GOVERNMENT OF TAMILNADU

Naan Muthalvan - Project-Based Experiential Learning

Early Prediction For Chronic Kidney Disease Detection: A Progressive Approach to Health Management

Submitted by

TEAM ID: NM2023TMID22147

M. JOTHIKA (20326ER008)

K.DHIVYASHREE(20326ER005)

H.IMRANA PARVEEN (20326ER006)

P.JEYA CHITHRA(20326ER007)

Under the guidance of

Mrs. J. SUKANYA, MCA., M.Phil.,

Assistant Professor

PG and Research Department of Computer Science

M.V.MUTHIAH GOVERNMENT ARTS COLLEGE FOR WOMEN

B(Affiliated To Mother Teresa Women's University, Kodaikanal)

Reaccredited with "A" Grade by NAAC

DINDIGUL-624001.

APRIL - 2023

M.V.MUTHIAH GOVERNMENT ARTS COLLEG FOR WOMEN

(Affiliated to Mother Teresa Women's University, Kodaikanal)

Reaccredited with "A" Grade by NAAC

Dindigul - 624 001

PG & RESEARCH DEPARTMENT OF COMPUTER SCIENCEBONAFIDE

CERTIFICATE

This is to certify that this is a bonafide record of the project entitled, **"EARLY PREDICTION FOR CHRONIC KIDNEY DISEASE DETECTION"** done **by M. JOTHIKA (20326ER008) and K. DHIVYASHREE (20326ER005) H. IMRANA PARVEEN (20326ER006) and P. JEYACHITHRA (20326ER007).** This is submitted in partial fulfillment for the award of the degree of **Bachelor of Science in Computer science in M.V.MUTHIAH GOVERNMENT ARTS COLLEGE FOR WOMEN, DINDIGUL** during the period of December 2022 to April 2023.

Project Mentor(s) Head of the Department

Submitted for viva-voce Examination held on 11.04.2023

TABLE OF CONTENTS

ABSTRACT

Chronic Kidney Disease (CKD) or chronic renal disease has become a major issue with a steady growth rate. A person can only survive without kidneys for an average time of 18 days, which makes a huge demand for a kidney transplant and Dialysis. It is important to have effective methods for early prediction of CKD. Machine learning methods are effective in CKD prediction. This work proposes a workflow to predict CKD status based on clinical data, incorporating data prepossessing, a missing value handling method with collaborative filtering and attributes selection.

Out of the 11 machine learning methods considered, the extra tree classifier and random forest classifier are shown to result in the highest accuracy and minimal bias to the attributes. The research also considers the practical aspects of data collection and highlights the importance of incorporating domain knowledge when using machine learning for CKD status prediction.

1. INTRODUCTION

1.1 Overview:

Chronic Kidney Disease (CKD) is a major medical problem and can be cured if treated in the early stages. Usually, people are not aware that medical tests we take 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. The information says that it helps us to measure the severity of the problem, the predicted survival of the patient after the illness, the pattern of the disease and work for curing the diseaseIn todays world as we know most of the people are facing so many disease and as this can be cured if we treat people in early stages this project can use a pretrained model to predict the Chronic Kidney Disease which can help in treatments of peoples who are suffer from this disease.

1.2 Purpose

Chronic kidney disease (CKD) occurs when the kidneys gradually lose their function over time. There are several potential causes of CKD, including

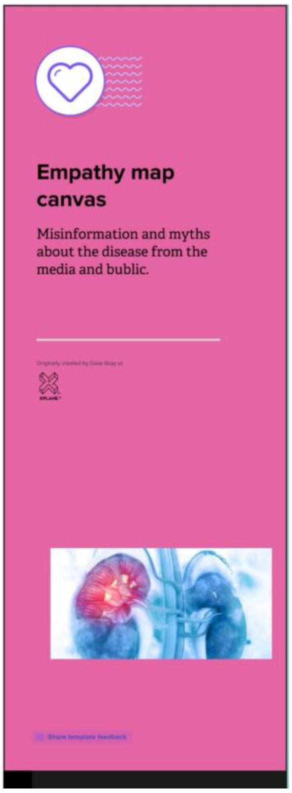
1. Diabetes: High blood sugar levels can damage the small blood vessels in the kidneys, making them less effective at filtering waste and excess fluid from the body.
2. High blood pressure: High blood pressure can damage the small blood vessels in the kidneys, as well as the delicate filtering units called nephrons.
3. Glomerulonephritis: This is an inflammation of the tiny filters in the kidneys, which can cause scarring and permanent damage.
4. Polycystic kidney disease: This is a genetic condition where cysts develop in the kidneys, impairing their function.

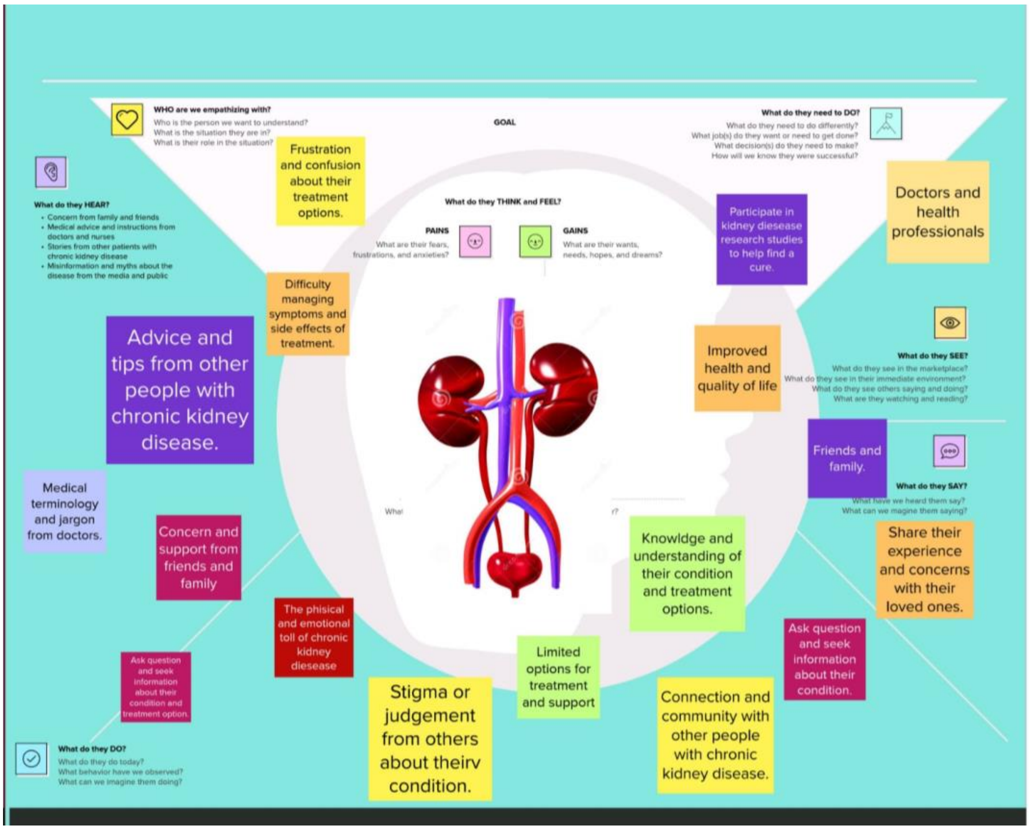
Other conditions: Certain autoimmune diseases, infections, and genetic disorders can also lead to CKD

The purpose of the kidneys is to filter waste and excess fluid from the blood, maintaining a healthy balance of electrolytes and other substances in the body.

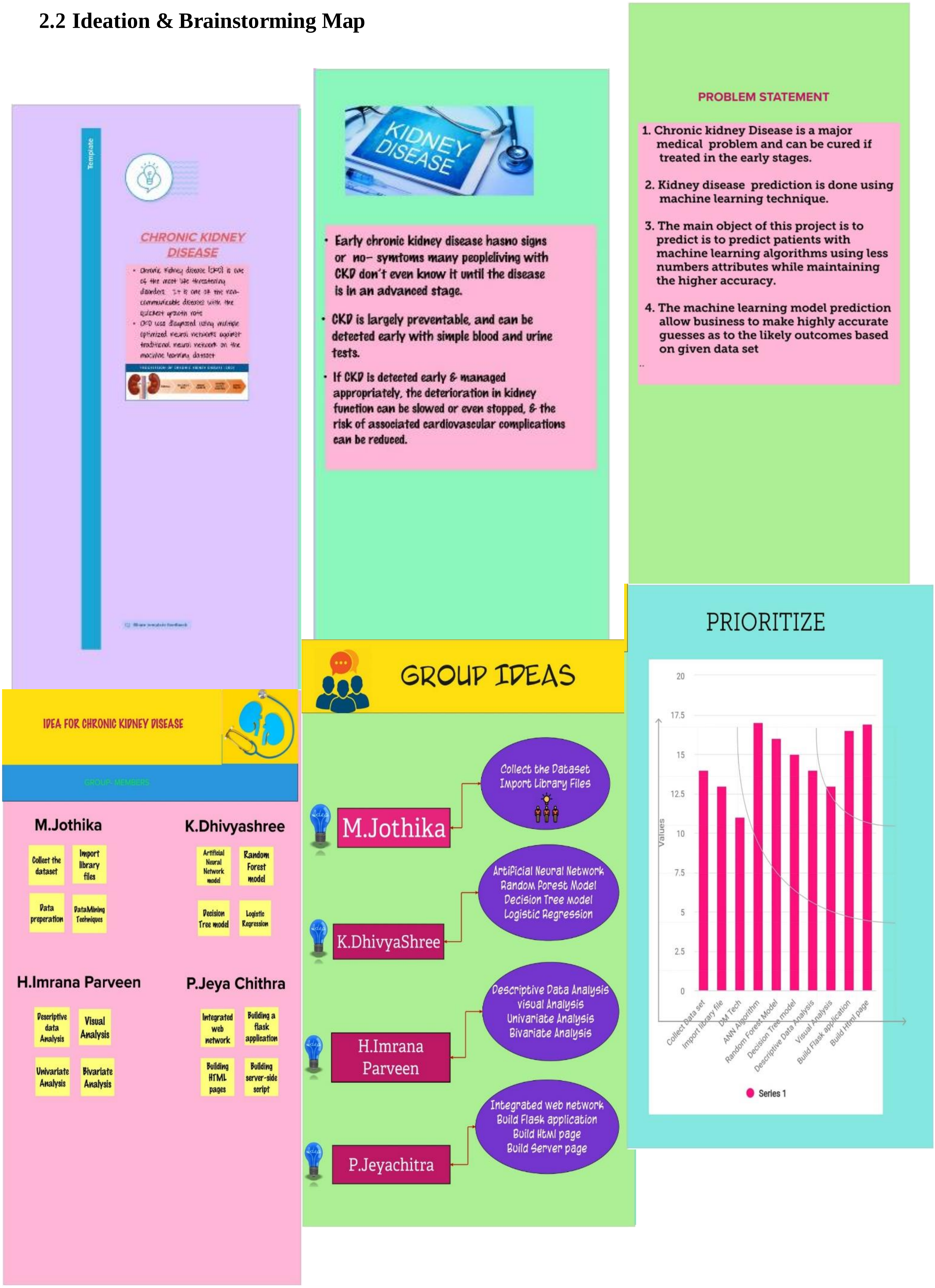
2. Problem definition & Design Thinking

2.1 Empathy Map



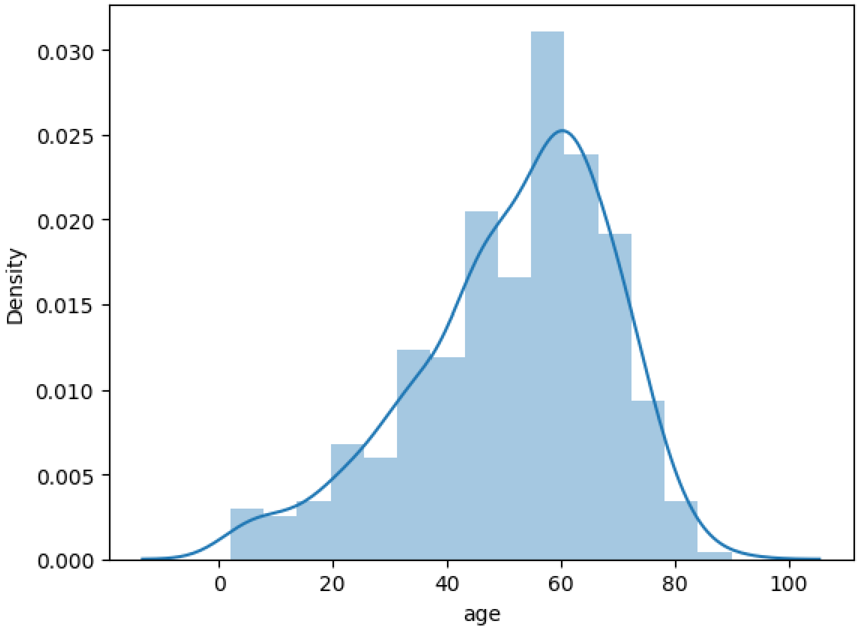


2.2 Ideation & Brainstorming Map

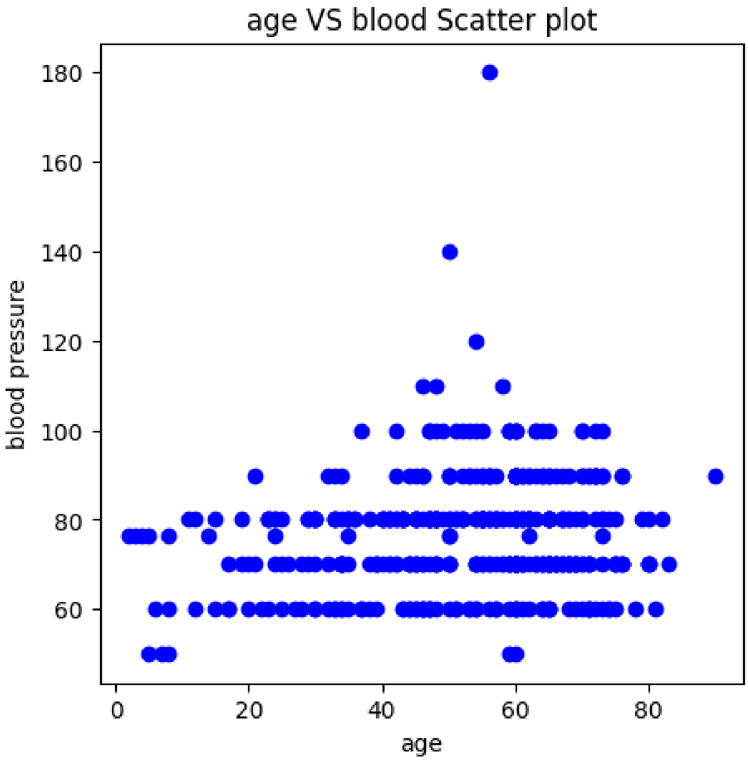


3. Result

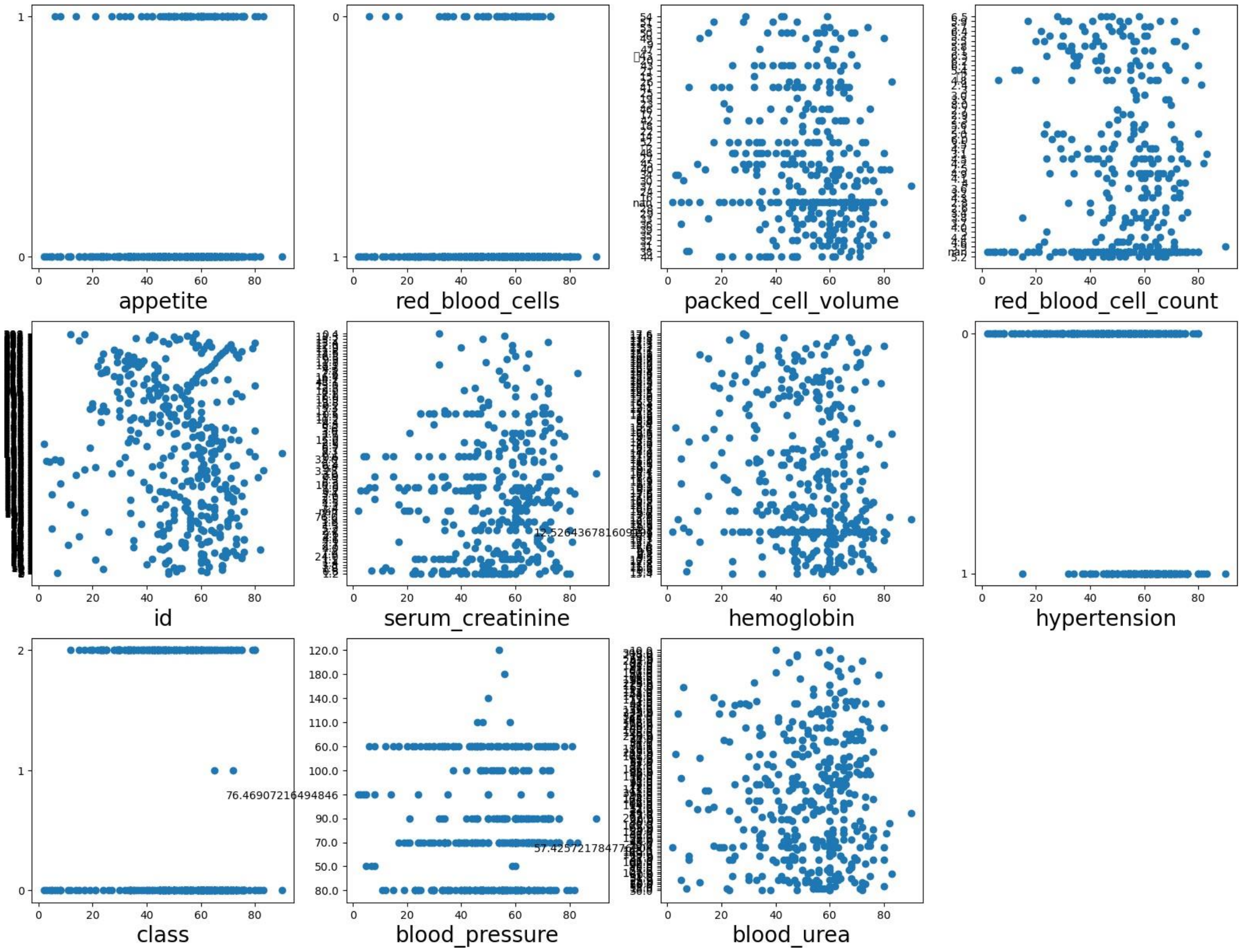
3.1 Univariate analysis

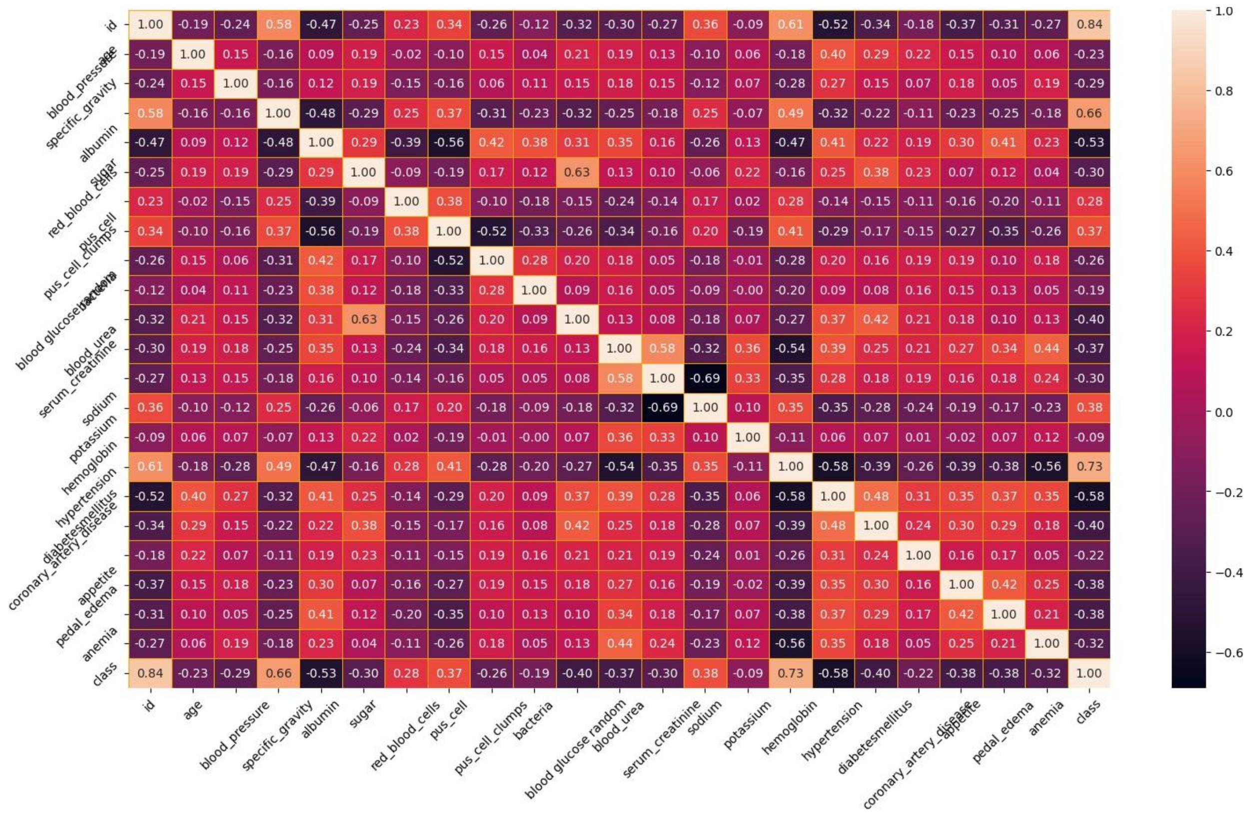


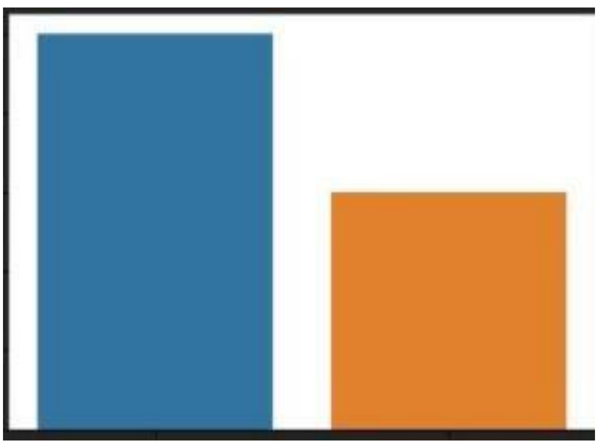
3.2 Bivariateanalysis

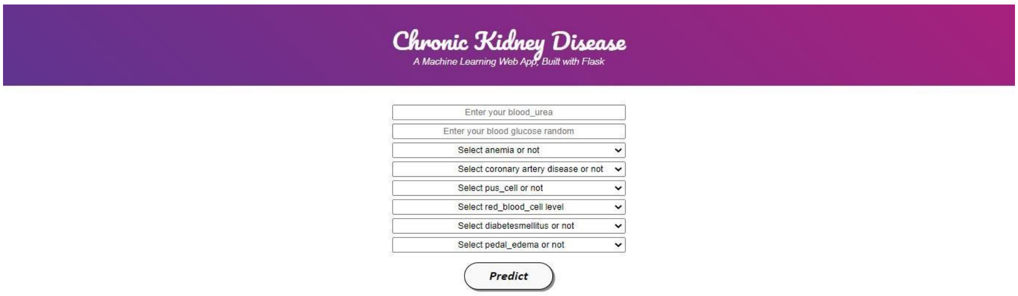


3.3 Multivariate analysis

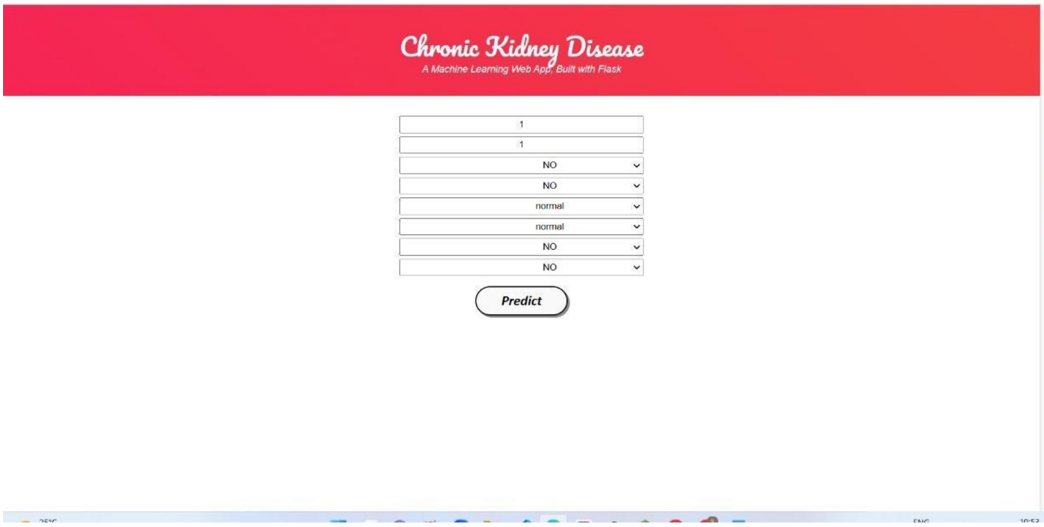



id
age
blood_pressure
specific_gravity
albumin
sugar
red_blood_cells
pus_cell
pus_cell_clumps
bacteria
blood glucose random
blood_urea
serum_creatinine
sodium
potassium
hemoglobin
hypertension
diabetesmellitus
coronary_artery_disease
appetite
pedal_edema
anemia
class
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6





Input - Now, the user will give inputs to get the predicted result after clicking onto the submit button.


Chronic Kidney Disease
A Machine Learning Web App, Built with Flask
1
1
NO
NO
normal
normal
NO
NO
Predict


Chronic Kidney Disease
A Machine Learning Web App, Built with Flask
Prediction: Great! You DON'T have Chronic Kidney Disease
IT'S TIME TO PARTYYYYYYY !

4. ADVANTAGES & DISADVANTAGES

ADVANTAGES

- Early detection and treatment of CKD can slow down or stop the progression of the disease, which can help prevent or delay the need for dialysis or a kidney transplant.
- CKD may be detected during routine blood or urine tests, which can lead to earlier detection of other health problems that may be contributing to the kidney disease.
- CKD may lead to changes in lifestyle that can improve overall health, such as following a healthy diet, quitting smoking, and increasing physical activity

DISADVANTAGES

- CKD can increase the risk of serious health problems, such as heart disease, stroke, and bone disease.
- The symptoms of CKD may be vague and nonspecific, making it difficult to detect until the disease has progressed to a later stage.
- Treatment for CKD can be expensive and time-consuming, requiring regular doctor visits, medications, and potentially dialysis or a kidney transplant.
- CKD can have a significant impact on quality of life, including fatigue, sleep disturbances, and changes in appetite or body weight.
- CKD can also lead to feelings of anxiety, depression, or stress, especially if it requires major lifestyle changes or interferes with daily activities.

5. APPLICATION

Chronic kidney disease (CKD) has several applications in healthcare, including:

1. Diagnosis: CKD can be diagnosed using blood and urine tests that measure kidney function and detect any abnormalities. These tests can help healthcare providers identify the disease early, allowing for early intervention and better outcomes.
2. Monitoring: Once diagnosed, CKD can be monitored over time to track changes in kidney function and determine the effectiveness of treatment. Monitoring may include regular blood and urine tests, imaging studies, and other diagnostic tests.
3. Treatment: CKD can be treated using a variety of approaches, including medications, lifestyle changes, and dialysis or kidney transplant in advanced cases. Healthcare providers may use a combination of these treatments to manage the disease and prevent complications.
4. Prevention: Individuals at risk of CKD, such as those with diabetes, high blood pressure, or a family history of kidney disease, may benefit from early detection and intervention to prevent the disease from developing.
5. Research: CKD research has led to advancements in understanding the disease, identifying risk factors, and developing new treatments. Ongoing research in this area may lead to further improvements in CKD diagnosis, treatment, and prevention.

6. CONCLUSION

Individuals at risk of CKD should be vigilant about monitoring their kidney function and taking steps to prevent the disease from developing or progressing. Healthcare providers play a critical role in diagnosing, monitoring, and treating CKD, and ongoing research in this area may lead to further improvements in CKD outcomes. Overall, managing CKD requires a comprehensive, multidisciplinary approach that involves healthcare providers, patients, and their families.

In conclusion, chronic kidney disease (CKD) is a serious medical condition in which the kidneys gradually lose their ability to function properly over time. While early detection and intervention can slow or stop the progression of the disease, CKD can increase the risk of serious health problems, such as heart disease and bone disease. Treatment for CKD can be expensive and time-consuming, and may include medications, lifestyle changes, and in some cases, dialysis or a kidney transplant.

7.FUTURE SCOPE

There are several areas of future research and development in chronic kidney disease (CKD) that have the potential to improve diagnosis, treatment, and outcomes. Here are some examples:

1. Early Detection: Researchers are exploring new biomarkers that can detect CKD at an earlier stage, before significant kidney damage has occurred. These biomarkers may include urine and blood tests, imaging studies, and other diagnostic tools.
2. Precision Medicine: Advances in genetics and personalized medicine may lead to more targeted treatments for CKD based on an individual's unique genetic makeup and other factors.
3. Regenerative Medicine: Researchers are exploring new techniques for repairing or regenerating damaged kidney tissue, such as stem cell therapy and tissue engineering.
4. Artificial Kidneys: Development of artificial kidneys that can replace the function of damaged kidneys, potentially reducing the need for dialysis or transplantation.
5. Telemedicine: Telemedicine and remote monitoring technologies may improve access to care for patients with CKD, especially in underserved areas or regions with limited access to specialists.
6. Patient Education: Improved patient education and engagement can help individuals with CKD better understand their condition, manage their symptoms, and adhere to treatment plans.

8.APPENDIX

A.SOURCE CODE

```
import pandas as pd
import numpy as np
from collections import Counter as c
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
from sklearn.metrics import accuracy_score,confusion_matrix
from sklearn.model_selection import train_test_split
from  sklearn.preprocessing  import  LabelEncoder
from sklearn.linear_model import LogisticRegression
import pickle
data=pd.read_csv("/content/kidney_disease.csv")
data.head()

data.columns

data.columns=['id','age','blood_pressure','specific_gravity','albumin',
                              'sugar','red_blood_cells','pus_cell','pus
_cell_clumps','bacteria',
                              'blood glucose random','blood_urea','seru
m_creatinine','sodium','potassium',
                               'hemoglobin','packed_cell_volume','white
_blood_cell_count','red_blood_cell_count',
                               'hypertension','diabetesmellitus','coron
ary_artery_disease','appetite',
                               'pedal_edema','anemia','class']
data.columns

data.info()

data.isnull().any()

data['blood glucose random'].fillna(data['blood glucose random'].mean()
,inplace=True)
data['blood_pressure'].fillna(data['blood_pressure'].mean(),inplace=Tru
e)
data['blood_urea'].fillna(data['blood_urea'].mean(),inplace=True)
data['hemoglobin'].fillna(data['hemoglobin'].mean(),inplace=True)
data['packed_cell_volume'].fillna(data['packed_cell_volume'].mean(),inp
lace=True)
data['potassium'].fillna(data['potassium'].mean(),inplace=True)
data['red_blood_cell_count'].fillna(data['red_blood_cell_count'].mean()
,inplace=True)
```

```
data['serum_creatinine'].fillna(data['blood_urea'].mean(),inplace=True)
data['sodium'].fillna(data['sodium'].mean(),inplace=True)
data['white_blood_cell_count'].fillna(data['white_blood_cell_count'].me
an(),inplace=True)


data['age'].fillna(data['age'].mode()[0],inplace=True)
data['hypertension'].fillna(data['hypertension'].mode()[0],inplace=True
)
data['pus_cell_clumps'].fillna(data['pus_cell_clumps'].mode()[0],inplac
e=True)
data['appetite'].fillna(data['appetite'].mode()[0],inplace=True)
data['albumin'].fillna(data['albumin'].mode()[0],inplace=True)
data['pus_cell'].fillna(data['pus_cell'].mode()[0],inplace=True)
data['red_blood_cells'].fillna(data['red_blood_cells'].mode()[0],inplac
e=True)
data['coronary_artery_disease '].fillna(data['coronary_artery_disease']
.mode()[0],inplace=True)
data['bacteria'].fillna(data['bacteria'].mode()[0],inplace=True)
data['anemia'].fillna(data['anemia'].mode()[0],inplace=True)
data['sugar'].fillna(data['sugar'].mode()[0],inplace=True)
data['diabetesmellitus '].fillna(data['diabetesmellitus '].mode()[0],in
place=True)
data['pedal_edema'].fillna(data['pedal_edema'].mode()[0],inplace=True)
data['specific_gravity'].fillna(data['specific_gravity'].mode()[0],inpl
ace=True)


data.isnull().sum()

catcols=set(data.dtypes[data.dtypes=='O'].index.values)
print(catcols)

for i in catcols:
      print("Columns :",i)
      print(c(data[i]))
      print('*'*120+'\n')

catcols.remove('red_blood_cell_count')
catcols.remove('packed_cell_volume')
catcols.remove('white_blood_cell_count')
print(catcols)

catcols=['anemia','pedal_edema','appetite','bacteria','class','coronary
_artery_disease','diabetesmellit
'hypertension','pus_cell','pus_cell_clumps','red_blood_cells']

from sklearn.preprocessing import LabelEncoder
for i in catcols:
```

```
        print("LABEL ENCODING OF:",i)
        LEi = LabelEncoder()
        print(c(data[i]))
        data[i] = LEi.fit_transform(data[i])
        print(c(data[i]))
        print("*"*100)

contcols=set(data.dtypes[data.dtypes!='O'].index.values)
print(contcols)

for i in contcols:
      print("Continous Columns:",i)
      print(c(data[i]))
      print('*'*120+'\n')

contcols.remove('specific_gravity')
contcols.remove('albumin')
contcols.remove('sugar')
print(contcols)

contcols.add('red_blood_cell_count')
contcols.add('packed_cell_volume')
contcols.add('white_blood_cell_count')
print(contcols)

contcols.add('specific_gravity')
contcols.add('albumin')
contcols.add('sugar')
print(contcols)

data['coronary_artery_disease']=data.coronary_artery_disease.replace('\
tno','no')
c(data['coronary_artery_disease'])

data['diabetesmellitus'] = data.diabetesmellitus.replace(to_replace={'\
tno':'no':'\tyes':'yes':'yes':'}
c(data['diabetesmellitus'])

data.describe()
sns.distplot(data.age)
import matplotlib.pyplot as plt
fig=plt.figure(figsize=(5,5))
plt.scatter(data['age'],data['blood_pressure'],color='blue')
plt.xlabel('age')
plt.ylabel('blood pressure')
plt.title("age vs blood scatter plot")
plt.figure(figsize=(20,15), facecolor='white')
plotnumber = 1
```

```
for column in contcols:
    if plotnumber<=11:
        ax = plt.subplot(3,4,plotnumber)
        plt.scatter(data['age'],data[column])
        plt.xlabel(column,fontsize=20)

    plotnumber+=1
plt.show( )
plt.figure(figsize=(20,15), facecolor='white')
plotnumber = 1
for column in contcols:
    if plotnumber<=11:
        ax = plt.subplot(3,4,plotnumber)
        plt.scatter(data['age'],data[column])
        plt.xlabel(column,fontsize=20)

    plotnumber+=1
plt.show( )
f,ax=plt.subplots(figsize=(18,10))
sns.heatmap(data.corr(),annot=True,fmt=".2f",ax=ax,linewidths=0.5,linec
olor="orange")
plt.xticks(rotation=45)
plt.yticks(rotation=45)
plt.show()
sns.countplot(data['class'])
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
x_bal=sc.fit_transform(x)
selcols=['red_blood_cells','pus_cell','blood glurose random','blood_ure
a','pedal_edema','anemia','diabetesmellitus','coronary_artery_disease']
x=pd.DataFrame(data,columns=selcols)
y=pd.DataFrame(data,columns=['class'])
print(x.shape)
print(y.shape)
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random
_state=2)
```