EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

PROJECT REPORT

TEAM ID: PNT2022TMID01939

BRUNDHA B , ABINAYA V , ABINAYA K , PRATHEEKSHA Y

1. INTRODUCTION

1.1 PROJECT OVERVIEW:

Chronic kidney disease prediction is one of the most important issues in healthcare- analytics. The most interesting and challenging tasks in day-to-day lives as one third of the adult population is affected by chronic kidney disease (CKD), and millions die each year because they do not have access to affordable treatment. Chronic Kidney Disease can be cured, if treated in the early stages. The main aim of the project is to predict whether the patient have chronic kidney disease or not in a painless, accurate and fasterway based on certain diagnostic measurement like Blood Pressure (BP), Albumin (Al) etc.,and then appropriate treatment can be given based on the details provided by the model.

1.2 PURPOSE

The purpose of the project is to alert doctors for an early detection of kidney disease andhence ensure speedy recovery or prevention of kidney disease. This Project aims at creating a model for early detection of Chronic Kidney Disease using Machine Learning technology. The output is integrated with Flask. The front end developed in html is used to receive user input on various parameters needed to decide on the early detection of kidney disease. The same model is deployed into IBM cloud.

2. LITERATURE SURVEY:

2.1 EXISTING SYSTEM

Presently kidney disease is detected at late stages in many countries leading to loss of precious lives. There are very few means to identify them at an early stage. Most of the user details remain unverified and it's difficult to track the fake users. The user interface

of the application is not user friendly and the user must have a device with an androidoperating system with an active internet connection to interact with this application.

2.2 REFERENCES

- 1. "What Is ChronicKidney Disease?". NationalInstitute of Diabetesand Digestive and Kidney Diseases. June 2017. Retrieved 19 December 2017.
- 2. "Kidney Failure". MedlinePlus. Retrieved 11 November 2017.
- 3. "Chronic Kidney DiseaseTests & Diagnosis". National Institute of Diabetes and Digestive and Kidney Diseases. October 2016. Retrieved 19 December 2017.
- 4. "Kidney Failure". NationalInstitute of Diabetesand Digestive and Kidney Diseases. Retrieved 11 November 2017.
- 5. "Managing Chronic KidneyDisease". National Institute of Diabetes and Digestive and Kidney Diseases. October 2016.
- 6. "Eating Right for Chronic KidneyDisease | NIDDK". National Institute of Diabetesand Digestive and Kidney Diseases. Retrieved 5 September 2019.
- 7. "Mineral & Bone Disorder in Chronic KidneyDisease". National Institute Of Diabetes and Digestive and Kidney Diseases. November 2015. Retrieved19 December 2018

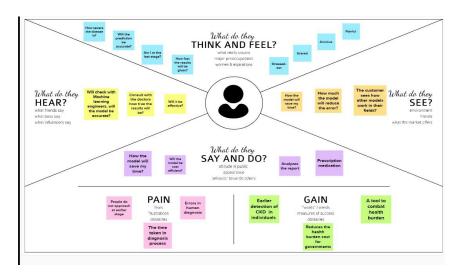
2.3 PROBLEM STATEMENT

Chronic Kidney Disease (CKD) is a serious medical condition that, if diagnosed early enough, is curable. Most individuals are unaware that the various medical tests we undergo for various reasons may provide important information aboutkidney disorders. As a result, characteristics of numerous medical tests are examined to see which characteristics might contain useful information about the disease. According to the information, doing so enables us to assess the problem's severity, and we utilize this data to create a machine learning model that forecasts chronic kidney disease.

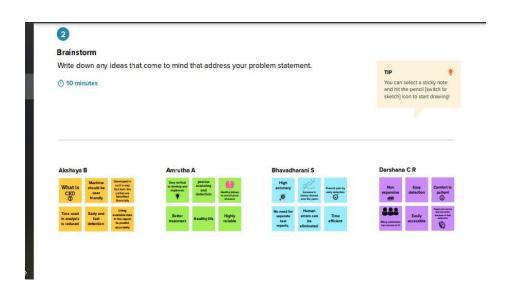
If chronic kidney disease is addressed early on, it may be cured. This project's primary goal is to more accurately and quickly identify whether a patient has chronic kidney disease using diagnostic data including Blood Pressure (Bp), Albumin, and other parameters (Al).

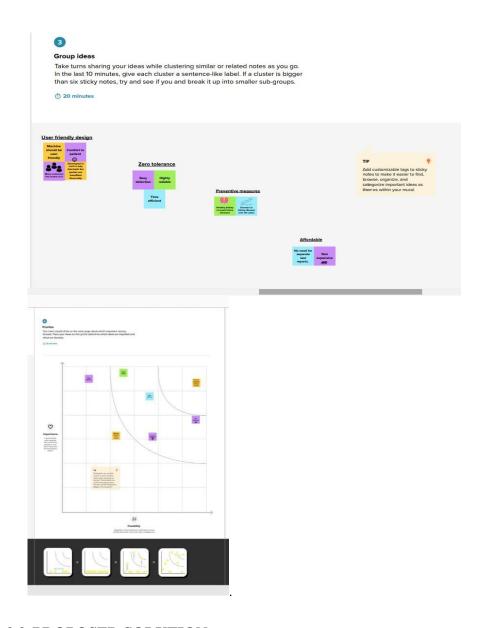
3. IDEATION & PROPOSED SOLUTION:

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAINSTORMING





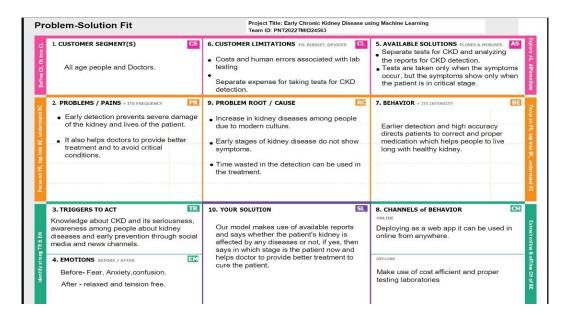
3.3 PROPOSED SOLUTION

S.no	Parameter	Description
1.	Problem Statement(Problemto be solved)	 Chronic Kidney Disease (CKD) is a major medical problem hence Chronic kidney disease prediction is one of the most important issues in healthcare analytics. 10% of the population worldwide is

		affected by chronic kidney disease (CKD),and millions die each year because they do not have access to affordable treatment. • Chronic kidney Disease can be cured, if treated in the early stages.
2.	Solution Description	 The idea is to detect the presence of kidney disease through machine learning based classification models. Early detection of chronic kidney disease is identified throughvarious ML Algorithms such as Logistics Regression, Random Forest, Decision Tree, Support Vector Machines and KNN. Using these techniques, each algorithm's effectiveness is evalutated. A web app is developed that asks basic user details about kidney details and result is produced
3.	Novelty	 Aims to find the best machine learning model for the early prediction of chronic kidney disease by analyzing the essential parameters and comparing their predictive accuracies. Then collaborate the best machine learning model to an interactive user- interface which helps in the earlydetection of CKD and provide cure.
4.	Social Impact	The main aim of this application is early prediction of chronic kidney

		disease that can possibly stop orslow the progression of this diseaseto the end stage.
5.	Business Model	 The widespread use of Machine Learning of predicting the CKD inthe Medical Industries promotes medical innovation, lowers medical expenses, and improves medical quality. To cure the CKD patients atearly stages. We can generate revenue through direct customers or can also collaborate with the health care sector and generate revenue from their customers.
6.	Scalability of solution	 Early prediction of CKD using Machine Learning that is more efficient to analyze the disease so that it can be cured on time. We can also use image data and apply Deep Learning techniques such as Multilayer Perceptron(MLP)etc., which will provide an improvedaccuracy than machine learning.

3.4 Problem Solution Fit



4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form
FR-2	User Confirmation	Confirmation via password.
FR-3	Obtain Information	The system should be able to get the information forpredicting the disease.
FR-4	Displaying Result	The system must be able to display whether the user isaffected or not

4.2 NON FUNCTIONAL REQUIREMENT

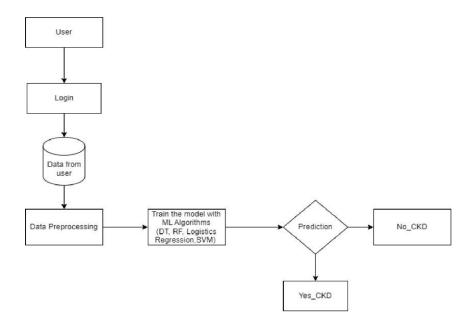
FR No.	Non-Functional Requirement	Description		
NFR-1	Usability	Access to use the application is permitted only to theregistered users		

NFR-2	Security	Authentication is done for the security process.
NFR-3	Reliability	The user gets the correct and predicted value and standard results.
NFR-4	Performance	Lowering the total load time of prediction and user interaction
NFR-5	Availability	Easily available to everyone.
NFR-6	Scalability	To be able to change things.

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored



5.2 SOLUTION AND TECHNICAL ARCHITECTURE

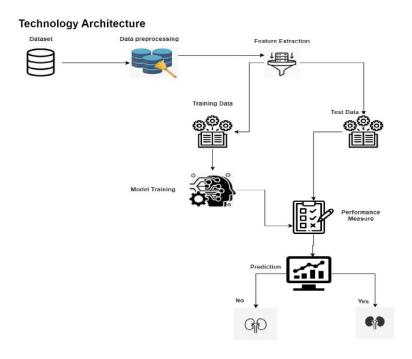


Table-1: Components & Technologies:

S.No	Component	Description	Technology
1	User Interface	How user interacts with application	HTML, CSS,Python Flask
2	Application Logic-1	Get input from the user	HTML,CSS,Python Flask
3	Application Logic-2	Predicts based on the provided input	Python
4	Application Logic-3	Displays the predicted Result	Python,HTML,CSS,Flask
5.	Machine Learning Model	Random Forest, Regression techniques, Decision tree and SVM	Classification Algorithms

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
------	-----------------	-------------	------------

1.	Open-Source Frameworks	List the open-source frameworks used	Google colab,Jupyter notebook,IBM cloud and Flask.
2.	Scalable Architecture	Model can be scalable	Python
3.	Availability	It is used as a website(UI) or available in cloud	IBM cloud
4.	Performance	High accuracy	Machine Learning Classification techniques

5.3 USER STORIES

User Type	Functional Requireme nt (Epic)	User Story Number	User Story	Acceptance criteria	Priority	Release
Customer	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	can access account	High	Sprint-1
	Login	USN-2	As a user, I can log into the application	Login into account	High	Sprint-1
Data Entry	Enter data	USN-3	Enter symptoms and clinical data	Enter ClinicalData	High	Sprint-2
Customer (View Result)	View Result	USN-4	Result can be viewed by the user	View Result	Medium	Sprint-3
Administrator	Predict Disease	USN-5	As the admin, I build ML models to predict CKD	I deploy models within websites	High	Sprint-1

6. PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND SCHEDULING

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

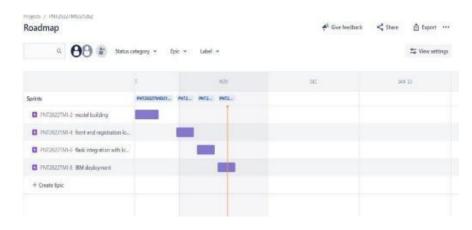
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data Collection	USN-1	Collect the appropriate dataset for detection of Chronic Kidney Disease.	10	High	Amrutha A
Sprint-1	Data Pre-processing	USN-2	Transform the collected data into useful readable format through data pre-processing	10	High	Bhavadharani S
Sprint-2	Model Building	USN-3	Identifying the variations in the test reports of a diseased person and calculating the severeness of damage	10	Medium	Akshaya B
Sprint-2	Model Building	USN-4	Splitting the available dataset as training and testing datasets for the model	10	Medium	Darshana C R
Sprint-3	Training and Testing	USN-5	Training the model using classification algorithm and testing the performance of the model.	10	High	Akshaya B
Sprint-3	Application Building	USN-6	Develop HTML code for user interface and Python code for connection of model	10	High	Darshana C R
Sprint-4	Application Building	USN-7	Running of Flask with connection of model to get the detection result from the model	10	Medium	Bhavadharani S
Sprint-4	Implementation of Application	USN-8	Completion of Application and Deployment in IBM Cloud	10	Medium	Amrutha A

6.2 SPRINT DELIVERY SCHEDULE

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.3 REPORTS FROM JIRA



7. CODING AND SOLUTIONING

7.1 FLASK DEPLOYMENT

Using Flask we are locally deploying our machine Learning model. Flask acts as a web Framework . Additionally we have app.py file to locally deploy the model

Home.html

```
Flask \templates \temp
```

Home.css

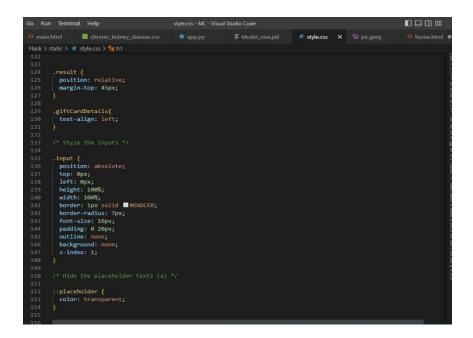
```
Run Terminal Help stylecss-ML-VsualStudio Code

main.html chronic kidney_disease.csv  app.py  Model_one.pkl  # stylecss  X  pic.jpeg  home.html

ask > static > # stylecss > % th

classication is a separate  in the care is a separate  in
```

```
Flesk > static > # style.css > % sh1
93
94
95
96
96
97
width: inherit;
98
padding: 20px 90px;
100
101
102
*title {
103
font-size: 50px;
104
margin-bottom: 50px;
105
106
}
107
108
**userProfile{
109
width: 400px;
110
display: block;
111
box-shadow: 0 10px 25px □rgba(92, 99, 105, .2);
112
margin: 0 auto;
113
padding: 10px 20px 10px 20px;
114
}
115
**inputContainer {
117
position: relative;
118
height: 45px;
119
width: 90%;
120
margin-bottom: 17px;
121
122
123
124
**result {
position: relative;
margin-top: 45px;
125
position: relative;
margin-top: 45px;
```



```
# style.css
    ○ home.html • ▷ 🏻 …

■ Model_one.pkl # style.css × □ pic.jpeg
    249 .nav-item a {
250 display: block;
251 color: white;
252 font-size:20px;
                                                                    text-align: center;
padding: 10px 20px;
text-align: centre;
padding: 16px 20px;
text-decoration: none;
text-
                                                     .home-page-image{
  width: 100%;
  padding-ton: 25m
         275 padding-top: 25px;
276 }
```

```
app.py
from flask import Flask, render_template, request, send_from_directory
         app = Flask(__name__)
       import pickle
model = pickle.load(open(r'Model_one.pkl', 'rb'))
         @app.route('/static/<path:path>')
         def send_report(path):
    return send_from_directory('static', path)
      @app.route('/')
def helloworld():
    return render_template("home.html")
gapp.route('/login', methods = ['POST'])
def login():
    a = request.form["age"]
    b = request.form["bp"]
    c = request.form["alb"]
    e = request.form["alb"]
    e = request.form["Busen"]
    f = request.form["Busen"]
    h = request.form["busenia"]
    h = request.form["busenia"]
    i = request.form["busenia"]
    i = request.form["busenia"]
    i = request.form["busenia"]
    l = request.form["sodium"]
    l = request.form["posium"]
    n = request.form["posium"]

                                                                                                                                                                                                                                                                                                  app.py >...
f = request.form["RBC"]
g = request.form["bacteria"]
h = request.form["bgr"]
i = request.form["bu"]
j = request.form["sodium"]
l = request.form["sodium"]
l = request.form["haemo"]
m = request.form["pcv"]
o = request.form["rbc-count"]
o = request.form["hypertension"]
p = request.form["pe"]
                            t = \texttt{[[float(a), float(b), float(c), float(d), float(e), float(f), float(g), float(h), float(i), float(j), float(
         @app.route('/admin')
def admin():
    return "Hey Admin How are you?"
       app.run(host='localhost', port=5000)
if __name __= ' __mate__
                               app.run(debug = True)
```

7.2 IBM DEPLOYMENT

Now after Locally deploying our machine Learning model we deployed our model into IBM deployment

```
OPEN EDITORS 1 unsaved
                                app_ibm.py
   chronic_kidney_...
    ■ Model_one.pkl F...
                               # style.css Flask\sta...
   pic.jpeq Flask\sta...
• home.html Flask...
                          header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
                               import pickle
root_path = os.path.dirname(__file__)
file_path = os.path.join(root_path, "Model_one.pkl")
 # style.css
v templates
                               @app.route('/static/<path:path>')
def send_report(path):
    return send_from_directory('static', path)
                               @app.route('/')
def helloworld():
    return render_template("home.html")
                                def login():
                                    a = request.form["age"]
b = request.form["bp"]
```

```
Chronic kidney ...
                                        app.py Flask
                                                                                                                                           def login():
                                                                                                                                              jef logIn():
    a = request.form["age"]
    b = request.form["bp"]
    c = request.form["sg"]
    d = request.form["alb"]

    home.html Flask...

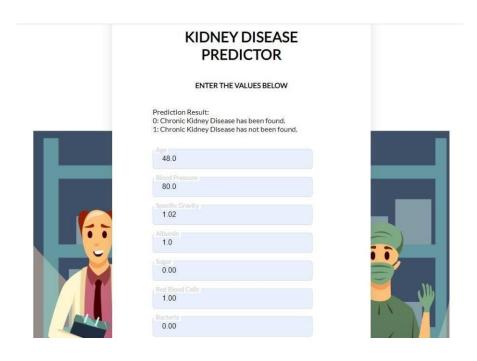
                                                                                                                                                       e = request.form["sugar"
f = request.form["RBC"]
                      ∨ ML
                                                                                                                                                   f = request.form["B8C"]
g = request.form["bacteria"]
h = request.form["bgr"]
i = request.form["bt"]
j = request.form["sc"]
k = request.form["scdium"]
l = request.form["pcv"]
m = request.form["pcv"]
n = request.form["pcv"]
 ö
                               # style.css
                                                                                                                                                     n = request.form["rbc-count"]
o = request.form["hypertension"]
                                                                                                                                                       p = request.form["pe"]

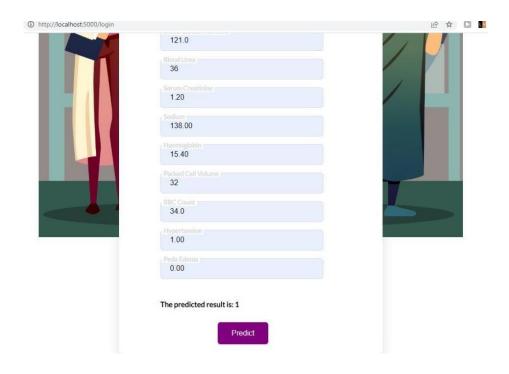
    app.py
    ■ Model_one.pkl

                                                                                                                                                      t = [[float(a), float(b), float(c), float(d), float(e), float(f), float(g), float(h), float(i), float(j), float(j)
                                                                                                                                                          payload_scoring =( {"input_data": [{"fields": ('f0','f1','f2','f3','f4','f5','f6','f7','f8','f9','f10','f11',
                                                                                                                                                        response_scoring = requests.post('https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/1fc2c8b6-0274-4b9f-85fa-9b
headers=('Authorization': 'Bearer ' + mltoken))
print("Scoring response")
print(response_scoring_json())
                                                                                                                                                           pred=response_scoring.json()
> OUTLINE
                    > TIMELINE
                                                                                                                                                          output=pred['predictions'][0]['values'][0][0]
                                                                                                                                                                                                                                                                                                                                                                  s: 4 UTF-8 CRLF () Python 3.11.0 64-bit @ Go Live 🔊 0m 🔘 Flow
```

8. TESTING

8.1 TESTCASES







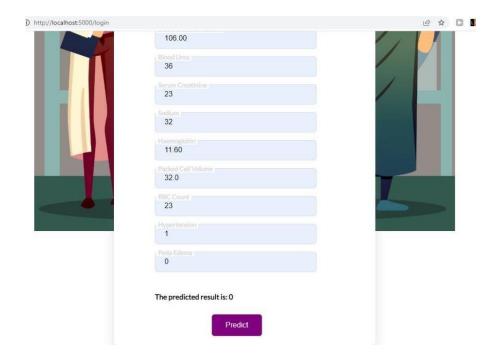
KIDNEY DISEASE **PREDICTOR**

ENTER THE VALUES BELOW

- Prediction Result:
 0: Chronic Kidney Disease has been found.
 1: Chronic Kidney Disease has not been found.

51		
ood Pressure		
60		
necific Gravity -		
1.00		
3.0		
4.0		
1.00		
0.00		





8.2 USER ACCECPTANCE TESTING

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [Early Detection of Chronic Kidney Disease] project at the time of the release to User Acceptance Testing (UAT).

2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	3	1	1	1	6
Duplicate	4	0	2	0	6
External	2	2	0	1	5
Fixed	1	1	1	1	4
Not Reproduced	0	0	0	0	0
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	10	4	4	3	21

3. Test Case Analysis

	t cases that have passed, failed, and untested			
Section	Total Cases	Not Tested	Fail	Pass
Home Screen	1	0	0	1
User Input	3	0	0	3
Chronic Kidney Disease testing	2	0	0	2
No Chronic Kidney Disease testing	2	0	0	2
Version Control	2	0	0	2

9. RESULTS

9.1 PERFORMANCE METRICS

```
[54] print("Accuracy of the model is: %3f " % accuracy_score(y_test,y_predic))

Accuracy of the model is: 0.9900000

[55] print(classification_report(y_test, y_predict))

precision recall f1-score support

0 1.00 0.96 0.98 57
1 0.96 1.00 0.98 43

accuracy 0.98 100
macro avg 0.98 0.98 0.98 100
weighted avg 0.98 0.98 0.98 100
```

10. ADVANTAGES & DISADVANTAGES

Advantages:

Chronic kidney disease (CKD) is one of the most critical health problems due to its increasing prevalence. It is also known as chronic renal disease which is a condition characterized by a gradual loss of kidney function over time. A better testing method which could possibly detect CKD in the early stages would be much more useful using machine learning algorithm

- Greater cost reduction in hospitals for testing
- Helps in early diagnosis of the disease
- Chances of recovery is higher

Disadvantages:

Even Though the CKD prediction model web application consists of a lot of advantages but it comes with certain disadvantages here are some of them .

- Chances of prediction to be wrong for least number of time which can cause problems
- Vast feature in dataset on discovery of time for the disease making the model inefficient to keep up the metrics
- Since it's a web application it requires scaling of web applications to handle concurrent requests after a certain threshold.

11. CONCLUSION

The benefit of this approach is that the prediction process takes far less time doctors to initiate treatment at the earliest for patients with CKD and further to classify larger populations of patients within a shorter span. Because the dataset used in this paper is tiny with 400 examples, we prefer to work with larger datasets in the future or compare the results of this dataset with a different dataset with the same. In addition, to help minimize the incidence of CKD, we try to predict if a person

with this syndrome chances chronic risk factors such as hypertension, family history of kidney failure and diabetes using the appropriate dataset. Early prediction is very crucial for both the experts and the patients to prevent and slow down the progress of chronic kidney disease to kidney failure.

12. FUTURE SCOPE

This work will be considered as the basement for the healthcare system for CKD patients. Also, extension to this work is that implementation of Machine learning provides high- quality performance. The hope is that it would encourage people to seek early treatment for chronic renal disease and to make improvements in their lives.

13. APPENDIX

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 CKD status prediction.

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-7593-1658893656

10.