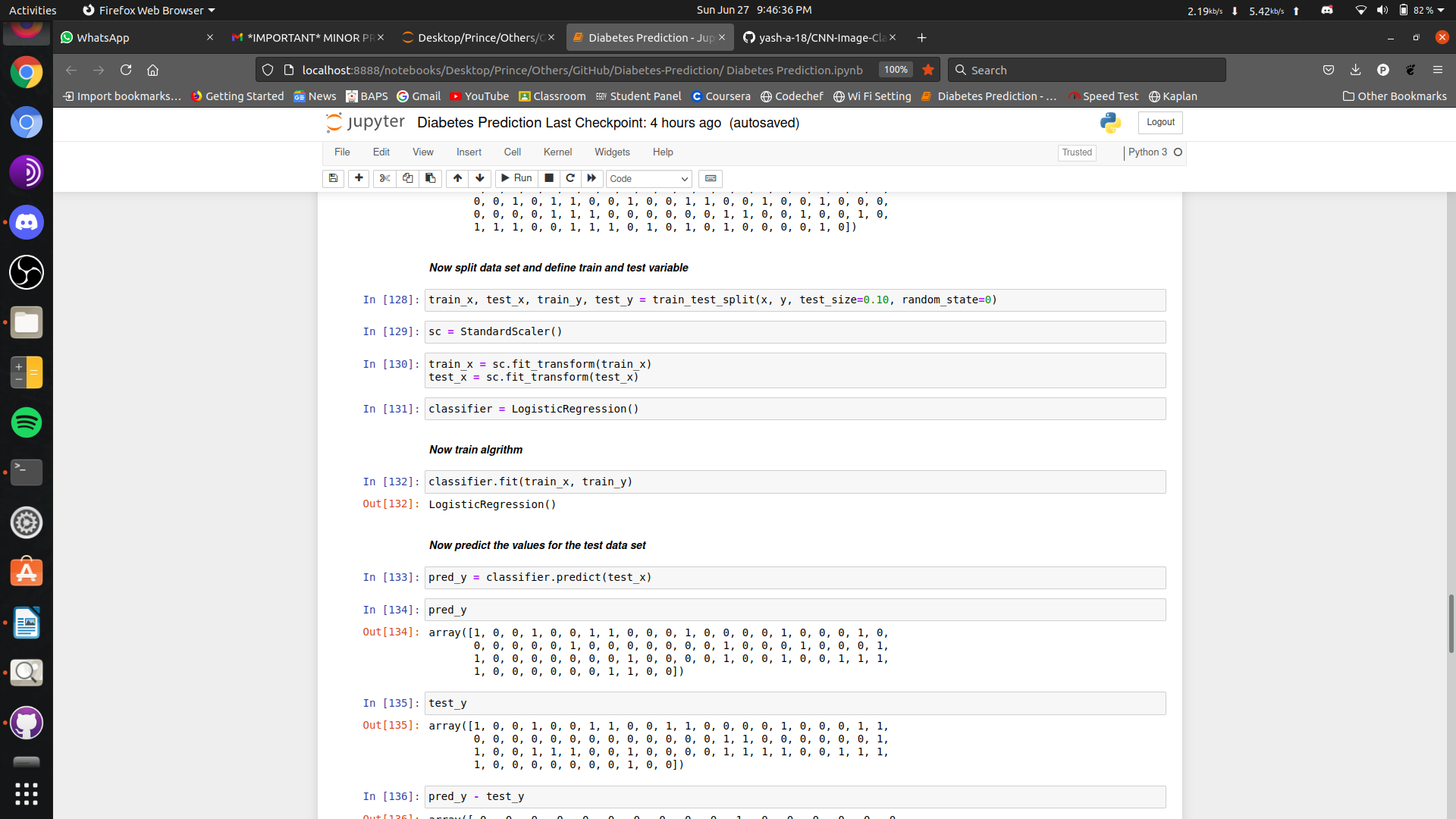


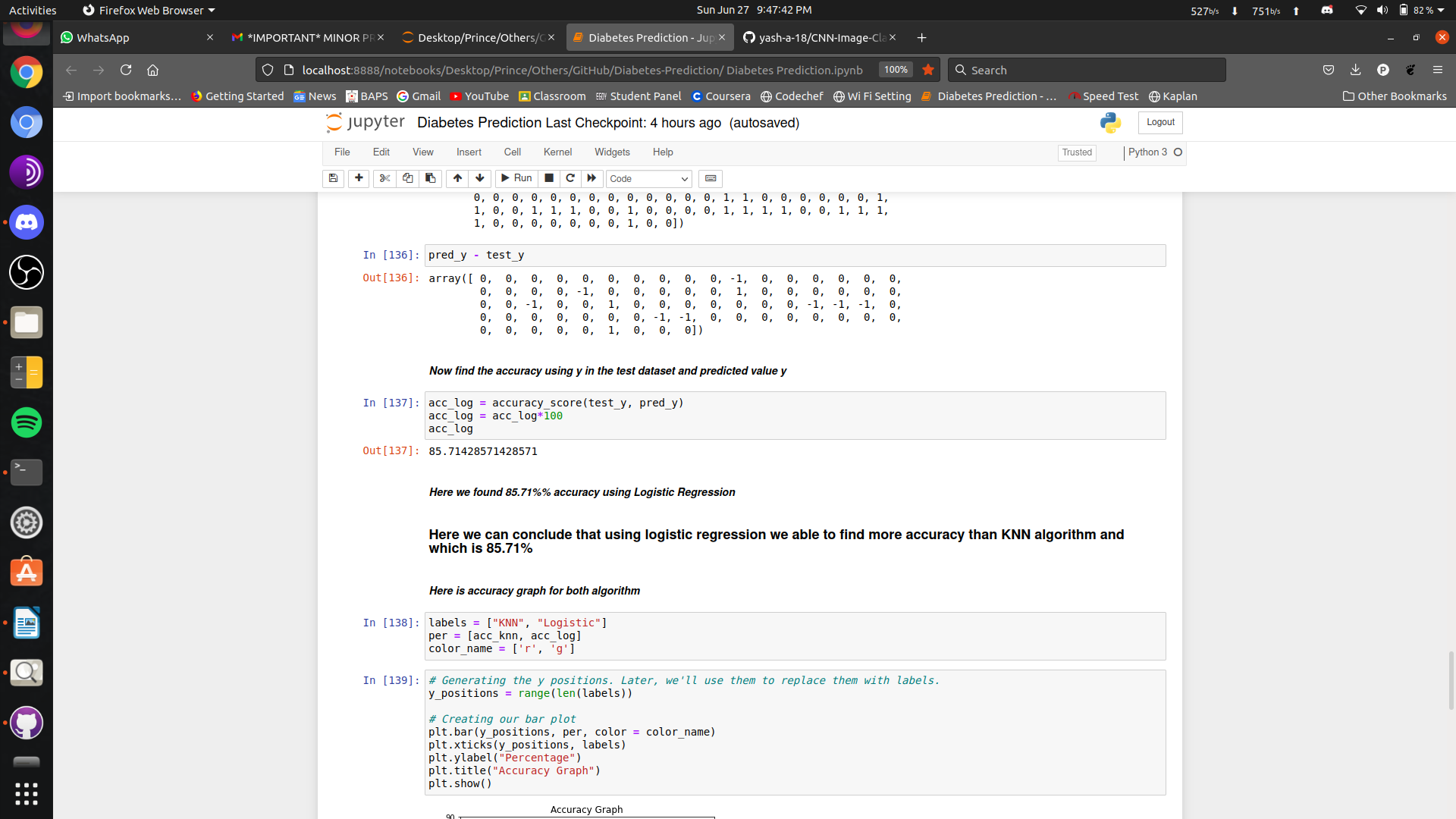


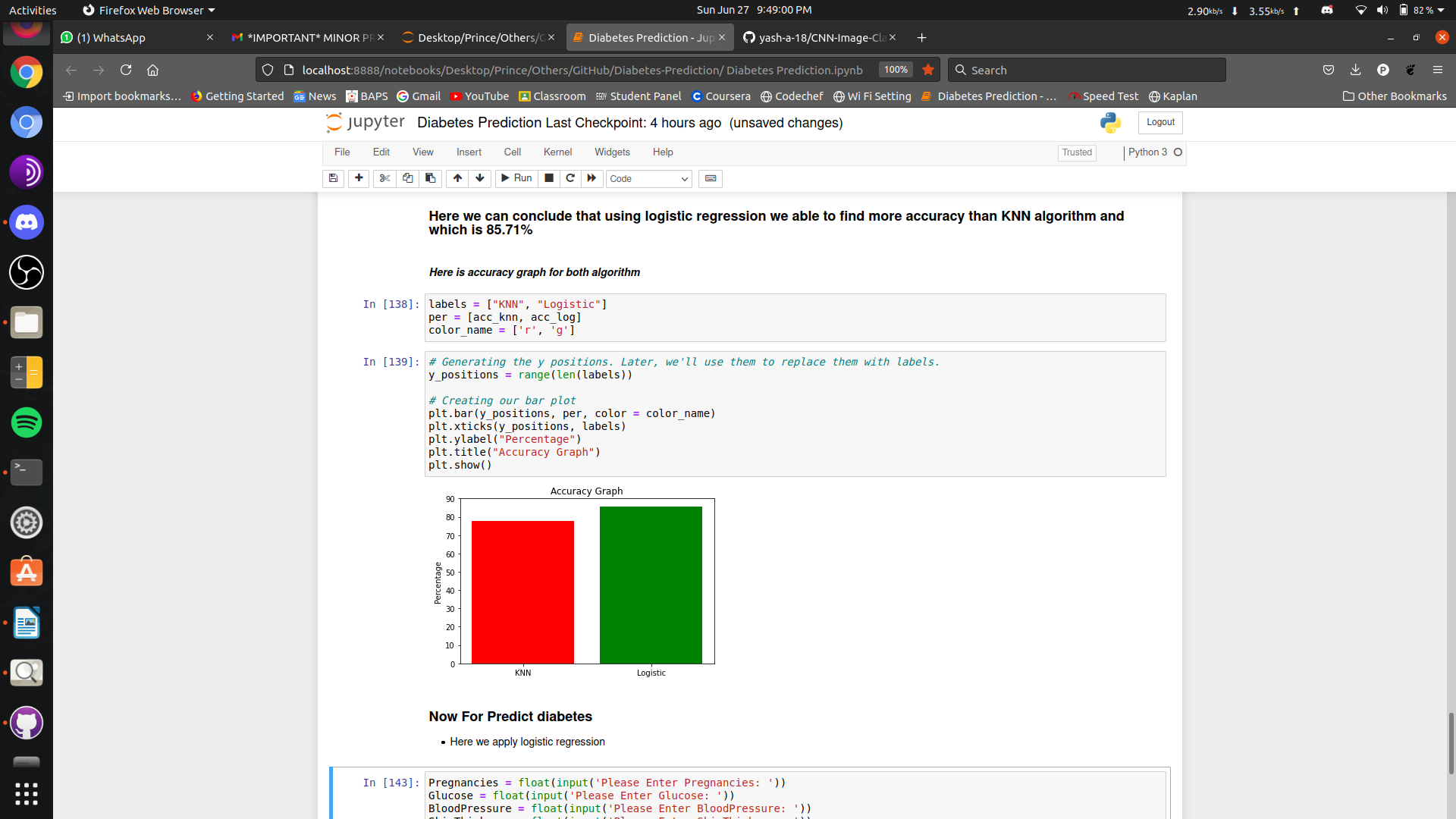
* Second method is Logistic regression.



* Now split data and apply logistic algorithm and find accuracy.







**Conclusion:**

Here we can conclude that using logistic regression we able to find more accuracy than KNN algorithm and which is 85.71%