**RSIP Career Plus ML 044**

**PROJECT TITLE**

**Chronic kidney disease analysis**

**Category: Machine Learning**

**Skills Required:**  
Python ,Python Web Frame Works, Python For Data Analysis, Python For Data Visualization,DataPreprocessingTechniques,MachineLearning,Classification Algorithms

**TEAM MEMBERS:**

                                                                               RAGUL .R

                                   KAVIN KUMAR.P

                                      YUVAANESH.VB

                                       LOGESH.VS

**Project Description:**

The definition and classification for [chronic kidney disease](https://www.sciencedirect.com/topics/medicine-and-dentistry/chronic-kidney-disease) was proposed by the National Kidney Foundation Kidney Disease Outcomes Quality Initiative (NKF-KDOQI) in 2002 and endorsed by the Kidney Disease: Improving Global Outcomes (KDIGO) in 2004. This framework promoted increased attention to chronic kidney disease in clinical practice, research and public health, but has also generated debate. It was the position of KDIGO and KDOQI that the definition and classification should reflect patient prognosis and that an analysis of outcomes would answer key questions underlying the debate.

Machine learning tasks are classified into several broad categories. In supervised learning, the algorithms build a mathematical model from a set of data that contains both the inputs and desired outputs. Classification algorithms and regression algorithms are types of supervised learning. Classification algorithms are used when the outputs, are restricted to a limited set of values. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. In unsupervised learning, the algorithm builds a mathematical model from a set of data which contains only inputs and no desired output labels. Unsupervised learning can discover patterns in the data, and can group the inputs into categories.

**INTRODUTION:**

Chronic kidney disease includes conditions that damage your kidneys and decrease their ability to keep you healthy by doing the jobs listed. If kidney disease gets worse, wastes can build to high levels in your blood and make you feel sick. You may develop complications like high blood pressure, anemia (low blood count), weak bones, poor nutritional health and nerve damage. Also, kidney disease increases your risk of having heart and blood vessel disease. These problems may happen slowly over a long period of time. Chronic kidney disease may be caused by diabetes, high blood pressure and other disorders. Early detection and treatment can often keep chronic kidney disease from getting worse. When kidney disease progresses, it may eventually lead to kidney failure, which requires dialysis or a kidney transplant to maintain life.

**Proposed Solution:**

The main objective of this research is to use classification algorithms to identify the liver patients from healthy individuals. This project also aims to compare the classification algorithms based on their performance factors.

To serve the medicinal community for the diagnosis of chronic kidney disease among patients..

The following algorithms are trained for this purpose, to choose the best performing classifier.

* Logistic Regression
* K-Nearest Neighbors

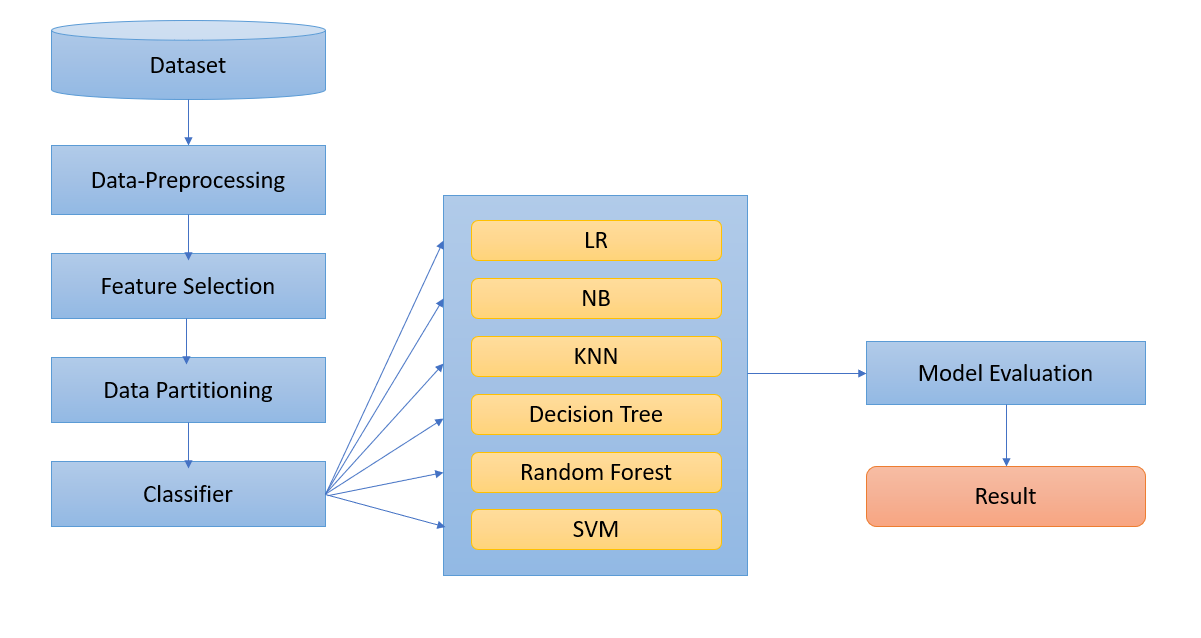
The results of each of the classification algorithm is summarized in the table shown below.

|  |  |
| --- | --- |
| **Model** | **Accuracy Score** |
| Logistic Regression | 0.925 |
| KNN | 0.8325 |

As clearly summarized in the table, Logistic Regression gave the best result.

**THEORITICAL ANALYSIS:**

**BLOCK DIAGRAM**



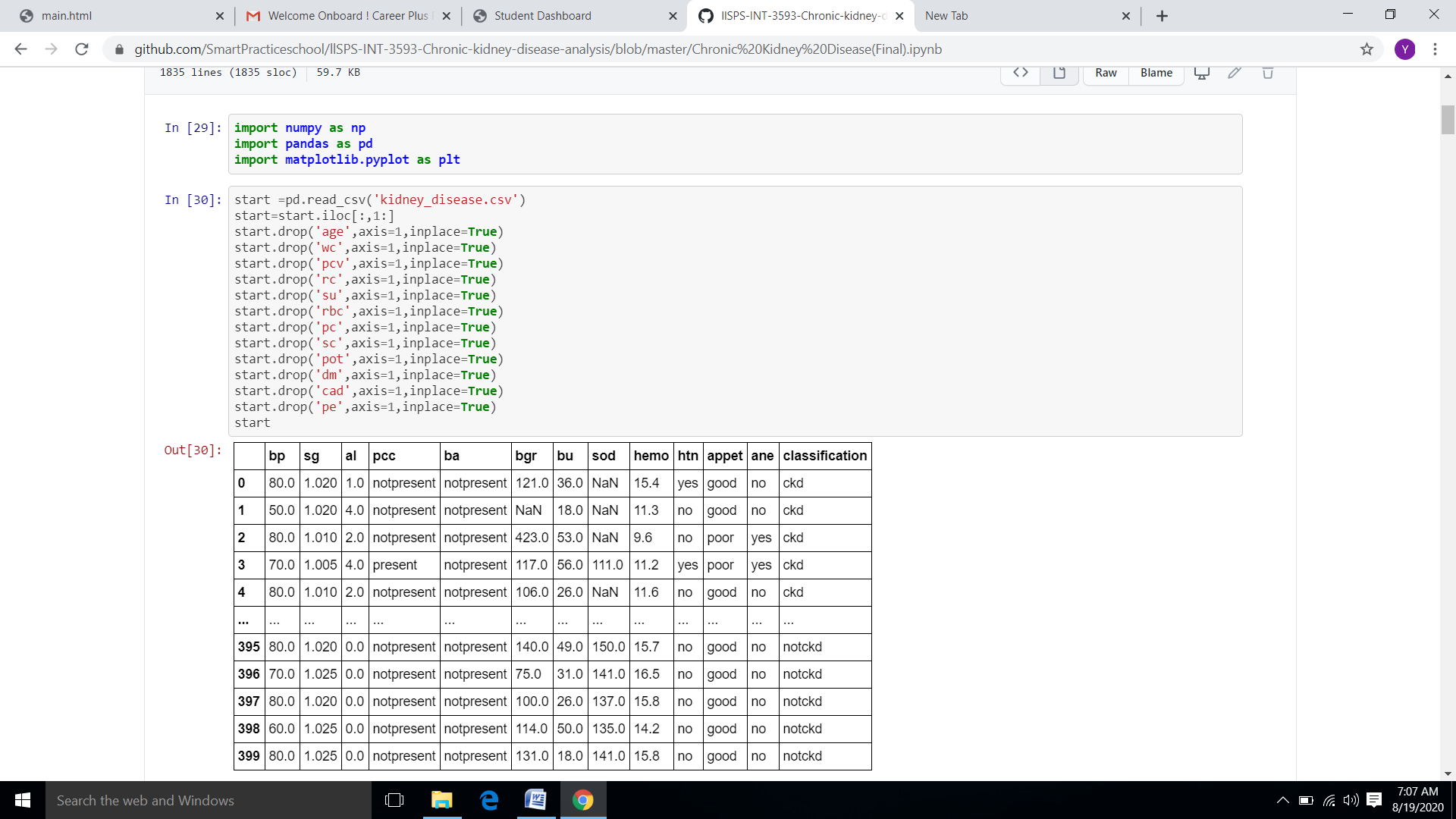
**HARDWARE/SOFTWARE DESIGNING:**

The steps followed in developing the model:

* Data Collection: The dataset was downloaded from the UCI ML Repository.
* Data Analysis: Evaluating cleanliness of the dataset by looking for any irrelevant data and handling missing data.
* Search for any trends, relations and correlations.
* Developing a model where the patient can be identified to be having  chronic kidney disease or not.

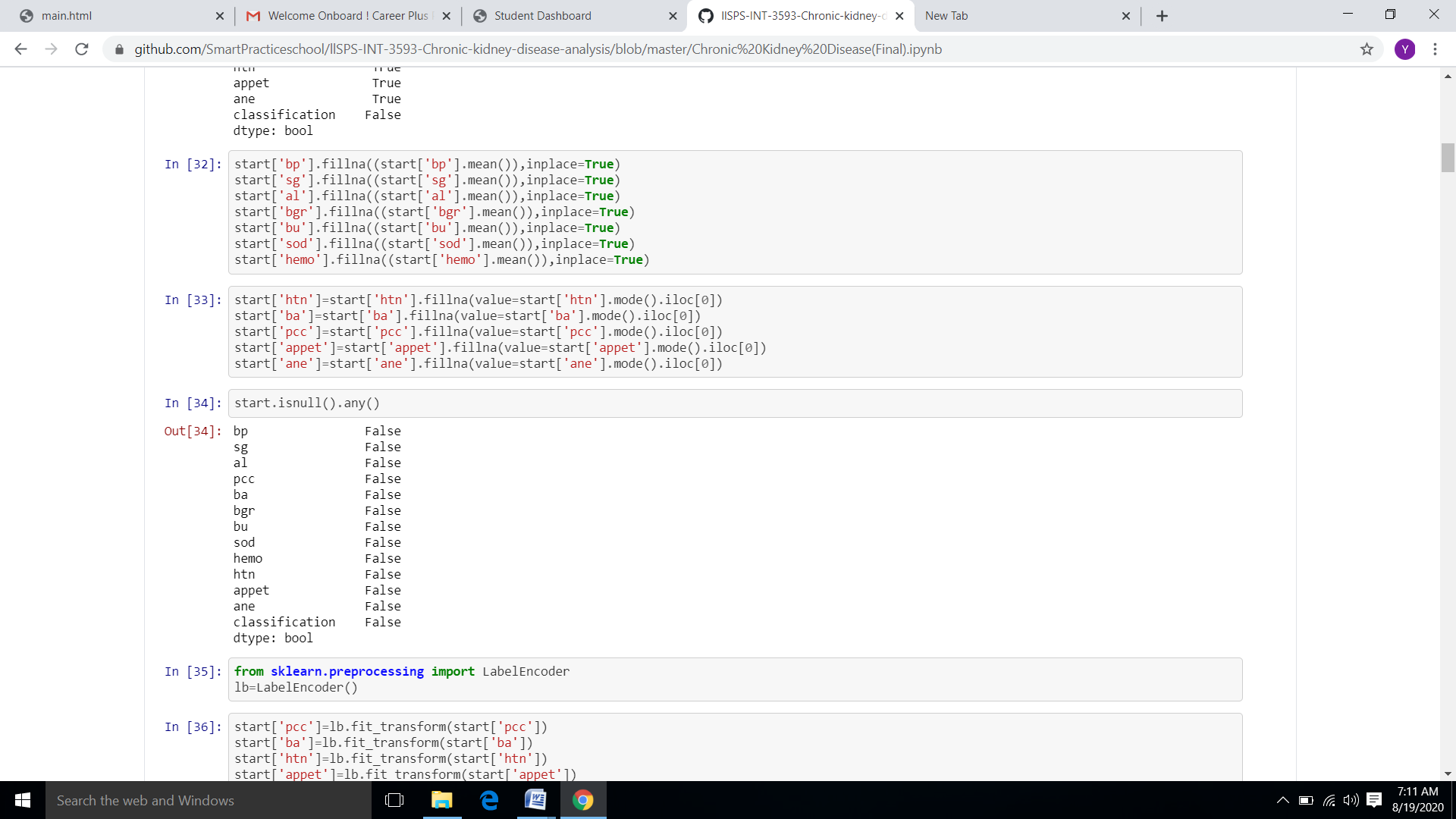
**Importing the Libraries:**

            Pandas,Numpy,Matplotlib libraries are imported .Numpy is the numerical python library used for doing the mathematical calculation.Matplotlib is visualization library.Pandas is used for data manipulation.



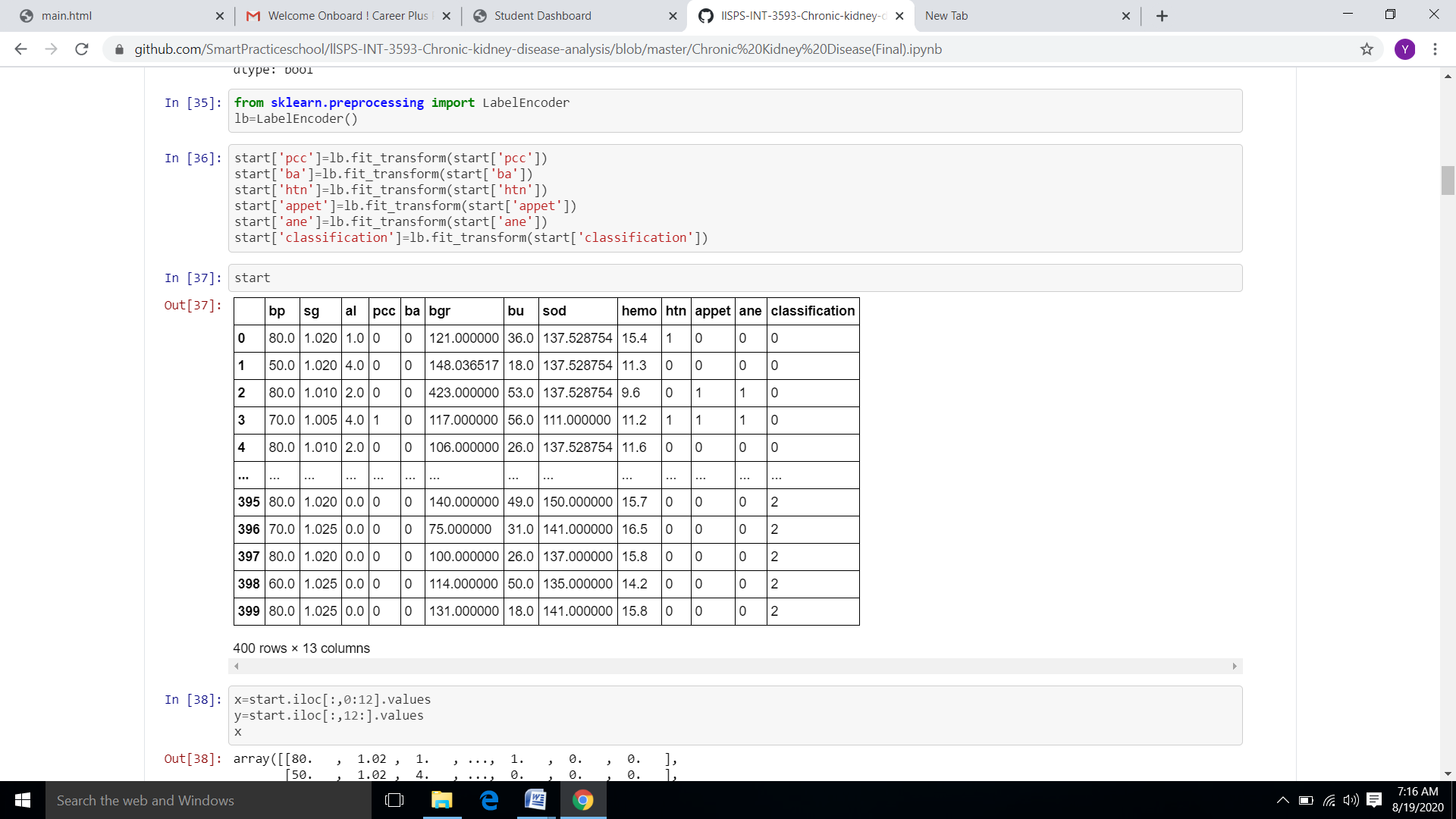
**Handling Missing By Mean and Mode**:

Since there are lot of unfilled columns and rows in the datset we nee to handle the missing values using the mode ,if it is categorical column else by using mean.



**Column Transformation:**

Since here most of the categorical columns has only two categories.Label encoder is enough for transforming the categorical columns into binary format.Later the values are used for carrying on the operations.

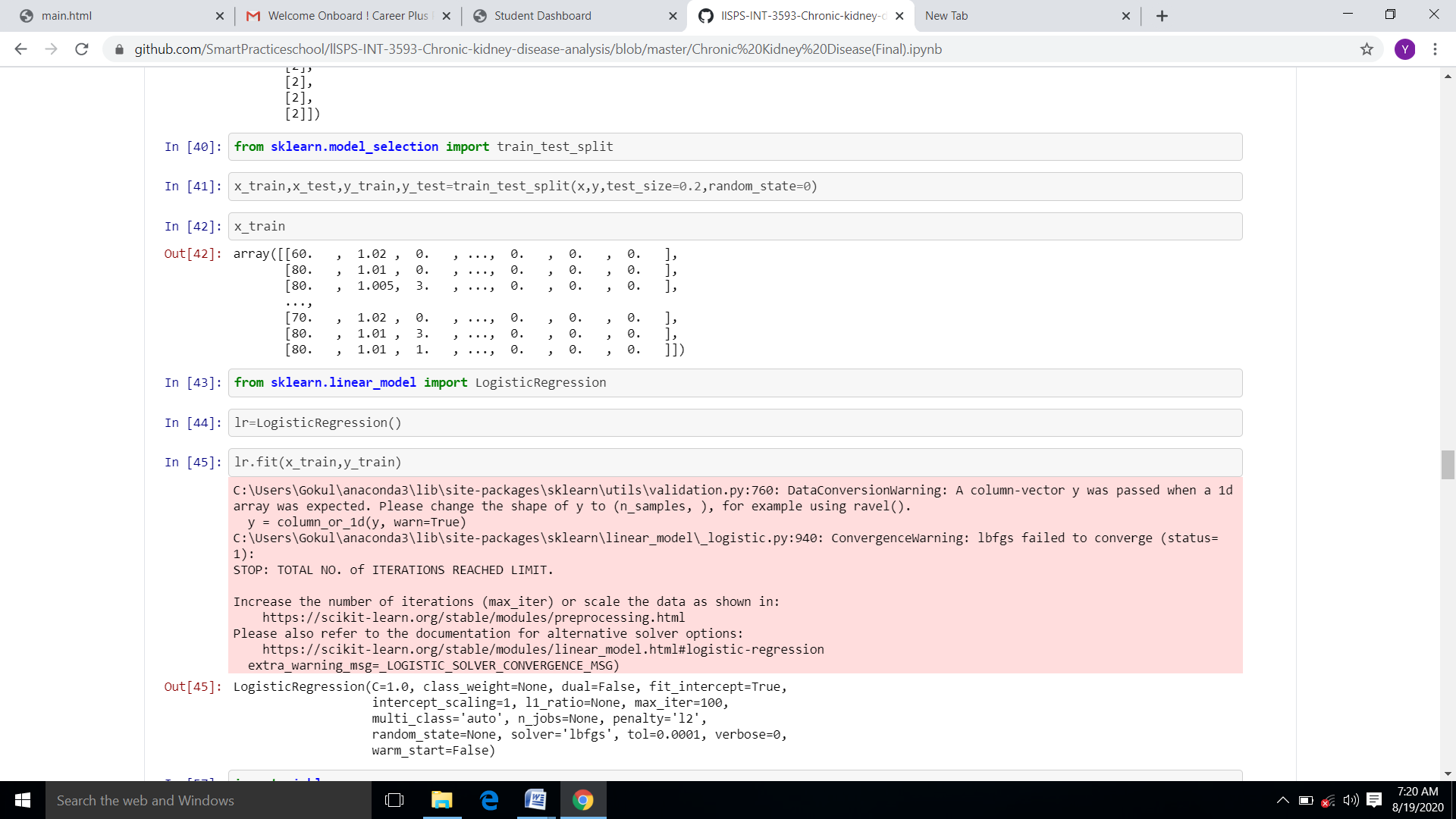


**Building a Model:**

            Since the outcome is binary and we have a reasonable number of examples at our disposal compared to number of features, this approach seems suitable. Since for this data, it already knows the output beforehand, it continuously adjusts the weights such that when these weights summed up with their features are introduced in the logistic function, the results are as near as possible to the actual ones.

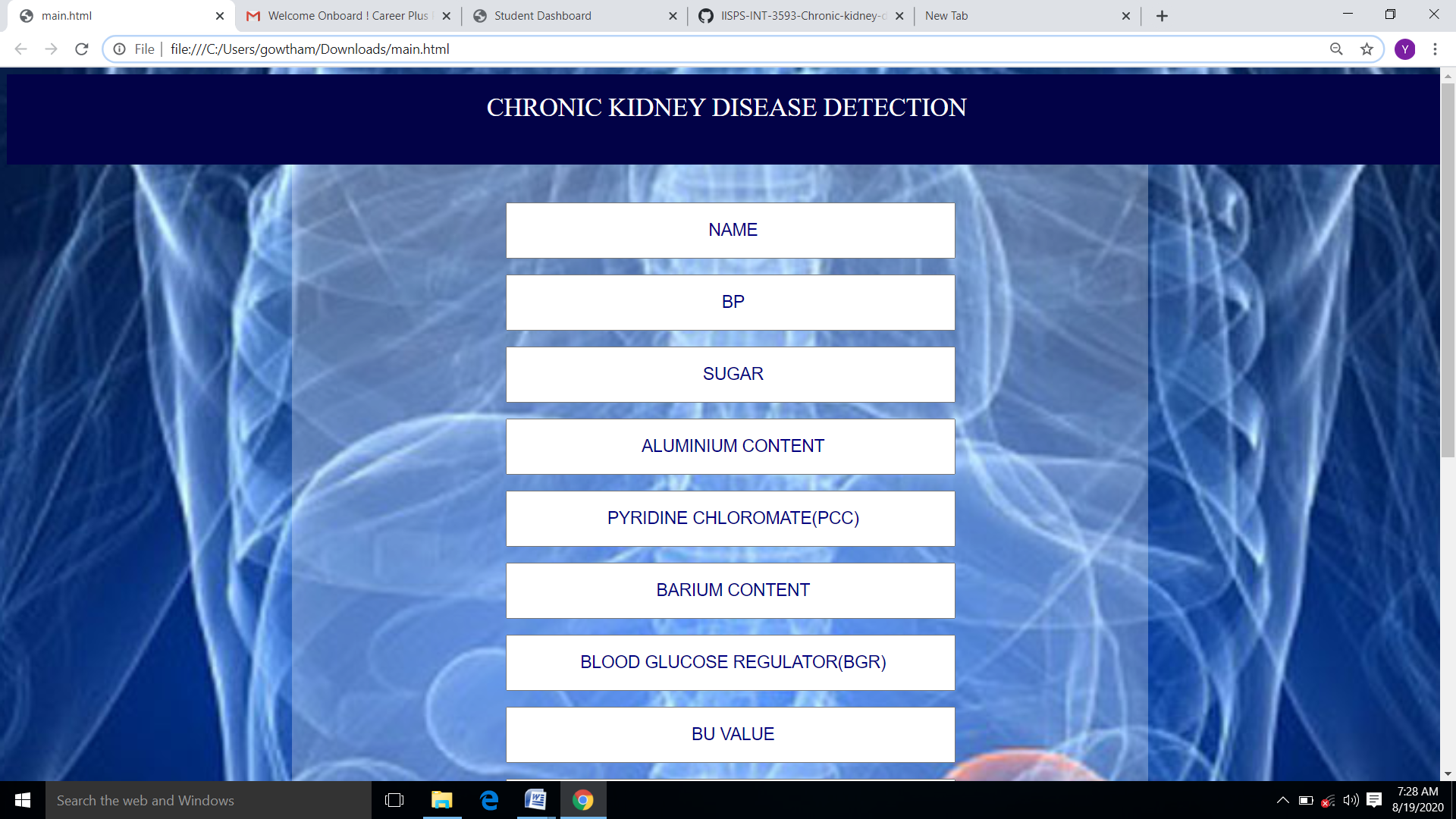
Once presented with a test value, it again inserts the value into our logistic function and returns the output as a number between 0 and 2, which represents the probability of that test value being in a particular class.

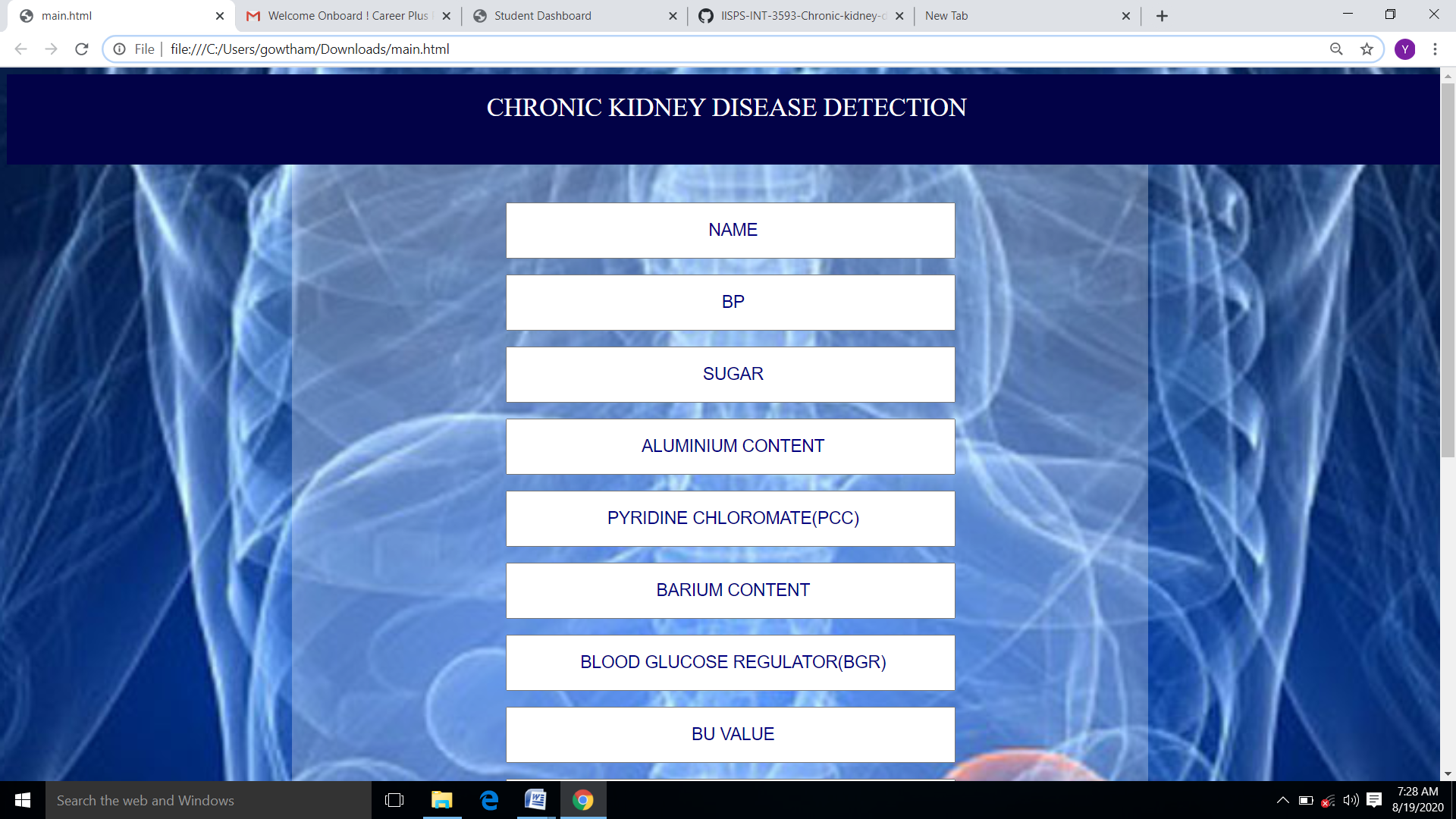
**Train and Testing Data:**



**Webpage:**

             The html code is created using visual studio code.The css used here is of type internal css.





**ADVANTAGES AND DISADVANTAGES:**

The benefits of this model are:

* Easy interface
* Straight forward results
* Accurate performance calculations

Disadvantages:

As our dataset is small, it’s training dataset is similar to test dataset. So it is difficult for the model to predict accurately for larger dataset.

**APPLICATIONS:**

This project makes it easier to predict whether the patient has chronic kidney disease or not.

**Future Scope:**

Database should be expanded on which the system will be tested much better. Also, the model requires further improvement mostly regarding feature selection of the chronic kidney disease into multiple components.

**Conclusion**:

        In this project, we have proposed methods for diagnosing chronic kidney disease in patients using machine learning techniques. The two machine learning techniques that were used include  Logistic Regression and  KNN. The system was implemented using all the models and their performance was evaluated. Performance evaluation was based on certain performance metrics. Logistic Regression was the model that resulted in the highest accuracy with an accuracy of 92%. A GUI, which can be used as a medical tool by hospitals and medical staff was implemented using Logistic Regression.

**Reference:**

1.Coresh J, Selvin E, Stevens L, et al. Prevalence of chronic kidney disease in the United States. *JAMA*. 2007;298:2038-2047.