**CHRONIC KIDNEY DISEASE ANALYSIS**

**A PROJECT REPORT ON REMOTE SUMMER INTERNSHIP PROGRAM-2020**

**ACKNOWLEDGEMENT**

I take this opportunity to express my deep gratitude and sincerest thanks to my trainer, for giving most valuable suggestions, helpful guidance and encouragement in the execution of this project work.

I would like to give a special mention to my mentor. Last but not the least I am grateful to all the members of SmartBridge (SmartInternz) platform for their support.

**CONTENT**

1. Introduction

1.1 Overview

1.2 Purpose

1. Literature Survey

2.1 Existing Problem

2.2 Proposed Solution

**3.** Theoretical Analysis

3.1 Block Diagram

3.2 Hardware/Software Designing

**4.** Experimental Investigations

**5.**Result

**6.** Advantages and Disadvantages

**7.**Applications

**8.** Conclusions

**9.** Future Scope

**10.** Bibliography

**INTRODUCTION**

**Overview**

The project Chronic Kidney Disease Analysis helps all the patients to know the valuable information regarding tests conducted during diagnosis. Being computerized the system will be more helpful and provide patients at ease.

**Purpose**

Chronic Kidney Disease (CKD) is a major medical problem and can be cured if treated it in the early stages. Usually, people are not aware that medical tests that are taken for different purposes could contain valuable information concerning kidney diseases. Consequently, attributes of various medical tests are investigated to distinguish which attributes may contain helpful information about the disease.

**LITERATURE SURVEY**

**Existing Problem**

The term “chronic kidney disease” means lasting damage to the kidneys that can get worse over time. If the damage is very bad, your kidneys may stop working.

Anyone can get CKD. Some people are more at risk than others. Some things that increase your risk for CKD include:

* Diabetes
* High blood pressure (hypertension)
* Heart disease
* Having a family member with kidney disease

Usually, people are not aware that medical tests that are taken for different purposes could contain valuable information concerning kidney diseases and if known beforehand can be cured at an early stage.

**Proposed Solution**

The solution that we are proposing to this problem is a Machine Learning Model that analyzes the information as follows:

* It helps us to measure the severity of the problem.
* The predicted survival of the patient after the illness.
* The pattern of the disease.
* Work for curing the disease.

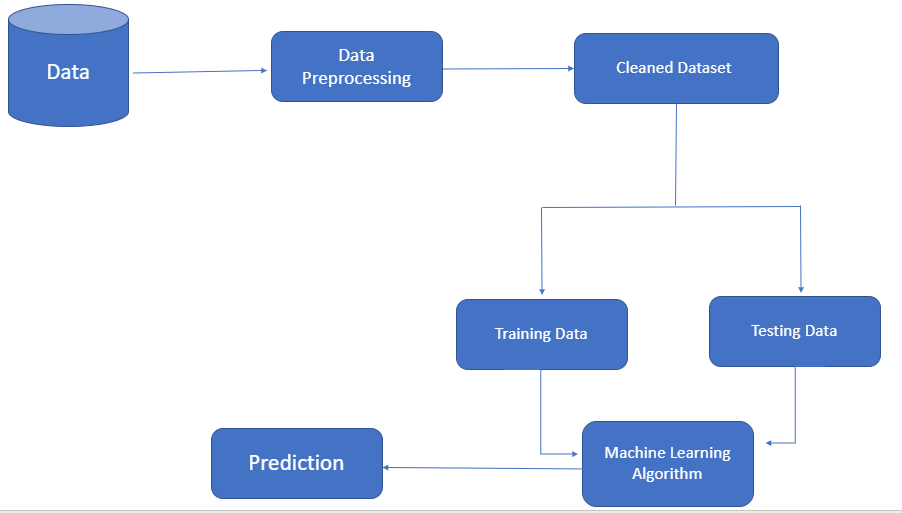
The algorithms used to train the Machine Learning Model are:

1. Logistic Regression
2. K-Nearest Neighbours

Following all these steps and finding the cure at an early stage will benefit many pateints and provide them with a healthy living.

**THEORETICAL ANALYSIS**

**Block Diagram**



**Hardware/Software Designing**

The steps followed in developing the model are as follows:

1. Data Acquisition : We have downloaded the dataset from kaggle.
2. Data Visualization : Visualizing the different columns of the dataset by plotting graphs .
3. Data Preprocessing : Taking care of any missing entries and of any relations within the dataset.
4. Model Building : Developing a model with the help of different classification algorithms to predict whether the patient is suffering from CKD or not.

**EXPERIMENTAL INVESTIGATIONS**

**Importing Libraries:**

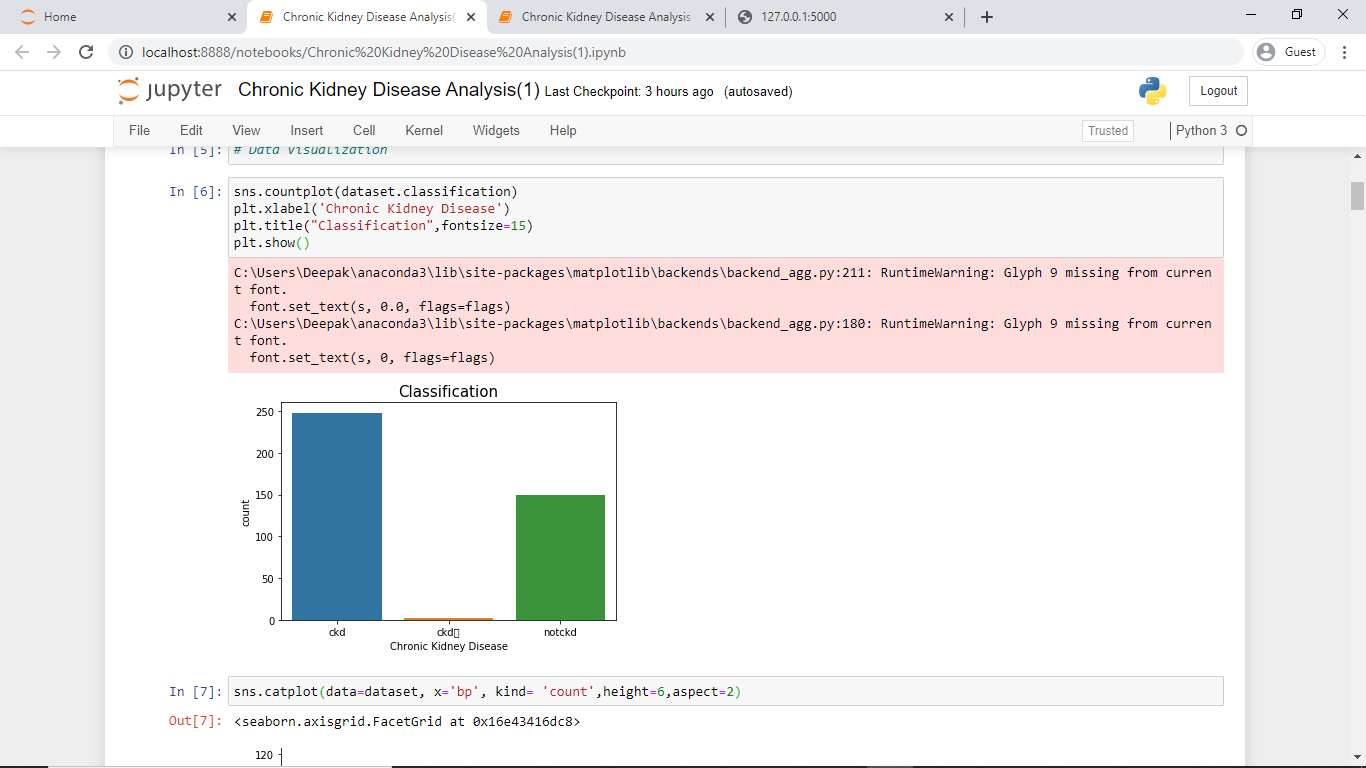
The first step of starting the analysis on the dataset is by importing the libraries numpy, pandas, matplotlib. Numpy is the numerical python library used for all sorts of mathematical calculations.Matplotlib is used for visualization of data. Pandas is used for data manipulatin.

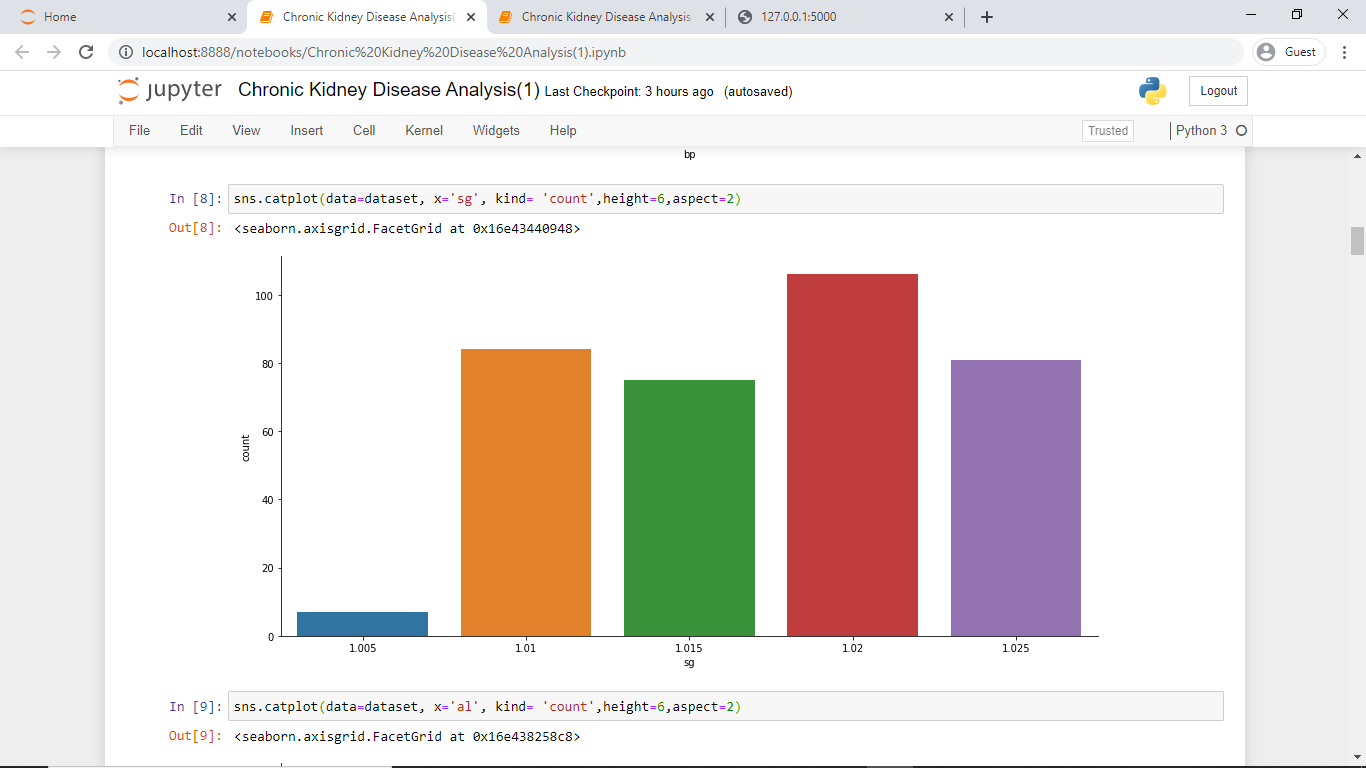
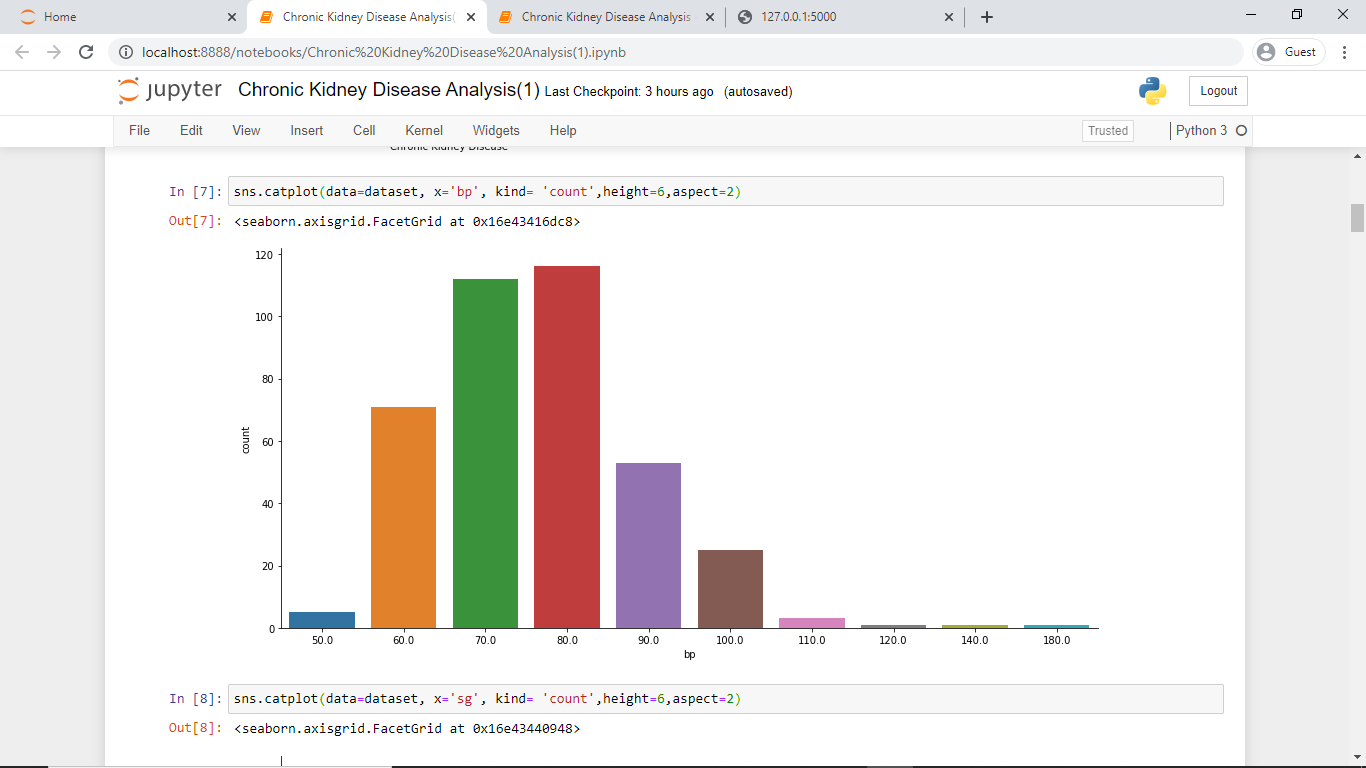
**Importing the Dataset:**

The dataset that was downloaded from kaggle is first uploaded in a .csv format and then imported in the file.

**Data Visualization:**

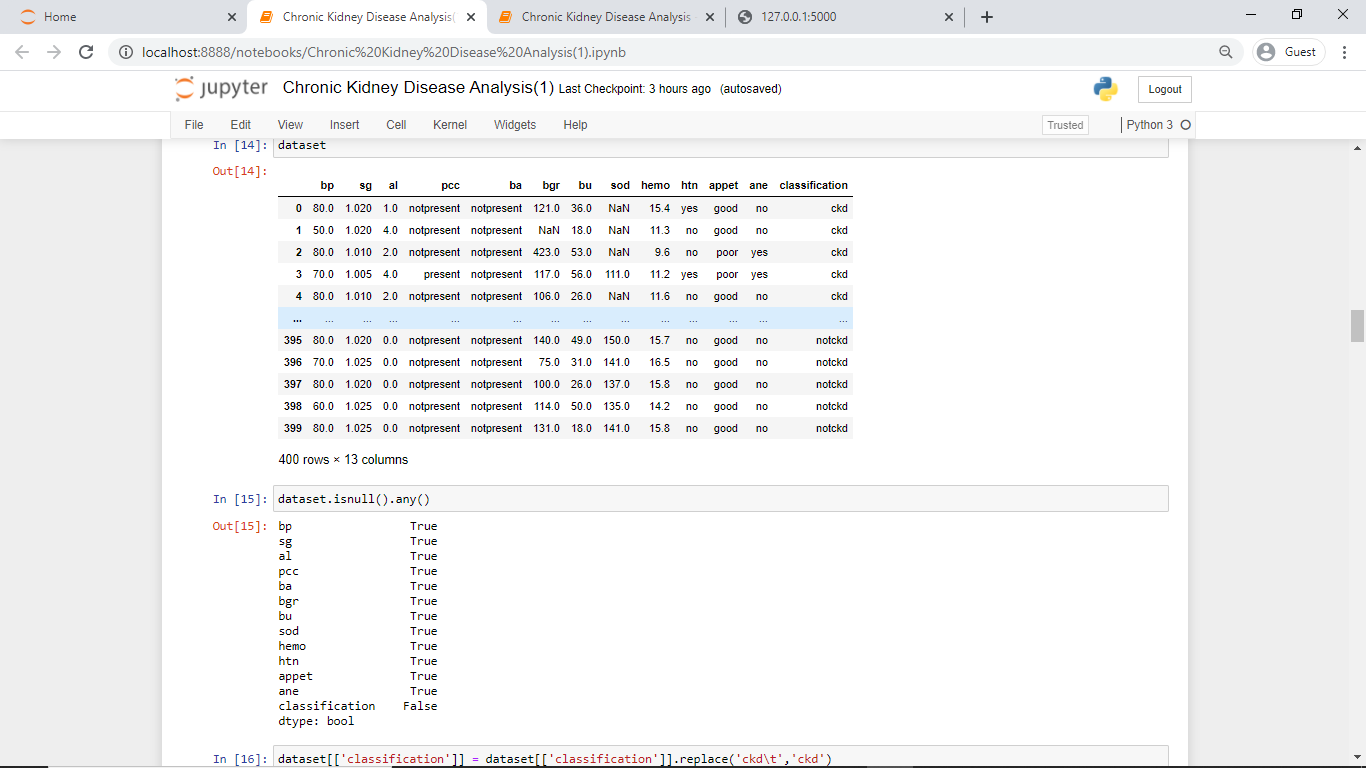
The raw data is visualized by plotting graphs of different input characteristics and the solution is proposed based on this analysis. This is an important step in analysing which category of algorithms(regression or classification) are to be applied on the model.





**Taking care of missing data:**

Dividing the data after visualization into dependent and independent variables and then finding out the missing data in the dataset , and then filling them by using different mode and mean methods as per requirement. For numerical values we use mean and for categorical values , mode is used.



**Applying One Hot Encoder and Label Encoder:**

It mainly did the work of converting categorical columns into binary elements for easy acceptibility by the algorithm.

**Feature Scaling:**

The main purpose of feature scaling is to convert all varying data in one format lying between 0-1.

**Splitting the dataset:**

The dataset is then divided into two sets:

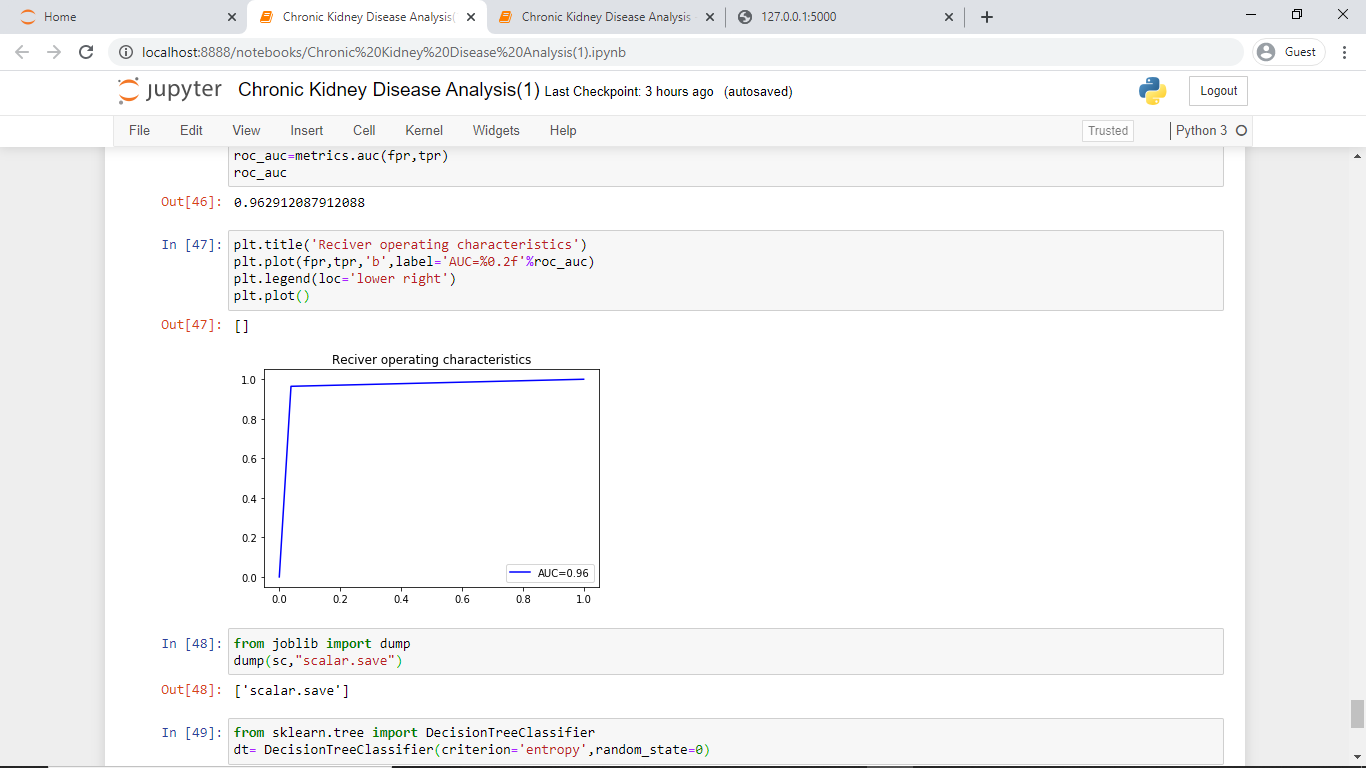
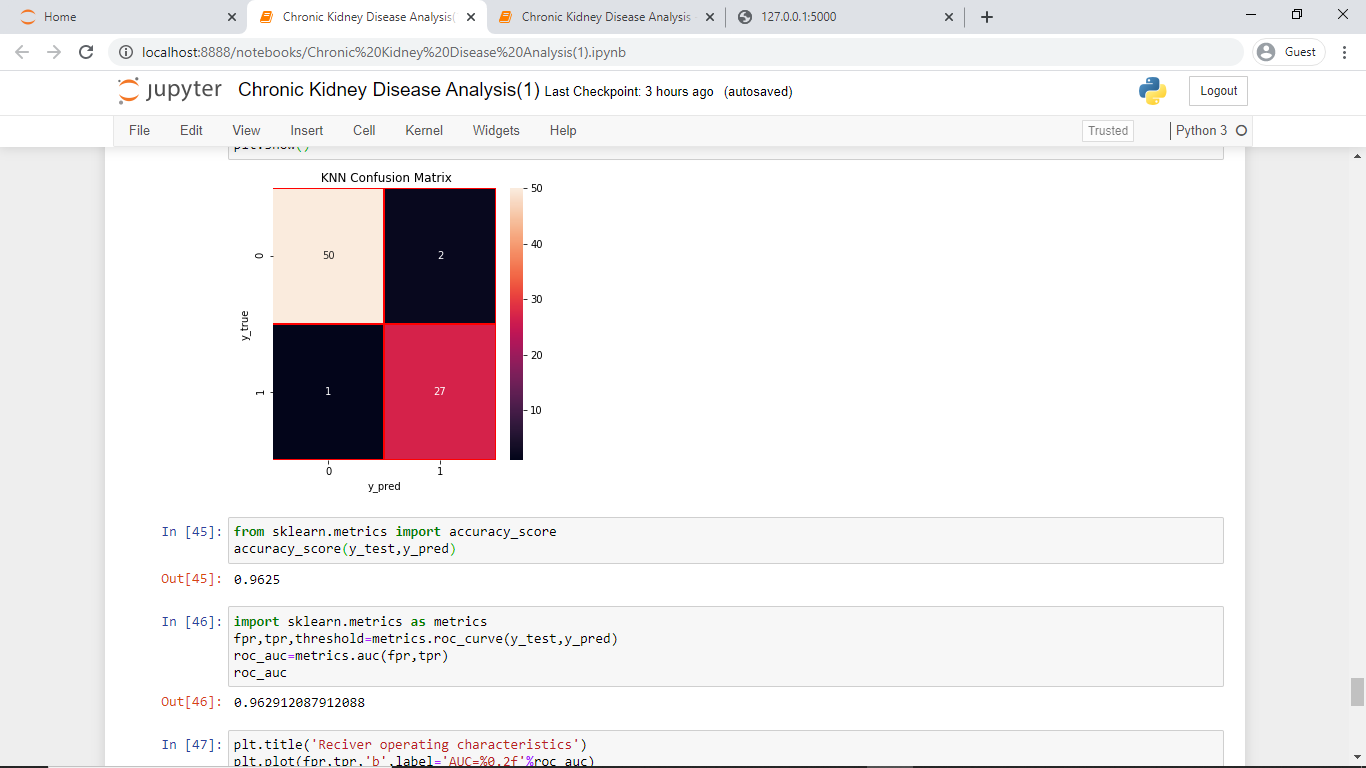
1. Training Data- It mainly comprises of 80% of the data. It is used to train the data to acquire desired output.
2. Testing Data- It comprises of the other 20% of the data that is tested on the model and on the basis of which accuracy of the model is predicted.

**Model Building:**

The main work here is of applying the different algorithms on the training set and checking the accuracy of the model on the basis of these algorithms. The evaluation is done using Confusion Matrix and by plotting the AUC curve.

The two algorithms that we used to build the model are:

1. Logistic Regression
2. K-Nearest Neighbors



**Application Building:**

The webpage is created using HTML and python using Flask framework.

**RESULT**

The Chronic Kidney Disease Analysis Model is analysed using the Logistic Regression and the predicted accuracy is 0.975. On using the KNN algorithm the predicted accuracy is 0.9625.

**ADVANTAGES AND DISADVANTAGES**

The advantages that the model proposes are:

1. User friendly interface.
2. Easy results with one click.
3. Accuracy in results.

The disadvantage is that the dataset was smaller as compared to real life scenario.

**APPLICATIONS**

The model can be applied in real life scenario and people suffering from the disease or having certain symptoms of the disease can surely benefit from it.The dataset should be expanded so that the model works better.

**CONCLUSION**

In this project, we have proposed methods for diagnosing CKD in patients using Machine Learning techniques.The two methods that we used are the Logistic Regression and the KNN algorithm.The system was implemented using the models and their performance was evaluated.Logistic Regression provided the highest accuracy of 0.975. A GUI was made based on the model for use in the medical industry.

**REFERENCE**

1.https://www.kidney.org/atoz/content/about-chronic-kidney-disease

2.https://www.kidney-international.org/article/S0085-2538(15)50698-4/fulltext

3. Kaggle