**1.Title of the project**

**1.1 Introduction**

**1.2 Objectives of Research**

**1.3 Problem statement**

**2.Review of Literature**

**3.Data Collection**

**4.Methodology**

**4.1 Exploratory Data Analysis**

**4.1.1Figures and Tabels**

**4.2 Statistical techniques and data visualization**

**4.3 Data Modeling Using Supervised ML techniques**

**5.Finding and Suggestions**

**6.Conclusion**

**1.Survival Analysis For Chronic Kidney Disease.**

**1.1Introduction**: Chronic kidney disease (CKD) is a long-term condition where the kidneys do not work as well as normal.

CKD does not usually cause any symptoms until it has reached an advanced stage. It is usually detected at earlier stages by blood and urine tests. The main symptoms of advanced kidney disease include:

* tiredness
* swollen ankles, feet or hands (due to water retention)
* shortness of breath
* nausea
* blood in urine

**1.2Objectives of Research:**

* Improve cardiovascular care in persons with chronic kidney disease
* Reduce kidney failure due to diabete
* Improve vascular access for hemodialysis patients
* Increase the proportion of patients with treated chronic kidney failure who receive a transplant
* Reduce the number of deaths among persons with chronic kidney disease

**1.3 Problem Statement:**

CKD is very common and is mainly associated with ageing. The older you get, the more likely you are to have some degree of kidney disease. It is estimated that about one in five men and one in four women between the ages of 65 and 74 has some degree of CKD.The most common cause of CKD is damage caused by other chronic conditions, such as high blood pressure (hypertension) and diabetes.

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease

**2.Review of Literature:**

Chronic kidney disease (CKD) is defined by persistent urine abnormalities, structural abnormalities or impaired excretory renal function suggestive of a loss of functional nephrons. The majority of patients with CKD are at risk of accelerated cardiovascular disease and death. For those who progress to end-stage renal disease, the limited accessibility to renal replacement therapy is a problem in many parts of the world. Risk factors for the development and progression of CKD include low nephron number at birth, nephron loss due to increasing age and acute or chronic kidney injuries caused by toxic exposures or diseases (for example, obesity and type 2 diabetes mellitus).

**3.Data Collection:**

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. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables includes the amount of bp, sg, al, su, rbc, pc, ba, sc, sod, pot, hemo, dm, cad and Outcome is classification.

**4.Methodology**

Brief Description of Algorithms Used:

**Support Vector Machine (SVM):**

SVM is one of the standard set of supervised machine learning model employed in

classification. Given a two-class training sample the aim of a support vector

machine is to find the best highest-margin separating hyper plane between the two

classes[26]. For better generalization hyper plane should not lies closer to the data

points belong to the other class. Hyper plane should be selected which is far from

the data points from each category.

**Decision Tree Classifier:**

Decision Tree is a supervised machine learning algorithm used to solve

classification problems. The main objective of using Decision Tree in this research

work is the prediction of target class using decision rule taken from prior data. It

uses nodes and internodes for the prediction and classification. Root nodes classify

the instances with different features. Root nodes can have two or more branches

while the leaf nodes represent classification. In every stage, Decision tree chooses

each node by evaluating the highest information gain among all the attributes.

**Random Forest:**

Random Forest is a flexible, easy to use machine learning algorithm that produces,

even without hyper-parameter tuning, a great result most of the time. It is also one

of the most used algorithms, because it’s simplicity and the fact that it can be used

for both classification and regression tasks. In this post, you are going to learn,

how the random forest algorithm works and several other important things about it.

**K-Nearest Neighbors:**

K-Nearest Neighbors (KNN) is one of the simplest algorithms used in Machine

learning for regression and classification problem. KNN algorithms use a data and

classify new data points based on a similarity measures (e.g. distance function).

Classification is done by a majority vote to its neighbors. The data is assigned to

the class which has the most nearest neighbors. As you increase the number of

nearest neighbors, the value of k, accuracy might increase.

**Logistic Regression:**

Logistic regression is named for the function used at the core of the method, the

logistic function. The logistic regression, also called the sigmoid function was

developed by statisticians to describe properties of population growth in ecology,

rising quickly and maxing out at the carrying capacity of the environment. It’s an

S-shaped curve that can take any real-valued number and map it into a value

between 0 and 1.

**Accuracy Measures:**

SVM, Decision Tree, Random Forest, KNN, Logistic Regression algorithms are

used in this research work. Experiments are performed using internal cross-

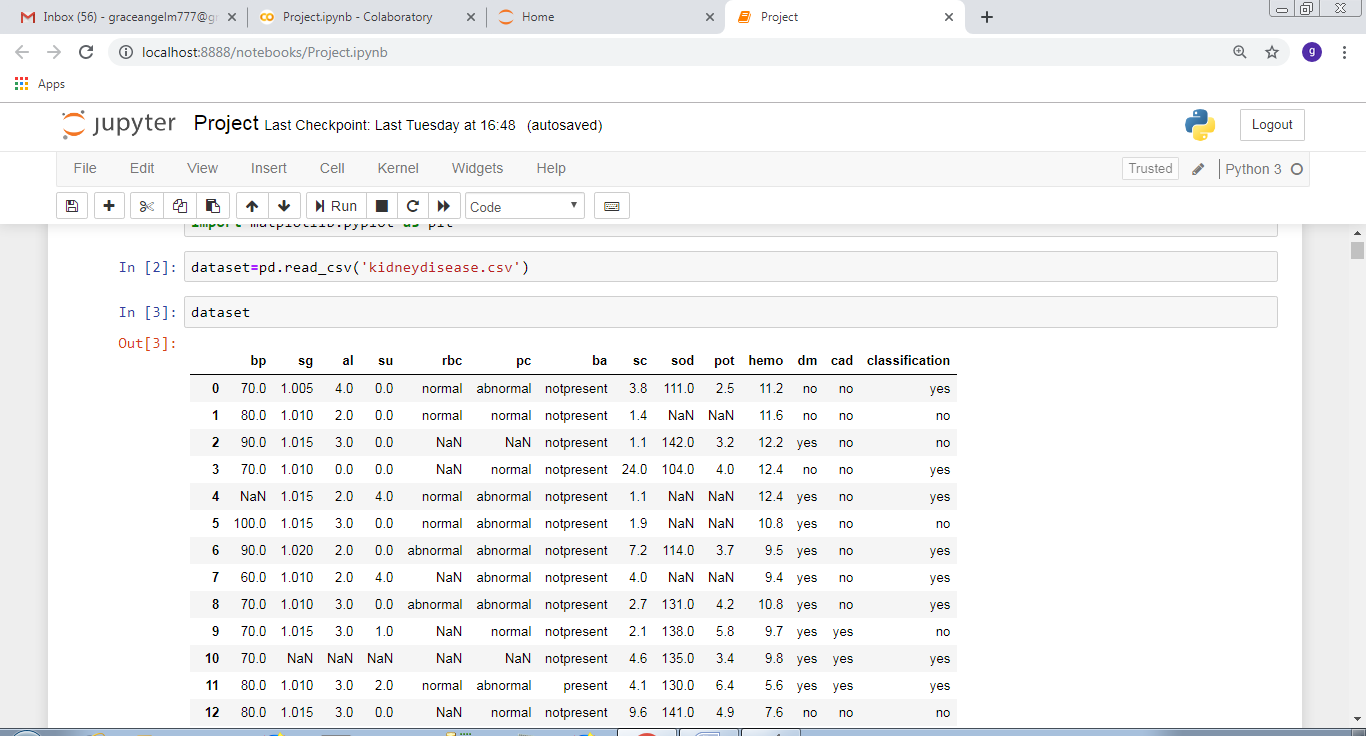
validation 10-folds. Accuracy, confusion matrix and ROC (Receiver Operating

Curve)measures are used for the classification of this work.

**4.1 Exploratory data analysis:**

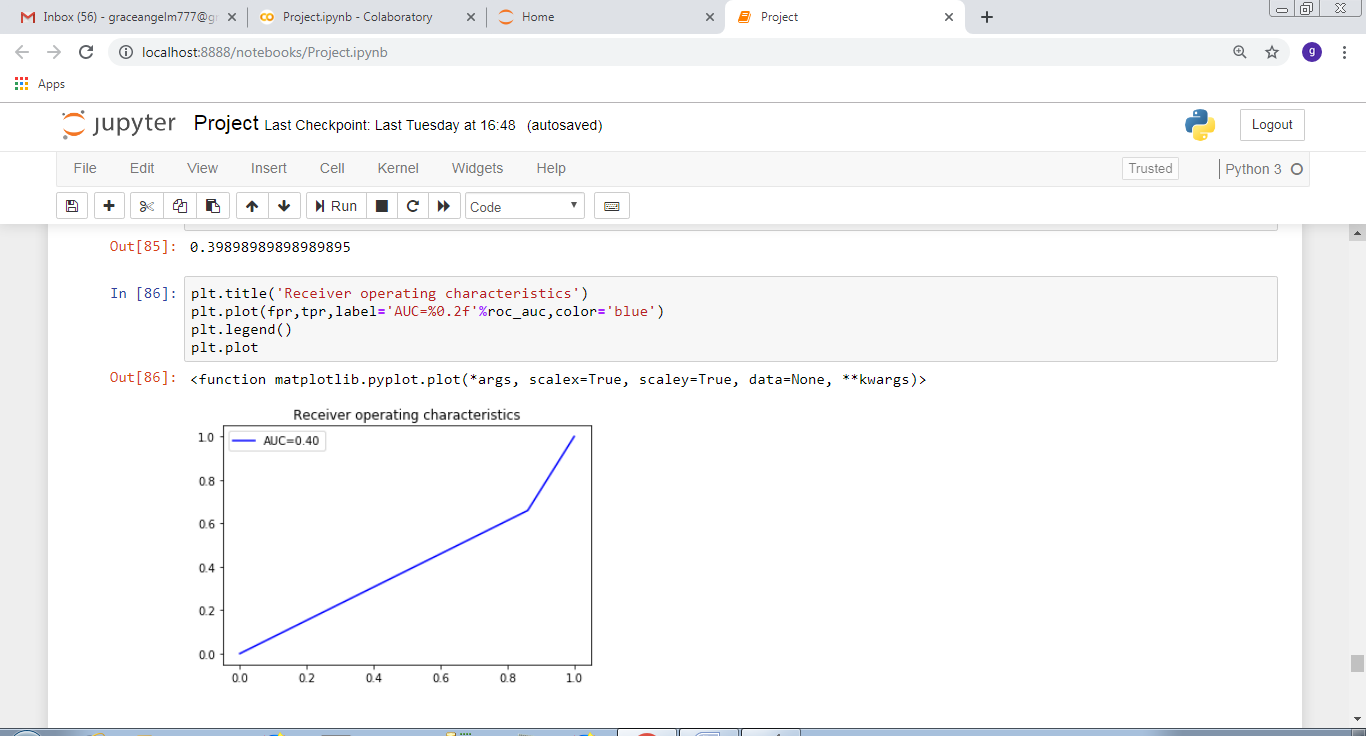
**4.1.1 Figures and tables:**

Dataset for Kidney Disease:

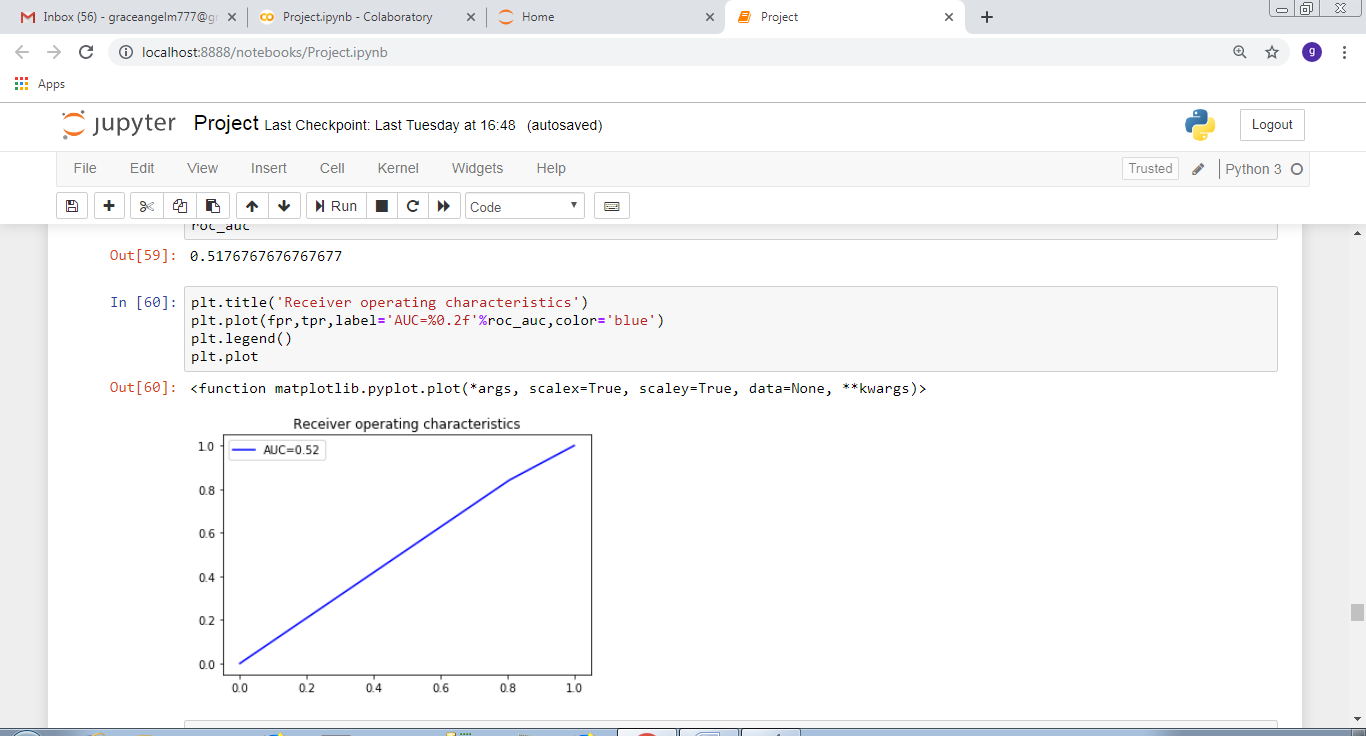


**AUC and ROC Curves:**

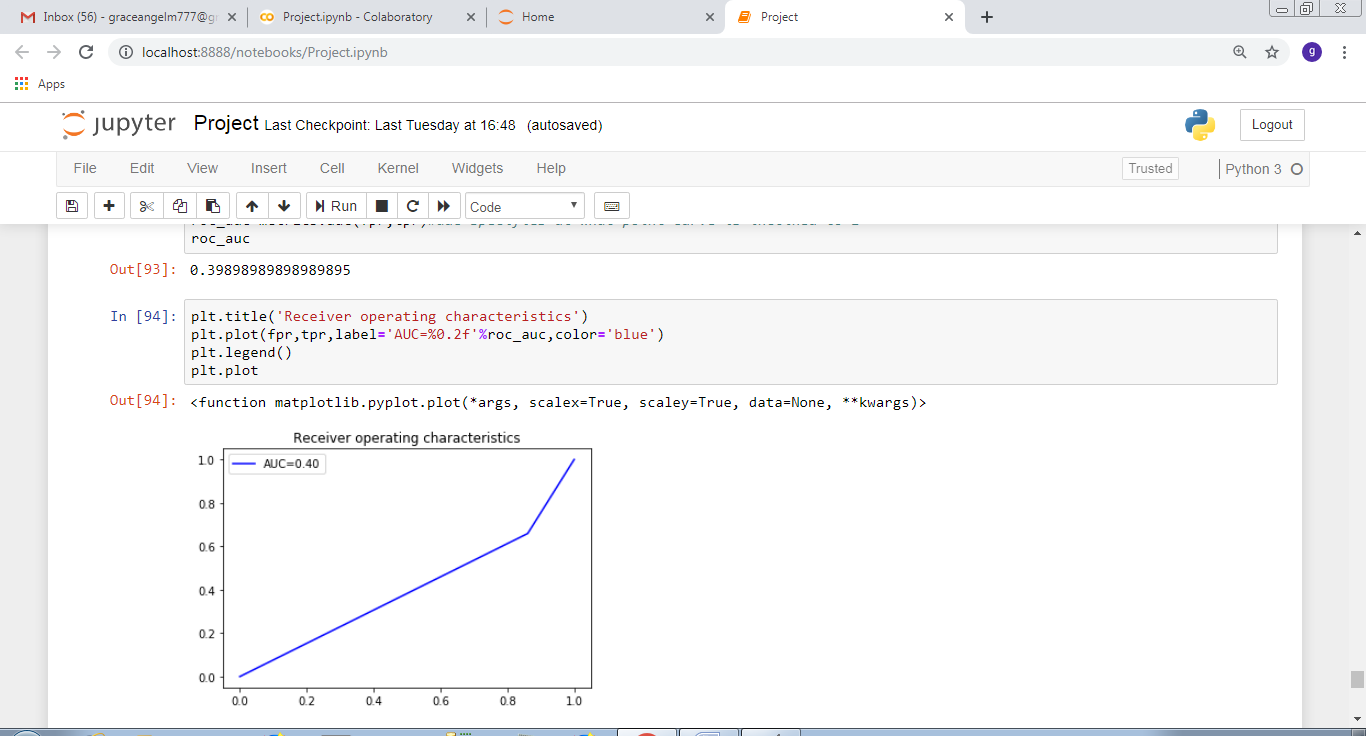
**Decision Tree:**

****

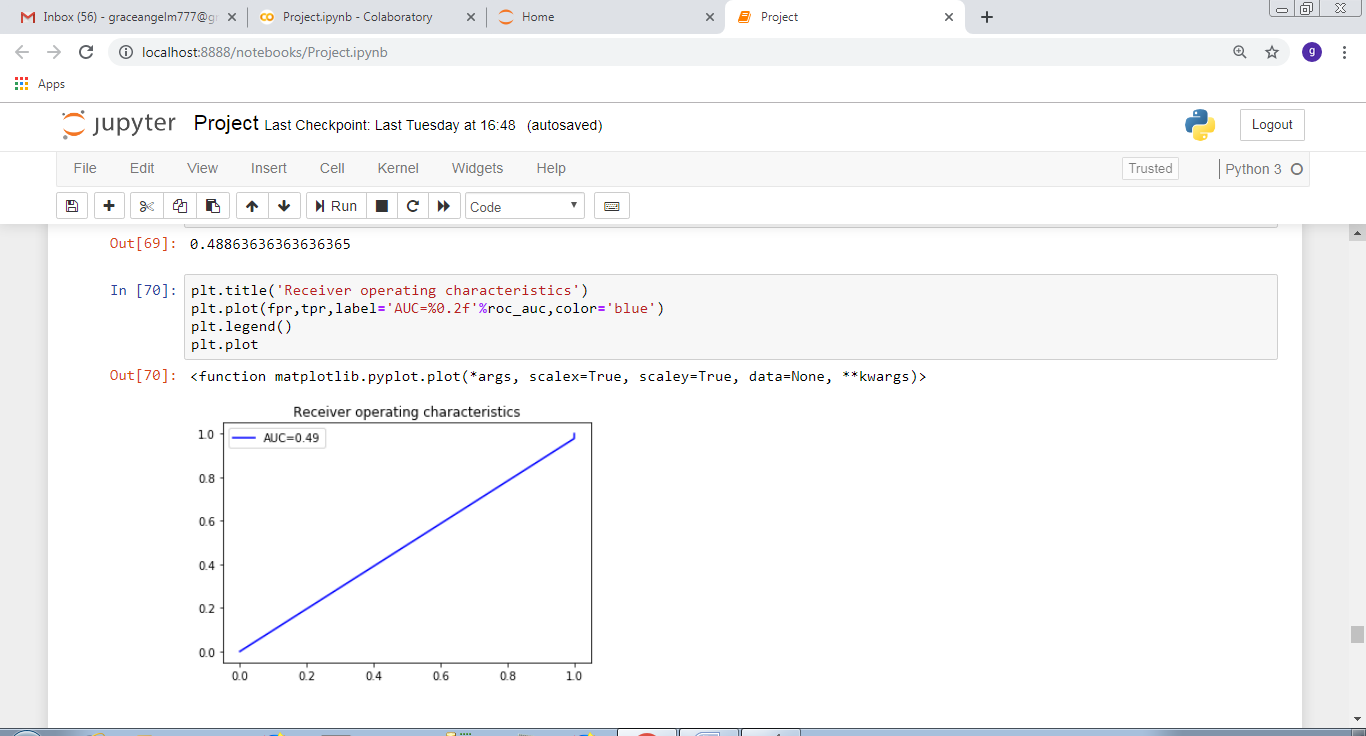
**KNN:(K-Nearest Neighbour):**

****

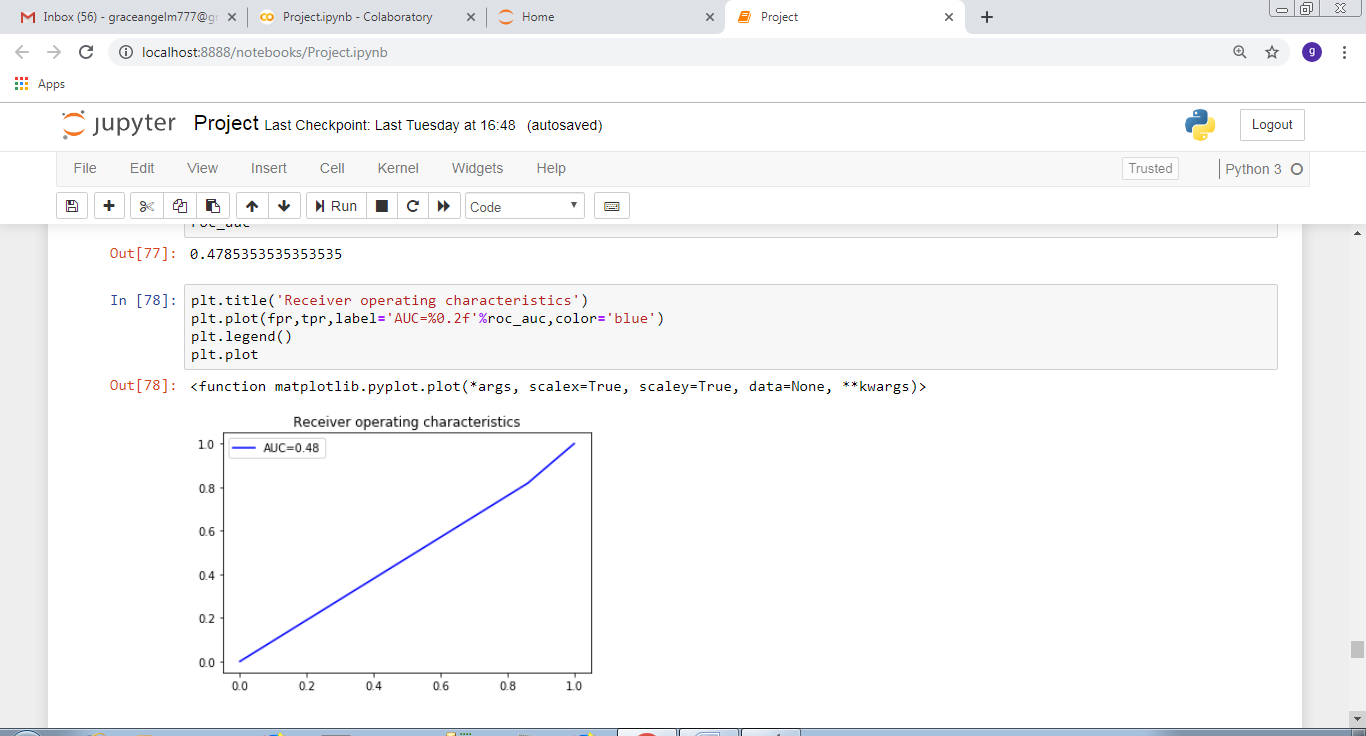
**SVM:(Support Vector Machine):**

****

**Logistic Regression:**

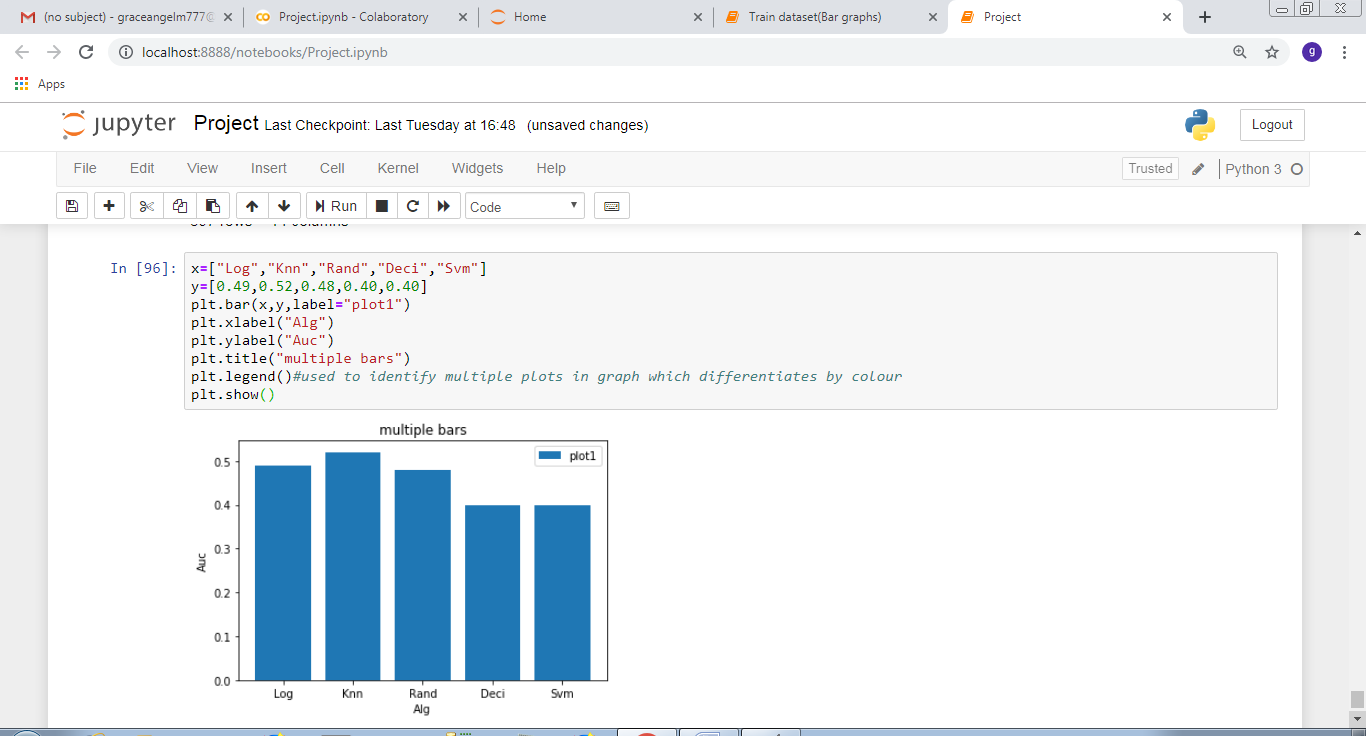
****

**Random Forest:**

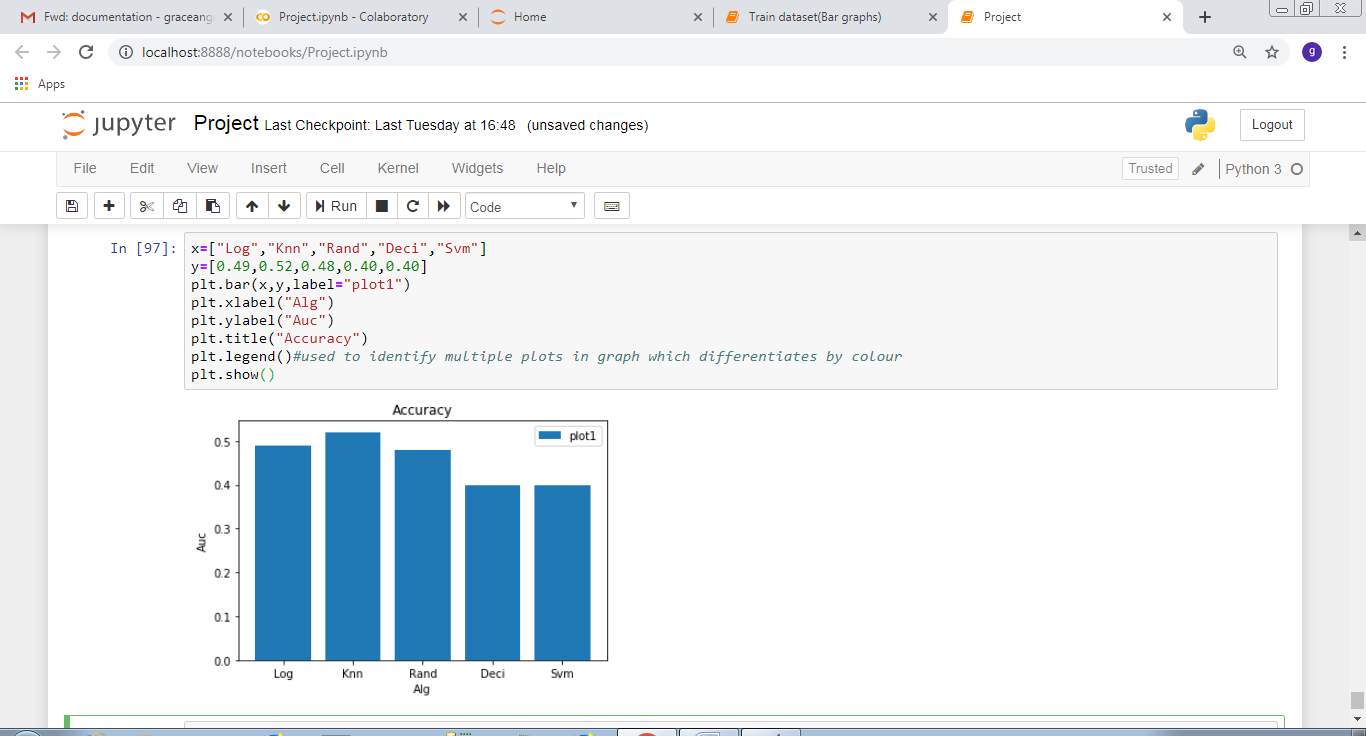
****

**Bar Graphs:**

**For AUC:**

****

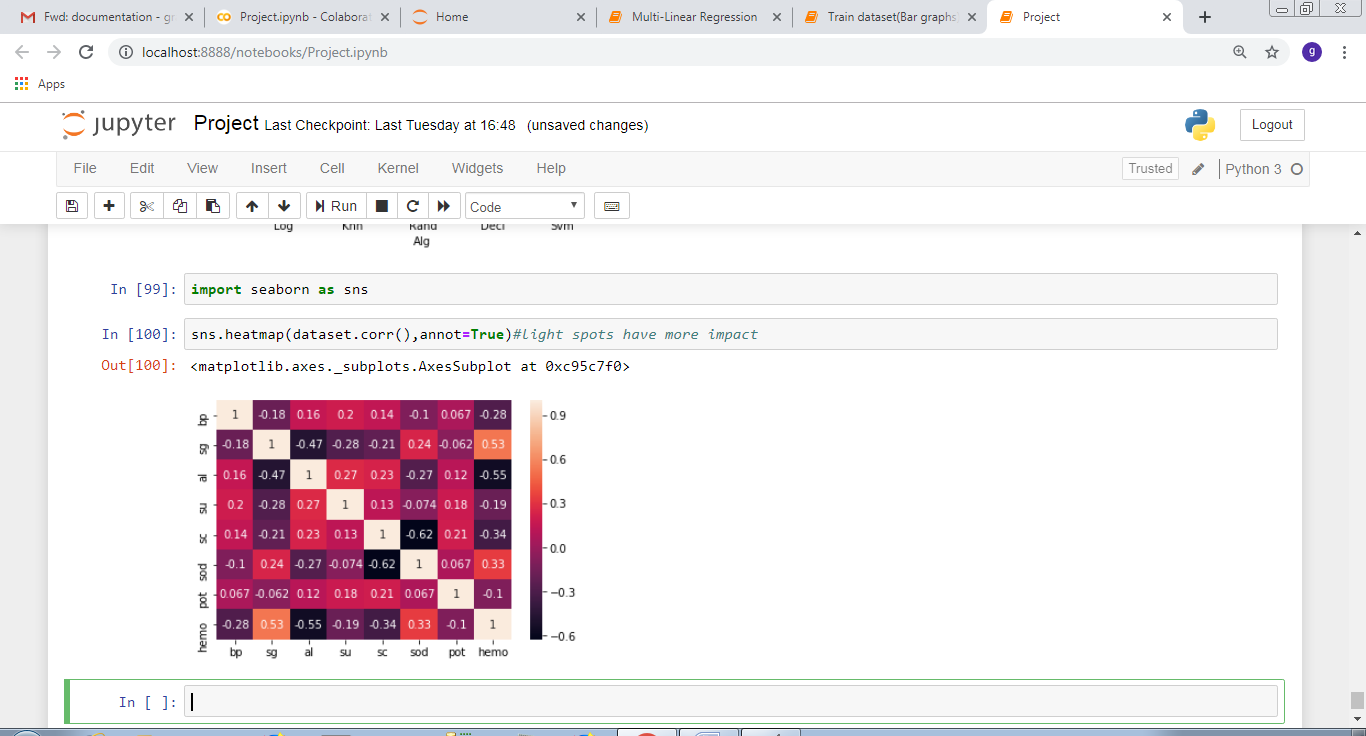
**For Accuracy:**

****

**4.2 Statistical techniques and data visualization:**

By importing matplotlib.pyplot library we have drawn graphs to demonstrate the AUC-ROC curves and by using bar graphs we have visualized the percentage levels of different techniques. we have used the co-relation function to demonstrate the impact of every

**Correlation graph:**

****

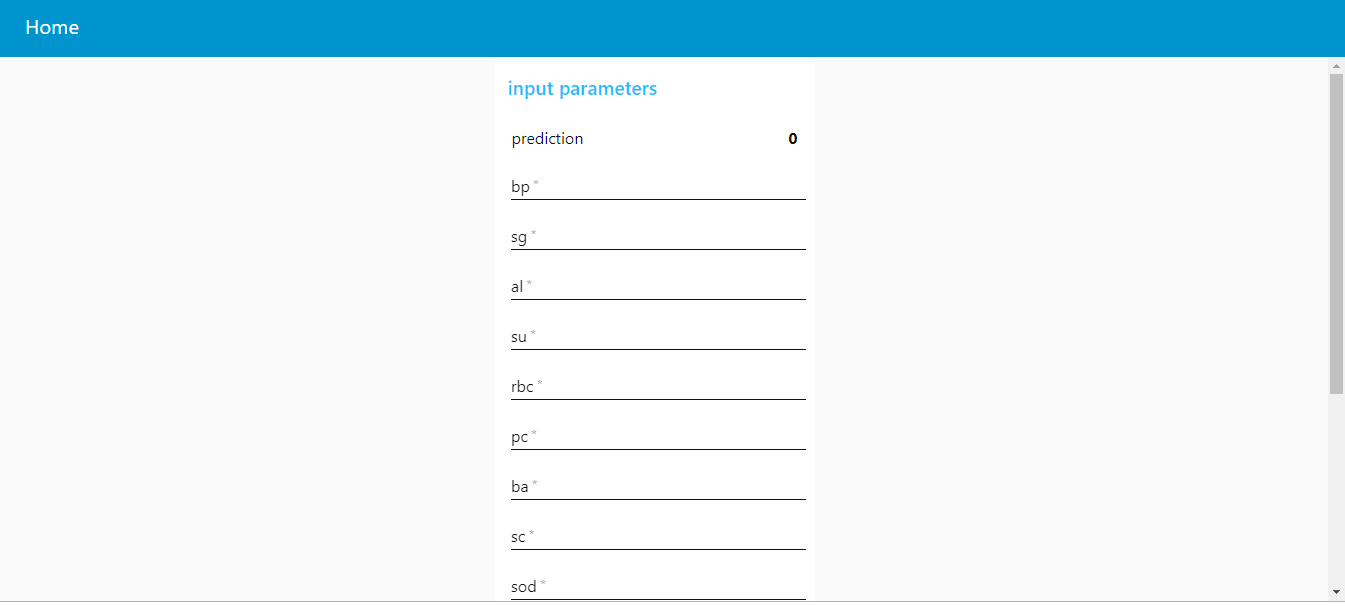
**4.3 Data modeling using supervised learning algorithms:**

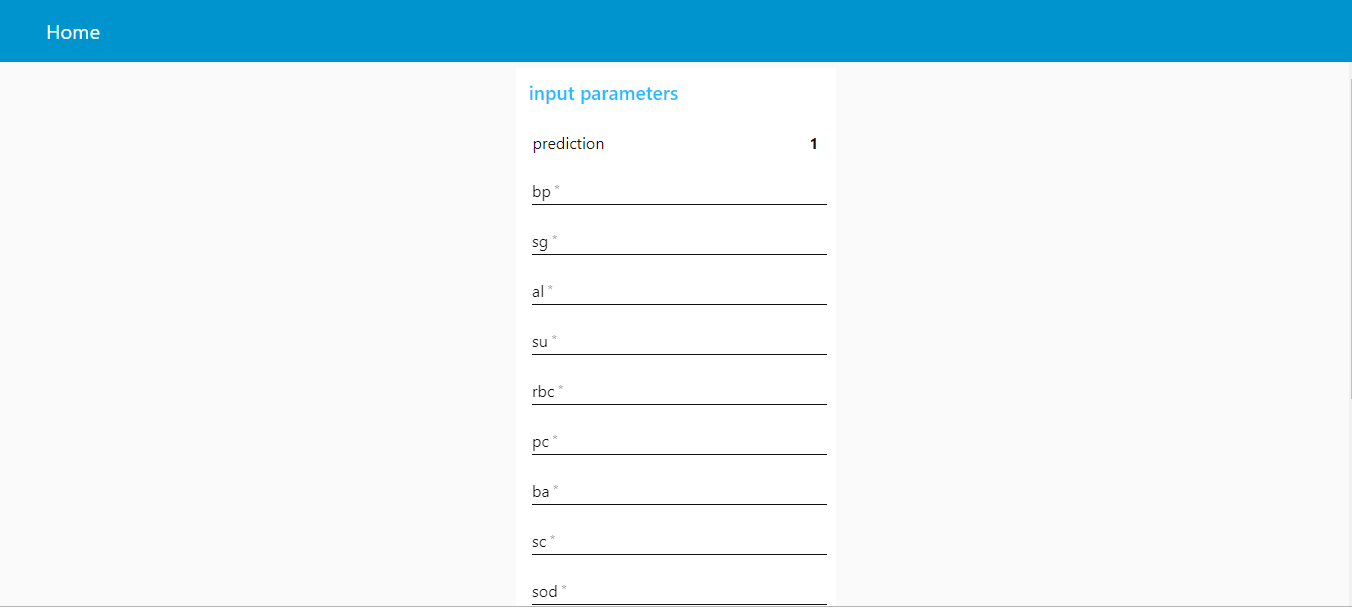
In general we have two types of learning algorithms, supervised and unsupervised learning algorithms. and in detail it consists of different techniques like,

* Support Vector Machines
* logistic regression
* decision trees
* k-nearest neighbor algorithm

since our model comes under supervised learning algorithm we applied every technique of the algorithm and based on the accuracy values we obtained we chosen k-nearest neighbour technique. and our model has  dependent variable which is classification(binary) means the output can be either a person is having disease or does not have disease.

**5.FINDINGS AND SUGGESTIONS:**

****

****

**6.CONCLUSION:**

One of the important real-world medical problems is the detection of kidney disease at its early stage. In this study, systematic efforts are made in designing a system which results in the prediction of kidney disease. During this work, five machine learning classification algorithms are studied and evaluated on various measures. Experiments are performed on kidney Database. Experimental results determine the adequacy of the designed system with an achieved accuracy of 0.52 using the K-Nearest Neighbour algorithm. In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases.

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease. Two classification methods were used, Naïve Bayes (NB) and

Support Vector Machines (SVM)

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease. Two classification methods were used, Naïve Bayes (NB) and

Support Vector Machines (SVM)

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease. Two classification methods were used, Naïve Bayes (NB) and

Support Vector Machines (SVM).

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease. Two classification methods were used, Naïve Bayes (NB) and

Support Vector Machines (SVM).

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease

Data mining and analytics are widely used to extract useful information from raw data. Today, data mining has

become an important field in healthcare that is used to detect unknown information in healthcare datasets and utilize

analytics to predict diseases. For chronic kidney disease (CKD), authors use predictive analytics models to predict

the disease based on its causes and develop models for its progression. A survey of the prediction models for CKD

was presented in [8]. The authors reviewed thirteen studies that describe 23 predictive models for kidney disease. In

[6], the authors discussed a text mining approach for CKD. The proposed approach extracts information from the

previously published literature. The use of artificial neural network models to predict kidney stones was presented in

[9]. Three different neural network algorithms were used and compared based on their accuracy, time to develop the

model, and size of training datasets. The use of data mining techniques for predicting kidney dialysis survivability

was discussed in [10]. Three predictive analytics models were used and compared. In [11], data mining classification

techniques were used to predict kidney disease