# EARLY DETECTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

**IBM-Project-41116-1660639550**

## NALAIYA THIRAN PROJECT BASED LEARNING ON PROFESSIONAL READLINESS FOR INNOVATION, EMPLOYMENT AND ENTERPRENEURSHIP

**PROJECT REPORT**

**S.Rubashree (951219104039)**
**S.Sudarmathi (951219104050)**
**S.Sathya@Dhanalakshmi (951219104041)**
**C.Raju (951219104037)**

## BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

# CHAPTER 1

# INTRODUCTION

## 1.1 PROJECT OVERVIEW

Every year, an increasing number of patients are diagnosed with late stages of renal disease . Chronic Kidney Disease, also known as Chronic Renal Disease, is characterized by abnormal kidney function or a breakdown of renal function that progresses over months or years. Chronic kidney disease is often found during screening of persons who are known to be at risk for kidney issues, such as those with high blood pressure or diabetes, and those with a blood family who has chronic kidney disease (CKD). As a result, early prognosis is critical in battling the disease and providing effective therapy. Only early identification and continuous monitoring can avoid serious kidney damage or renal failure. Machine Learning (ML) plays a significant part in the healthcare system, and it may efficiently aid and help with decision support in medical institutions. The primary goals of this research are to design and suggest a machine learning method for predicting CKD. Support Vector Machine (SVR), Random Forest (LR), Artificial Neural Network (ANN), and Decision Tree are four master teaching methodologies investigated (DT). The components are built using chronic kidney disease datasets, and the outcomes of these models are compared to select the optimal model for prediction.

**Keywords**: Chronic Kidney Disease (CKD), Machine Learning (ML), Support Vector Machine (SVR), Random Forest (LR), Artificial Neural Network (ANN), Decision Tree (DT).

### 1.2 PURPOSE

Every year, an increasing number of patients are diagnosed with late stages of renal disease. Chronic Kidney Disease, also known as Chronic Kidney Disease, is characterized by abnormal kidney function or a breakdown of renal function that progresses over months or years. Chronic kidney disease is often found during screening of persons who are known to be at risk for kidney issues, such as those with high blood pressure or diabetes, and those with a blood family who has chronic kidney disease (CKD). As a result, early prognosis is critical in battling the disease and providing effective therapy. Only early identification and continuous monitoring can avoid serious kidney damage or kidney failure. Machine Learning plays a significant part in the healthcare system, and it may efficiently aid and help with decision support in medical institutions. The primary goals of this research are to design and suggest a machine learning method for predicting CKD. Support Vector Machine (SVR), Random Forest (LR), Artificial Neural Network (ANN), and Decision Tree are four master teaching methodologies investigated (DT). The components are built using chronic kidney disease datasets.

# CHAPTER 2

# LITERATURE SURVEY

### 2.1 EXISTING PROBLEM

Non communicable illnesses are the leading cause of early death, and CKD is the leading non communicable disease. Chronic Kidney Disease is a major concern for the global health care system. People with CKD must focus on implementing proven, cost-effective therapies to as many people as possible while taking into consideration restricted needs, human and financial resources. Chronic kidney disease (CKD) is now wreaking havoc on society and is spreading at an alarming rate. Various efforts have been undertaken to advance early therapy to prevent the condition from progressing to chronic disease. Recent research suggests that some of the negative outcomes can be avoided with early identification and treatment.

**https://www.kidney.org/phi/form?version=health**

### 2.2 REFERENCES

1. **Title:** Prediction of Chronic Kidney disease using adaptive hybridized
   deep convolutional neural network on the internet of medical things platform
   **Source:** IEEE access

**Author:** G. chenetal
**Date:** October 2020

2. **Title:** Clinical practice guideline for diabetes management in chronic kidney disease
**Source:** Kidney Int
**Author:** P.T. Coatesetal
**Date:** October 2020

3.**Title:** Overview of clinical prediction models
**Source:** Ann Transt.Med
**Author:** L.Chen
**Date:** November 2019

4. **Title:** Prediction of chronic kidney diseases using deep artificial neural network technique
**Source:** Springer International Publishing
**Author:** H. kriplani
**Date:**June 2019

5.**Title:** Three types of Machine Learning Algorithms List of Common Machine Learning Algorithms
**Source:**IEEE Access
**Author:** Abdi
**Date:** November 2016

6. **Title:** Risk Prediction of Chronic Kidney Disease Using Machine Learning Algorithms
**Source:** ICCCNT
**Author:** N. A. Almansourteal
**Date:** October 2018

7. **Title:** Computer-Aided Diagnosis of Chronic Kidney Disease in Developing Countries
**Source:** IEEE Access
**Author:** M. Eliete Pinheiro
**Date:** October 2020

8. **Title:** Neural network and support vector machine for the prediction of chronic kidney disease
**Source:** Computer Bio Medical
**Author:** N. A. Almansoureta l
**Date:** October 2018

**9. Title:** Prediction of chronic kidney disease (CKD) using Data Science
**Source:** Conf. Intelligent Computer Control System
**Author:** N. V. Ganapathi Raju
**Date:** November 2019

**10. Title:** Prediction of chronic kidney disease using machine learning
**Source:** Advanced Science Technology
**Author:** M. V. Maheshwar redy
**Date:** December 2019

## 2.3 PROBLEM STATEMENT DEFINITION

Every year, an increasing number of patients are diagnosed with late stages of renal disease . Chronic Kidney Disease, also known as Chronic Renal Disease, is characterized by abnormal kidney function or a breakdown of renal function that progresses over months or years. Chronic kidney disease is often found during screening of persons who are known to be at risk for kidney issues, such as those with high blood pressure or diabetes, and those with a blood family who has chronic kidney disease (CKD). As a result, early prognosis is critical in battling the disease and providing effective therapy. Only early identification and continuous monitoring can avoid serious kidney damage or renal failure.

| | |
|---|---|
| Who does the problem affect? | Those who have a history of kidney failure in their families, diabetes, high blood pressure, or heart disease . |
| What are the boundaries of the problem? | Diagnosing kidney diseases using parameters like blood pressure and albumin . |
| What is the issue? | Kidney function is compromised by a disease or condition, leading to its damage over time |
| When does the issue | When a sickness or illness compromises kidney |

|  |  |
|---|---|
| occur? | function, causing kidney damage to worsen over several months or years |
| Where is the issue occurring? | The small blood veins in the kidneys might become strained by high blood pressure, thereby preventing normal functioning of the kidney.<br>Blood glucose levels that are too high can harm the kidneys' small filters. |
| Why is it important that we fix the problem | The progression of chronic kidney disease to an advanced state may be slowed or stopped with early detection and treatment |

# CHAPTER 3

# IDEATION & PROPOSED SOLUTION

## 3.1 EMPATHY MAP CANVAS

### THINK AND FEEL

- Tell physical Feeling to Doctor.
- They want to relief from chronic kidney disease.
- They feel tierd and weak of our body.

### HEAR

- To Visit Doctor.
- To Avoid Smoke and Drinks.
- To Give their Own Personal Suggestion.

### SAY AND DO

- Muscle Cramps.
- High Blood Pressure.

- Urinating Less Or More.
- Loss Of Appetite.

### SEE

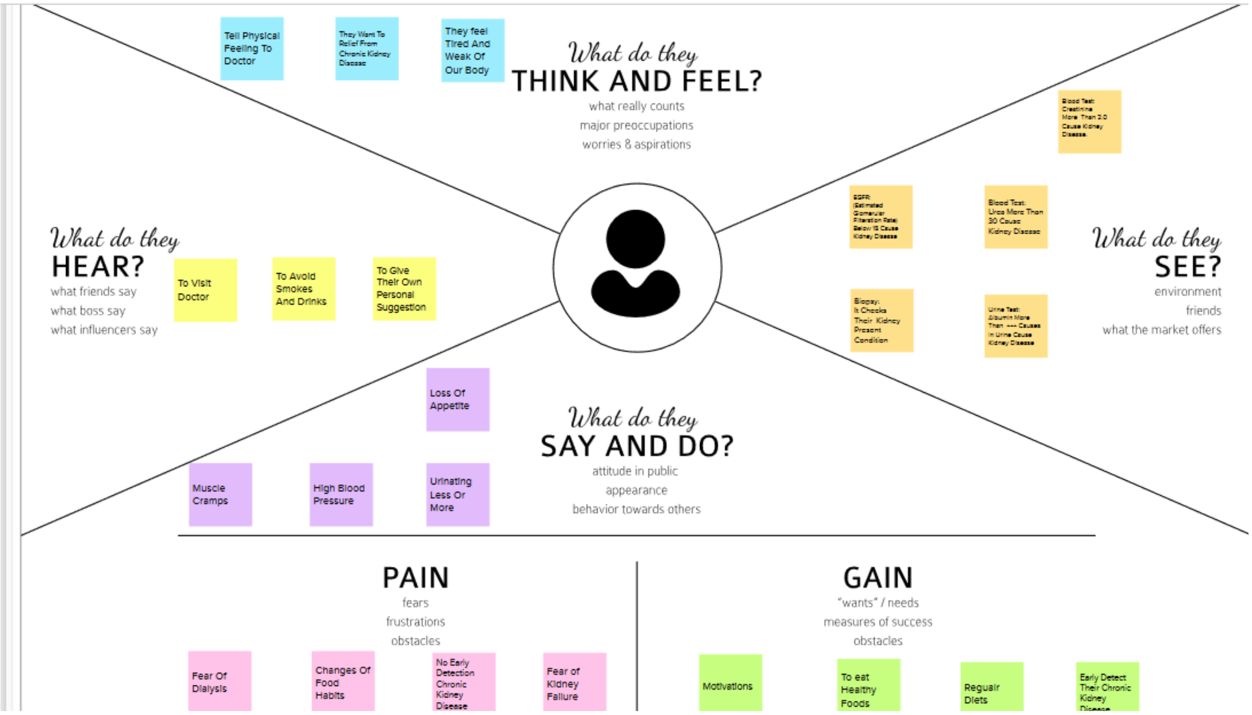
- Blood Test Creatinine More Than 2.0 Cause Kidney Disease.
- EGFR:(Estimated Glomerular Filteration Rate) Below 15 Cause Kidney Disease.
- Blood Test :Urea Morethan 30 Cause Kidney Disease.
- Biopsy: It Checks their Kidney Present Condition.
- UrineTest: Albumin More than +++ Causes in Urine Cause Kidney Disease

### GAIN

- Motivations.
- Eat healthy foods.
- Regular diets.
- Early detect their chronic kidney disease.

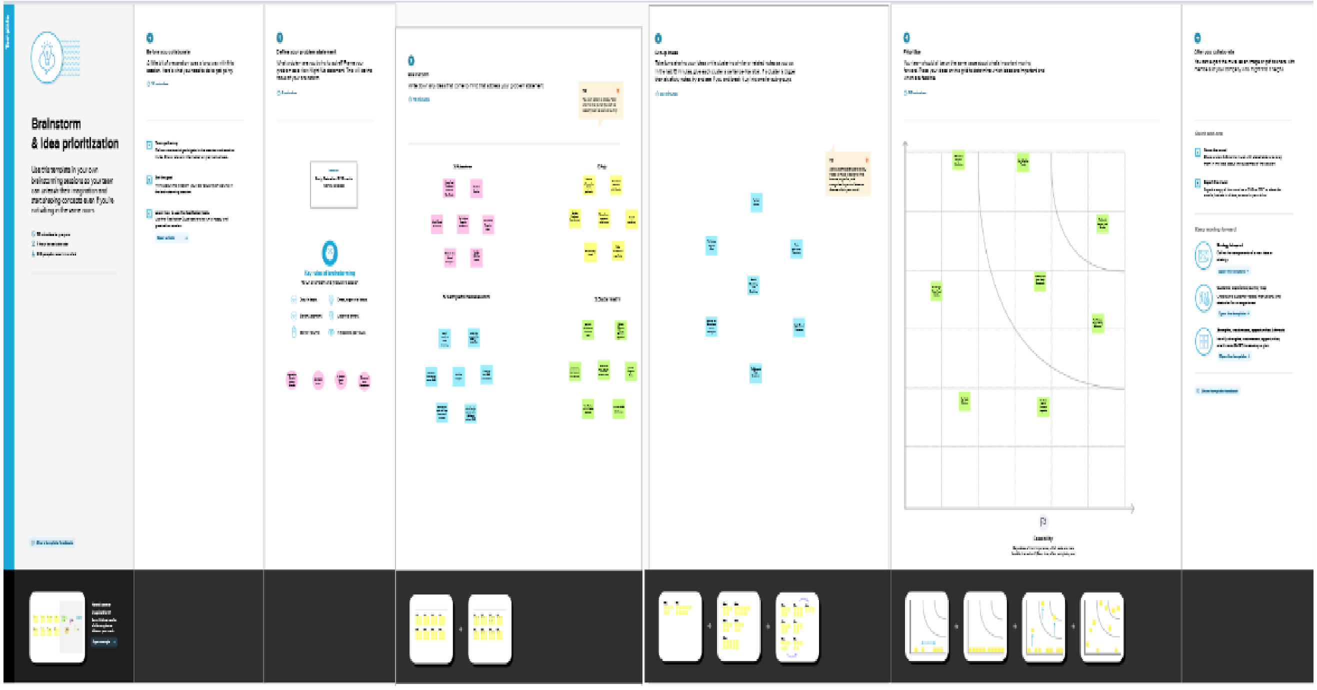
### PAIN

- Fear of dialysis.
- Change of food habits.
- Fear of kidney failure.
- No early detection of chronic kidney disease.


Tell Physical Feeling To Doctor
They Want To Relief From Chronic Kidney Disease
They feel Tired And Weak Of Our Body
What do they
THINK AND FEEL?
what really counts
major preoccupations
worries & aspirations
Blood Test: Creatinine More Than 2.0 Cause Kidney Disease.
EGFR: (Estimated Glomerular Filteration Rate) Below 15 Cause Kidney Disease
Blood Test: Urea More Than 30 Cause Kidney Disease
Biopsy: It Checks Their Kidney Present Condition
Urine Test: Albumin More Than +++ Causes In Urine Cause Kidney Disease
What do they
SEE?
environment
friends
what the market offers
What do they
HEAR?
what friends say
what boss say
what influencers say
To Visit Doctor
To Avoid Smokes And Drinks
To Give Their Own Personal Suggestion
Loss Of Appetite
Muscle Cramps
High Blood Pressure
Urinating Less Or More
What do they
SAY AND DO?
attitude in public
appearance
behavior towards others
PAIN
fears
frustrations
obstacles
Fear Of Dialysis
Changes Of Food Habits
No Early Detection Chronic Kidney Disease
Fear of Kidney Failure
GAIN
"wants" / needs
measures of success
obstacles
Motivations
To eat Healthy Foods
Regualr Diets
Early Detect Their Chronic Kidney Disease

## 3.2 IDEATION & BRAINSTORMING

- They feel tierd and weak .
- Collect information from the user.
- Visit doctor who has Chronic kidney disease.
- Test tha samples.
- To change their food habits.
- Eat healthy foods.



Brainstorm
& Idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

## 3.3 PROPOSED SOLUTION

| S.NO | Parameter | Description |
|---|---|---|
| 1 | Problem Statement(Problem to be solved) | Now a days, many people affected by kidneydisease it will be detected in severe stage cause death. Our project detects the kidney |

| | | |
|---|---|---|
| | | disease every accurate using web application. |
| 2 | Ideas/solution description | To detect the early stage of kidney disease in the form of web application. |
| 3 | Novelty/Uniqueness | User ca detect their disease in early stage. |
| 4 | Social Impact/Customer Satsifaction | 1. To reduce the death rate.<br>2. To avoiding progression of kidney failure.<br>3. To give awareness about kidney disease. |
| 5 | Business Model(Revenue Model) | 1. Can generate revenue through direct<br>customers<br>2. Can collaborate with health care sector<br>and generate revenue from their<br>customers |
| 6 | Scalability of the solution | Our web application has a high security database. |

## 3.4 PROBLEM SOLUTION FIT

### CUSTOMER SEGMENTS

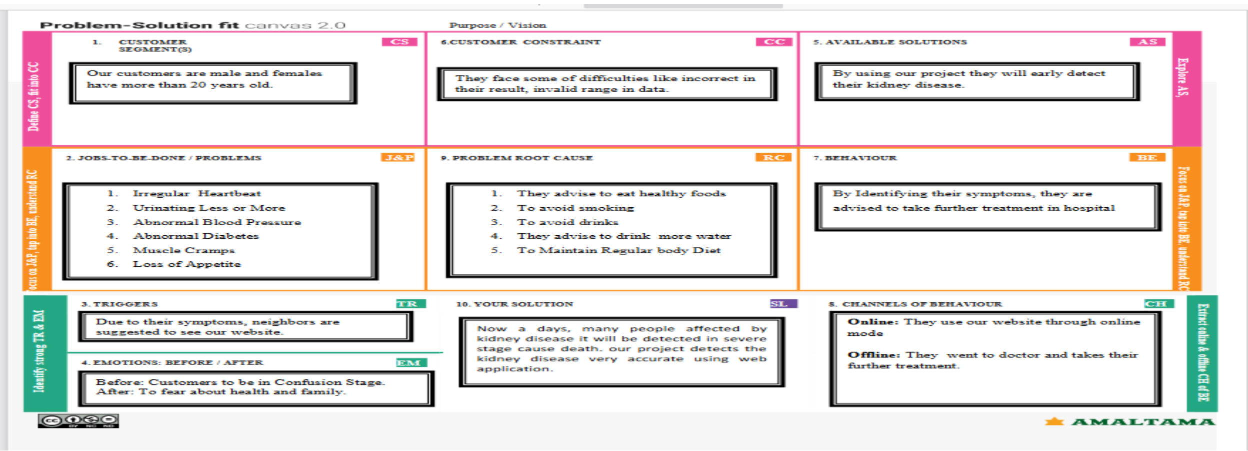
- Our customer is male or female more than 20 years old.

### PROBLEM ROOT CAUSE

- Advised to eat healthy foods
- To avoid smoking.
- To avoid drinks.
- Advised to drink more water.
- Follow regular diets.

## YOUR SOLUTION

- Now a days, many people affected by kidney disease it will detected in severe stage cause death. our project detects the disease accurately using web application.


Define CS, fit into CC
Focus on J&P, tap into BE, understand RC
Identify strong TR & EM
Explore AS,
Focus on J&P, tap into BE, understand RC
Extract online & offline CH of BE
CS
CC
AS
J&P
RC
BE
TR
EM
SL
CH
AMALTAMA

Problem-Solution fit canvas 2.0

Purpose / Vision

1. CUSTOMER SEGMENT(S)

Our customers are male and females have more than 20 years old.

6.CUSTOMER CONSTRAINT

They face some of difficulties like incorrect in their result, invalid range in data.

5. AVAILABLE SOLUTIONS

By using our project they will early detect their kidney disease.

2. JOBS-TO-BE-DONE / PROBLEMS

1. Irregular Heartbeat
2. Urinating Less or More
3. Abnormal Blood Pressure
4. Abnormal Diabetes
5. Muscle Cramps
6. Loss of Appetite

9. PROBLEM ROOT CAUSE

1. They advise to eat healthy foods
2. To avoid smoking
3. To avoid drinks
4. They advise to drink more water
5. To Maintain Regular body Diet

7. BEHAVIOUR

By Identifying their symptoms, they are advised to take further treatment in hospital

3. TRIGGERS

Due to their symptoms, neighbors are suggested to see our website.

4. EMOTIONS: BEFORE / AFTER

Before: Customers to be in Confusion Stage.
After: To fear about health and family.

10. YOUR SOLUTION

Now a days, many people affected by kidney disease it will be detected in severe stage cause death. our project detects the kidney disease very accurate using web application.

8. CHANNELS OF BEHAVIOUR

**Online:** They use our website through online mode

**Offline:** They went to doctor and takes their further treatment.

# CHAPTER 4

# REQUIREMENT ANALYSIS

## 4.1 FUNCTIONAL REQUIREMENTS

**Following are the Functional Requirements of proposed solution.**

| FR NO | Functional Requirements (Epic) | Sub requirements(Story/subtask) |
|---|---|---|
| FR-1 | User Registration | 1. Registration through Form<br>2. Registration through Gmail<br>3. Registration through LinkedIn. |
| FR-2 | User Confirmation | 1. Confirmation via Email<br>2. Confirmation via OTP |
| FR-3 | Data Entry | 1. Collect The data from the user<br>2. Fill the Form |
| FR-4 | Analysing the Data | 1. Analyse the data<br>2. Getting their results |
| FR-5 | Machine Learning Algorithms | 1. Decision Tree<br>2. Rainforest<br>3. Support Vector Machine<br>4. Artificial Neural Network |
| FR-6 | Phases | 1. Testing Phase<br>2. Training Phase |

## 4.2 NON-FUNCTIONAL REQUIREMENTS

**Following are the Non Functional Requirements of the Proposed Solutions.**

| NRF NO | Non-Functional Requiremnts | Description |
|---|---|---|
| NFR-1 | Usability | 1. Preventing loss of kidney disease.<br>2. Delaying oravoiding progression to kidney failure. |

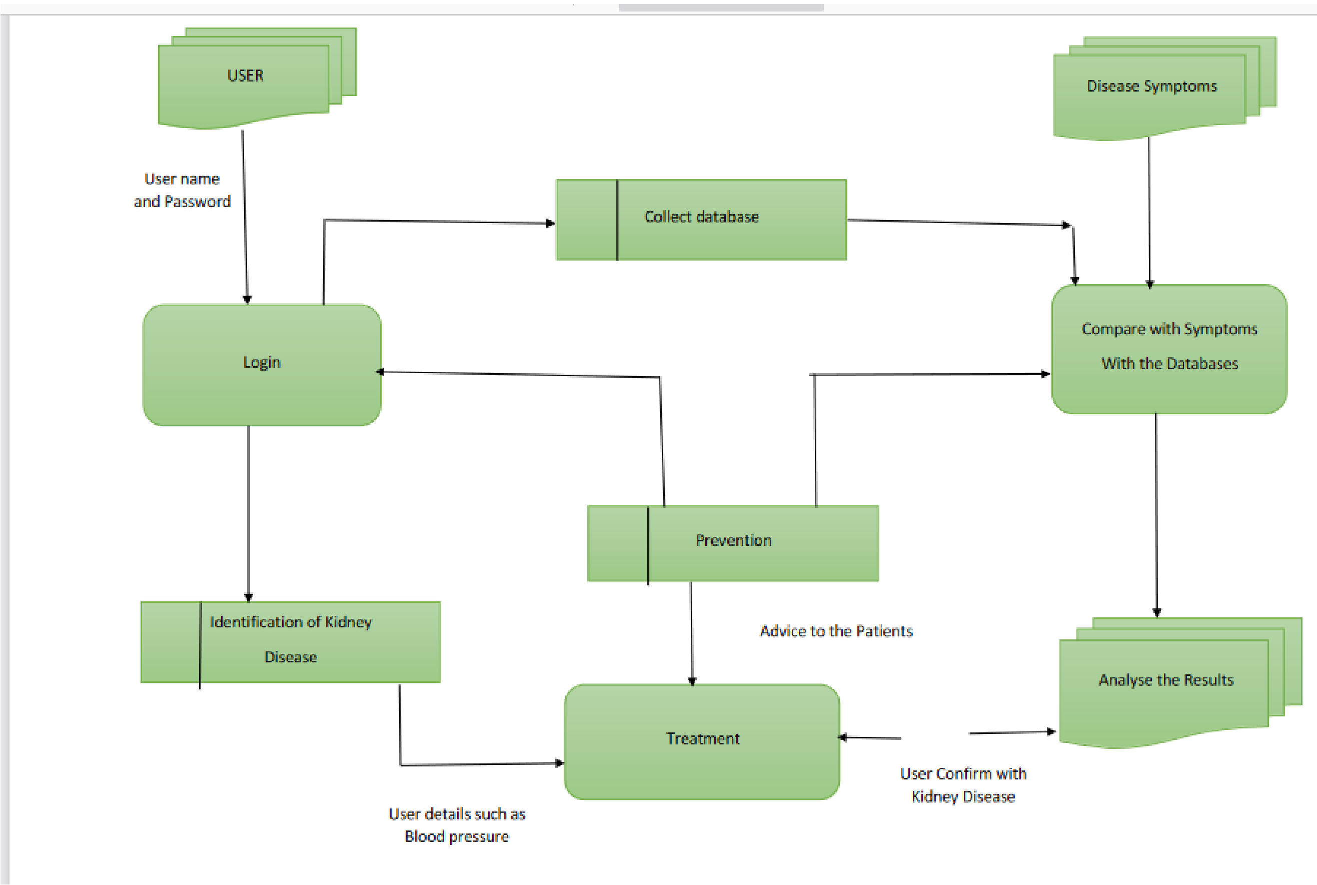
| | | |
|---|---|---|
| NFR-2 | Security | 1. Encrypt your data.<br>2. Focus the hosting service measure<br>3. Avoid security misconfigurations. |
| NFR-3 | Reliability | Result should be 99.99% accurate. |
| NFR-4 | Performance | Compare the data with symptoms to give results. |
| NFR-5 | Availability | It access at any time. |
| NFR-6 | Scalability | 1. Memory utilization.<br>2. CPU usage.<br>3. Networkinput/output.<br>4. Disk input/output. |

# CHAPTER 5

# PROJECT DESIGN

## 5.1 DATA FLOW DIAGRAMS

A Dataflow Diagrams(DFD) is a traditional visual representation of the information flows within a system .A neat and clear DFDcan depict the right amount of the system requirements graphically . it shows how data enters and leave the system, what changesthe information, and where data is stored.

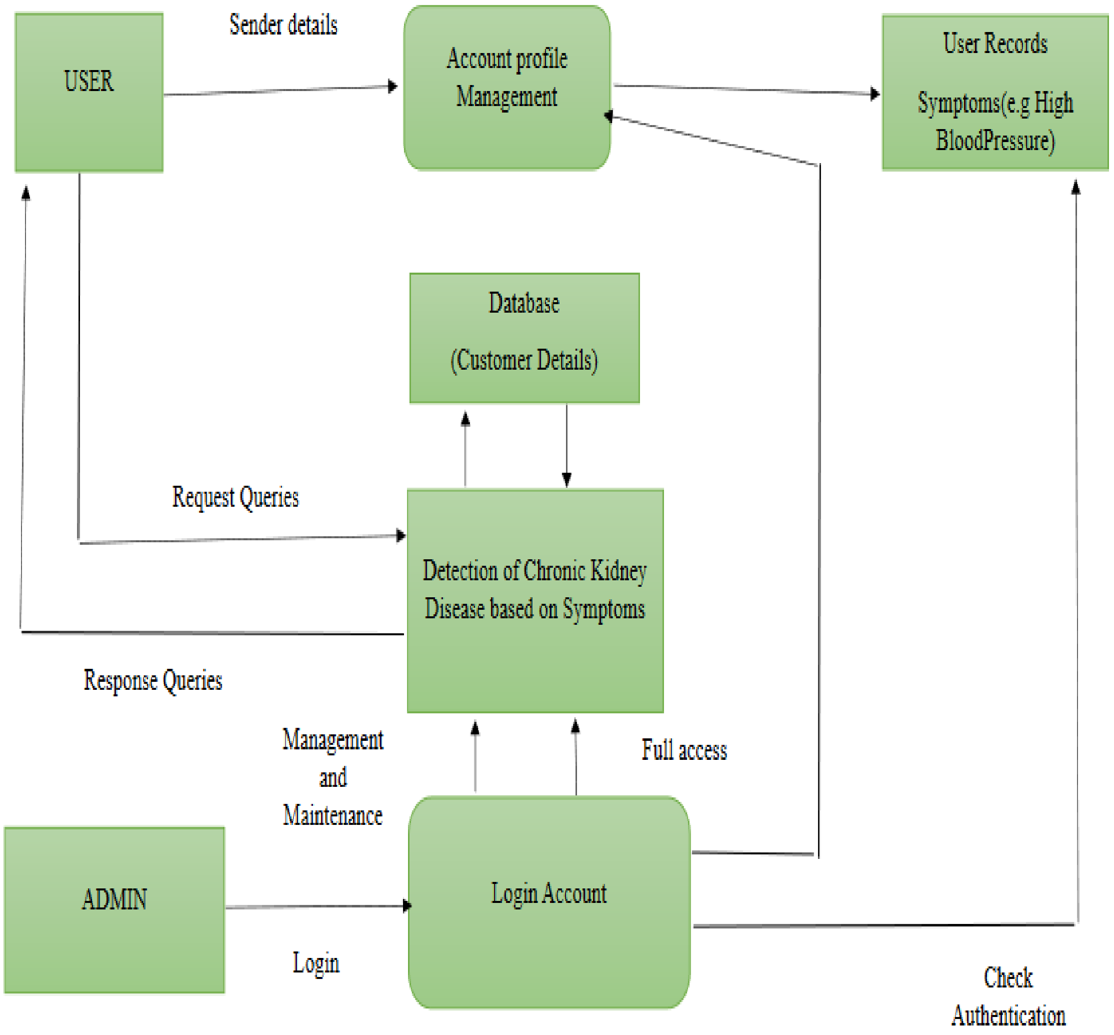

USER
Disease Symptoms
User name and Password
Collect database
Login
Compare with Symptoms With the Databases
Prevention
Identification of Kidney Disease
Advice to the Patients
Analyse the Results
Treatment
User Confirm with Kidney Disease
User details such as Blood pressure

## USER STORIES

| User Type | Functional Requirements | User Story Number | User story/Task | Acceptance criteria | priority | Release |
|---|---|---|---|---|---|---|
| Customer | Registration | USN-1 | As a user, I can register for the website by mail id, password. | access account | High | Sprint 1 |
| | Validation | USN-2 | As a user will receive a conformation mail. | Receive mail | High | Sprint 2 |
| | Login | USN-3 | User will login to the website by email id, password. | Create and login | High | Sprint 1 |
| | Collect database | USN-4 | User have to answer for 10 -20 questions. | Choose answers | High | Sprint 1 |
| | Output to user | USN-5 | User can view the result. | View results | High | Sprint 2 |
| | Remedies | USN-6 | Suggest some advices to the users | Accept the remedies | Medium | Sprint 2 |
| Administrator | Prediction | USN-7 | Admin will build machine learning<br>models to predict chronic kidney disease. | Deploy in the websites | High | Sprint 1 |

## 5.2 SOLUTION & TECHNICAL ARCHITECTURE

## SOLUTION ARCHITECTURE



## TECHNICAL ARCHITECTURE

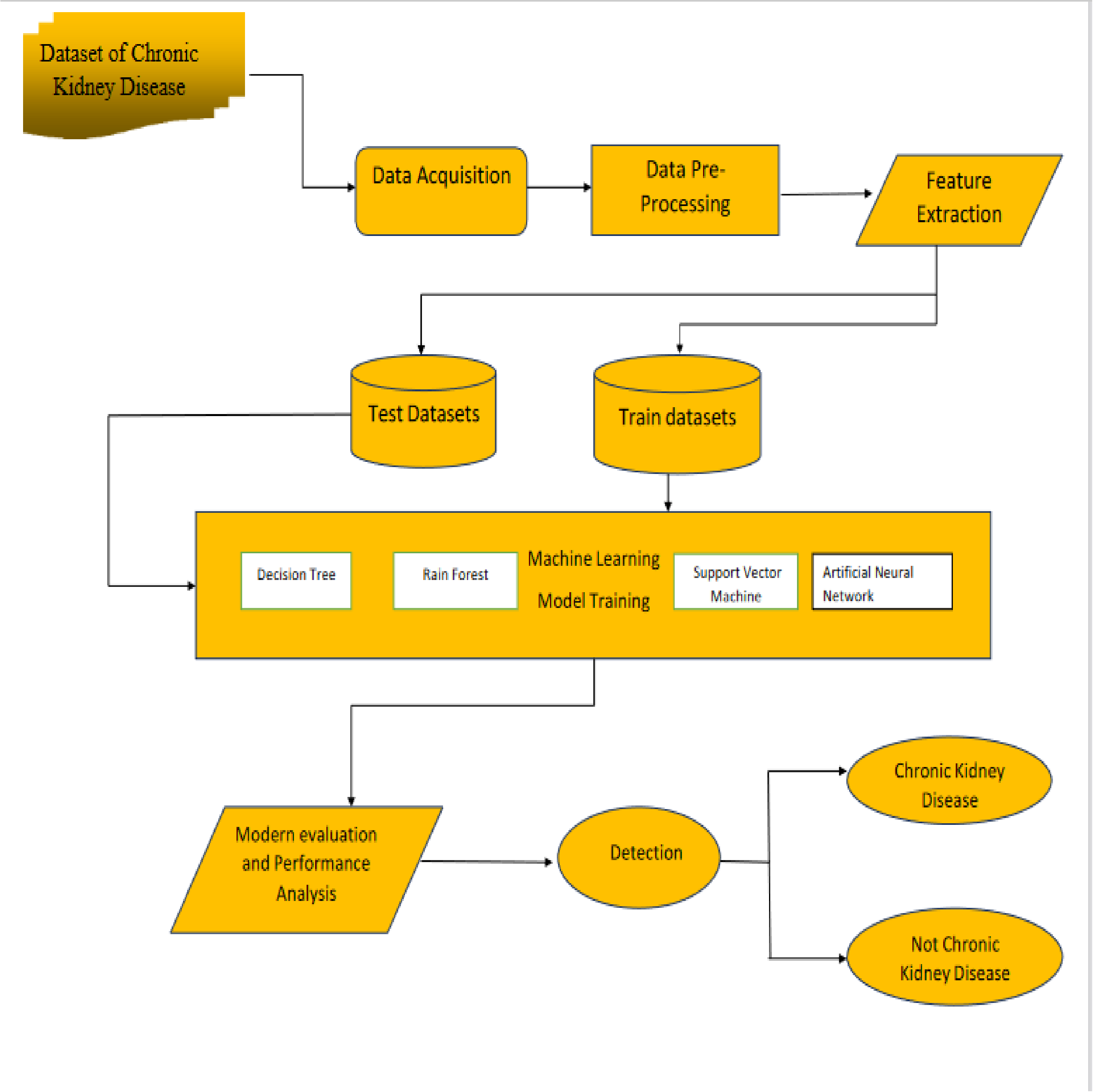


**TABLE 1: COMPONENTS & TECHNOLOGIES:**

| SI.NO | COMPONENTS | DESCRIPTION | TECHNOLOGY |
|---|---|---|---|
| 1 | User Registration | 1. Registration through Form<br>2. Registration through Gmail<br>3. Registration through LinkedIn. | HTML,<br>CSS and<br>Python flask. |
| 2 | User Confirmation | 1. Confirmation via Email<br>2. Confirmation via OTP | HTML,<br>CSS and<br>Python flask. |
| 3 | Data Entry | 1. Collect The data from the user<br>2. Fill the Form | HTML, CSS,<br>My Sql, Python flask<br>Backend Python. |
| 4 | Analysing the Data | 1. Analyse the data<br>2. Getting their results | Machine Learning Models. |
| 5 | Machine Learning Algorithms | 1. Decision Tree<br>2. Rainforest<br>3. Support Vector Machine<br>4. Artificial Neural Network | Based on Machine Learning Concepts. |
| 6 | Phases | 1. Testing Phase<br>2. Training Phase | Machine Learning and Deep learning concepts. |

**TABLE 2: APPLICATION & CHARACTERISTICS**

| SI NO | CHARACTERISTICS | DESCRIPTION | TECHNOLOGY |
|---|---|---|---|
| 1 | Usability | 1. Preventing loss of kidney disease.<br>2. Delaying or avoiding progression to kidney failure. | Cloud(Open Source Platform). |
| 2 | Security | 1. Encrypt your data.<br>2. Focus the hosting service measure<br>3. Avoid security misconfigurations. | Encryption and Authentication |
| 3 | Reliability | Result should be 99.99%accurate | Web Development. |

|  |  |  |  |
|---|---|---|---|
| 4 | Performance | Compare the data with symptoms to give results. | Machine Learning and Deep Learning Neural Networks. |
| 5 | Availability | It access at any time. | Machine Learning. |
| 6 | Scalability | 1. Memory utilization.<br>2. CPU usage.<br>3. Network input/output.<br>4. Disk input/output. | Performance Optimization. |

## 5.3 USER STORIES

| **User Type** | **Functional Requirements** | **User Story Number** | **User story/Task** | **Acceptance criteria** | **priority** | **Release** |
|---|---|---|---|---|---|---|
| Customer | Registration | USN-1 | As a user, I can register for the website by mail id, password. | access account | High | Sprint 1 |
|  | Validation | USN-2 | As a user will receive a conformation mail. | Receive mail | High | Sprint 2 |
|  | Login | USN-3 | User will login to the website by email id, password. | Create and login | High | Sprint 1 |
|  | Collect database | USN-4 | User have to answer for 10 -20 questions. | Choose answers | High | Sprint 1 |
|  | Output to user | USN-5 | User can view the result. | View results | High | Sprint 2 |
|  | Remedies | USN-6 | Suggest some advices to the users | Accept the remedies | Medium | Sprint 2 |
| Administrator | Prediction | USN-7 | Admin will build machine learning models to predict chronic kidney disease. | Deploy in the websites | High | Sprint 1 |

# CHAPTER 6

# PROJECT PLANNING & SCHEDULING

## 6.1 SPRINT PLANNING & ESTIMATION

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-1 | User Registration | USN-1 | As a user, I can register for the application by entering my name, mobile number, email, password, and confirming my password. | 10 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |
| Sprint-2 | | USN-2 | As a user, I can register for the application through Gmail | 5 | Medium | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |
| Sprint-1 | User Confirmation | USN-3 | As a user, I will receive confirmation email once I have registered for the application | 10 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |
| Sprint-2 | | USN-4 | As a user, I will receive confirmation otp to verify the identity. | 5 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |
| Sprint-2 | Data Collection | USN-5 | As a user, I will enter the input data for disease prediction in the form | 10 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |
| Sprint-3 | Provide output to the user | USN-6 | As a user, I will get the result of disease prediction in the dashboard. | 10 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |

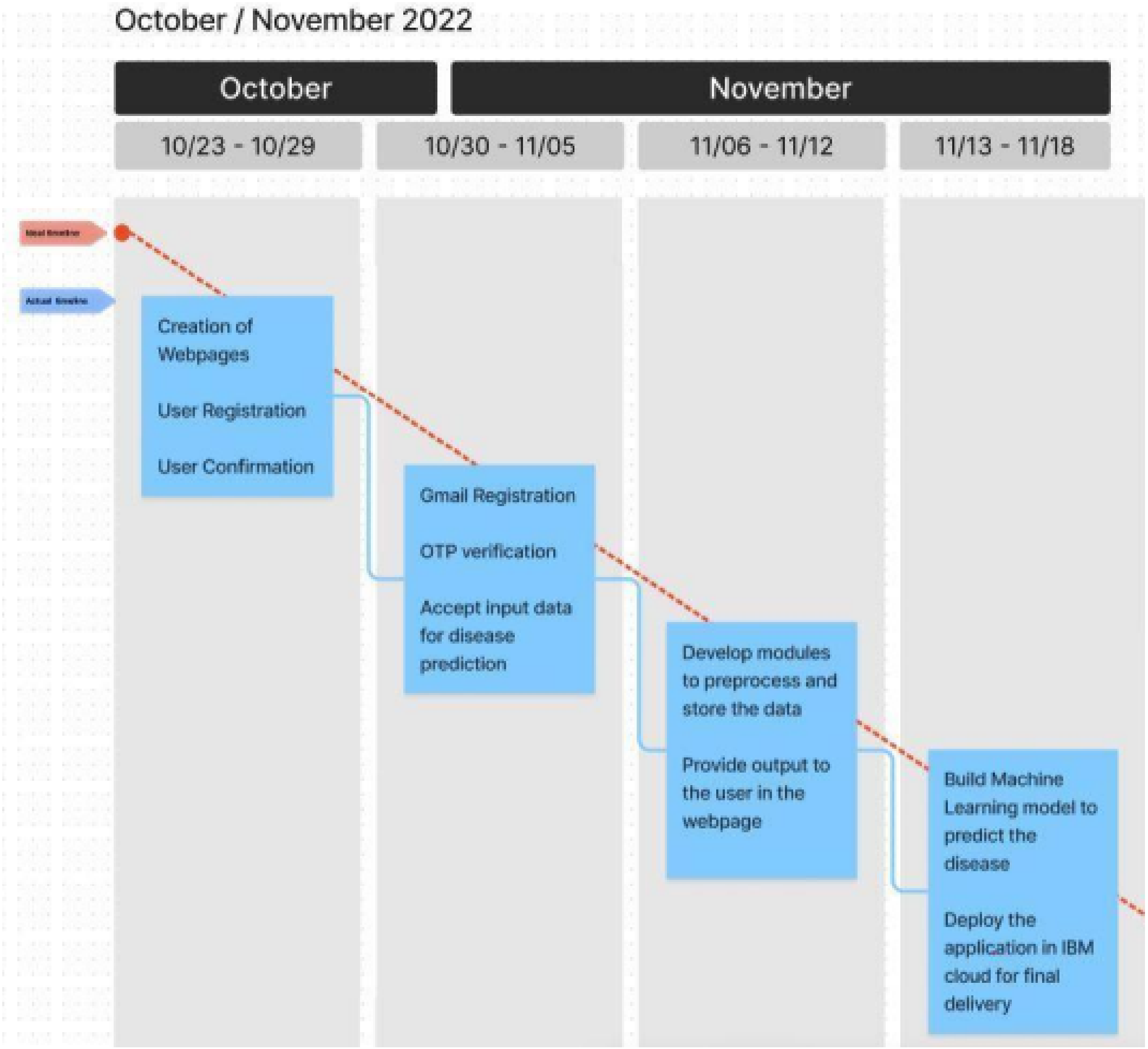
| | | | | | | |
|---|---|---|---|---|---|---|
| Sprint-3 | Data Analysis | USN-7 | As the admin, I will develop modules to preprocess and store the data. | 10 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |
| Sprint-4 | Prediction of disease | USN-8 | As the admin, I will build a Machine Learning model to predict the disease | 10 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |
| Sprint-4 | Final Delivery | USN-9 | Deploy the application in IBM cloud and make it available for use. | 10 | High | S.Rubashree<br>S.Sudarmathi<br>S.Sathya<br>C.Raju |

### 6.2 SPRINT DELIVARY SCHEDULE

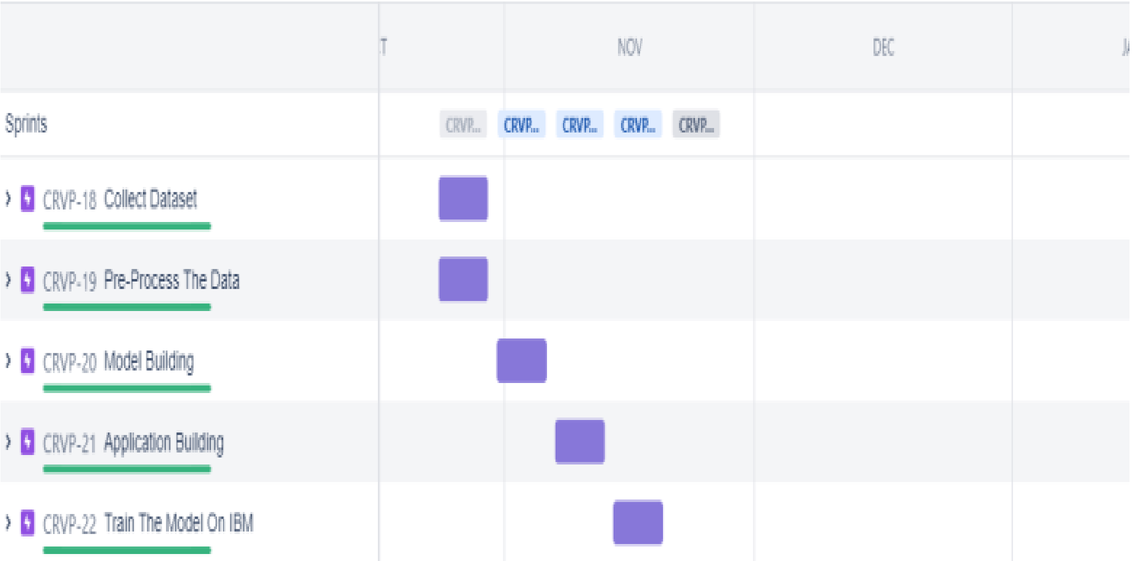
| **Sprint** | **Total Story Points** | **Duration** | **Sprint Start Date** | **Sprint End Date (Planned)** | **Story Points Completed (as on Planned End Date)** | **Sprint Release Date (Actual)** |
|---|---|---|---|---|---|---|
| Sprint -1 | 20 | 6 Day | 24 Oct 2022 | 29 Oct 2022 | 20 | 29 Oct 2022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

**Velocity:**
We have a 6-day sprint duration, and the velocity of the team is 20 (points per sprint). Theteam's average velocity (AV) per iteration unit (story points per day)
**AV = Sprint duration / velocity = 20 / 6 = 3.33**

## Burndown Chart



### 6.3 REPORTS FROM JIRA



Sprints

CRVP-18 Collect Dataset

CRVP-19 Pre-Process The Data

CRVP-20 Model Building

CRVP-21 Application Building

CRVP-22 Train The Model On IBM

## CHAPTER -7

## CODING & SOLUTIONING

### 7.1 FEATURE 1

### index.html

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
        <title>Flaskimio</title>
<style>
body {
background: linear-gradient(
           rgba(20,20,20, .75),
           rgba(20,20,20, .75)),
           url(
'https://pharmanewsintel.com/images/site/article_headers/_normal/Chronic_Disease_Manage.
png');
.container {
  border: 2px solid #ccc;
  padding: 10px;
  width: 20em;
height:21em;
background-color:white;
}


.hello{
opacity: 0.5;
}
</style>
</head>

<body>
<marquee bgcolor="white">IBM NALAIYA THIRAN</marquee>
<center><p style="font-size:60px;color:white;">Chronic Disease Prediction</p></center>
<form action="/val" method="post"><center>
  <label for="age">Age:</label><br>
  <input type="number" id="age" name="age"><br><br><br>


  <label for="bp">Blood Pressure:</label><br>
  <input type="number" id="bp" name="bp">
<br>
<br>
<br>
  <label for="sg">Urinary Specific Gravity:</label><br>
  <input type="number" id="sg" name="sg">
<br>
<br>
<br>
  <label for="al">al:</label><br>
  <input type="number" id="al" name="al">
<br>
```

```
<br>
<br>
  <label for="su">su:</label><br>
  <input type="number" id="su" name="su">
<br>
<br>
<br>
  <label for="rbc">rbc:</label><br>
  <input type="text" id="rbc" name="rbc">
<br>
<br>
<br>
  <label for="pc">pc:</label><br>
  <input type="text" id="pc" name="pc">
<br>
<br>
<br>
  <label for="pcc">pcc:</label><br>
  <input type="text" id="pcc" name="pcc">
<br>
<br>
<br>
  <label for="ba">ba:</label><br>
  <input type="text" id="ba" name="ba">
<br>
<br>
<br>
  <label for="bgr">bgr:</label><br>
  <input type="number" id="bgr" name="bgr">
<br>
<br>
<br>
  <label for="bu">bu:</label><br>
  <input type="number" id="bu" name="bu">
<br>
<br>
<br>
  <label for="sc">sc:</label><br>
  <input type="number" id="sc" name="sc">
<br>
```

```
<br><br>
  <label for="sod">sod:</label><br>
  <input type="number" id="sod" name="sod">
<br>
<br>
<br>
  <label for="pot">pot:</label><br>
  <input type="number" id="pot" name="pot">
<br>
<br>
<br>
  <label for="hemo">hemo:</label><br>
  <input type="number" id="hemo" name="hemo">
<br>
<br>
<br>
  <label for="pcv">pcv:</label><br>
  <input type="text" id="pcv" name="pcv">
<br>
<br>
<br>
  <label for="wc">wc:</label><br>
  <input type="text" id="wc" name="wc">
<br>
<br>
<br>
  <label for="rc">rc:</label><br>
  <input type="text" id="rc" name="rc">
<br>
<br>
<br>
  <label for="htn">htn:</label><br>
  <input type="text" id="htn" name="htn">
<br>
<br>
<br>
  <label for="dm">dm:</label><br>
  <input type="text" id="dm" name="dm">
<br>
<br>
<br>
```

```
  <label for="cad">cad:</label><br>
  <input type="text" id="cad" name="cad">
<br>
<br>
<br>
  <label for="appet">appet:</label><br>
  <input type="text" id="appet" name="appet">
<br>
<br>
<br>
  <label for="pe">pe:</label><br>
  <input type="text" id="pe" name="pe">
<br>
<br>
<br>
  <label for="ane">ane:</label><br>
  <input type="text" id="ane" name="ane">
<br>
<br></center>
  <center><button type="submit">Check</button></center>
</form>
</body>
</html>
```

**rename.html**

```
<html>
<head>
<style>
body { background-color: #E6E6FA;
}
</style>
</head>
<body >
<br>
<br>
<br>
<center><h1>{{answer1}}</h1></center>
```

```
<br>
<center><h1>{{answer2}}</h1></center>
<center><img alt="Qries" src="https://image.shutterstock.com/image-vector/cute-illustration-
sick-kidney-characters-260nw-1529381825.jpg width=500" height="500"></center>

</body>
</html>
```

**rename2.html**

```
<html>
<head>
<style>
body {
  background-color: #E6E6FA;
}
</style>
</head>
<body >
<br>
<br>
<br>
<center><h1>{{answer1}}</h1></center>
<br>
<center><h1>{{answer2}}</h1></center>
<center><img alt="Qries" src="https://media.istockphoto.com/vectors/vector-kidney-cartoon-
human-body-health-organ-smiling-mascot-on-vector-
id1136533246?k=20&m=1136533246&s=612x612&w=0&h=eWjhtoVOfX3lBF4fcsDD4ZLLzjeJ_A
```

```
x4cgdgdUexG1Q="
        width=500" height="500"></center>

</body>
</html>
```

### 7.2 FEATURE 2

```
import pickle
loaded_class = pickle. load(open('randomclass_chronic', 'rb'))
loaded_reg = pickle. load(open('randomreg_chronic', 'rb'))
import numpy as np
import pandas as pd
from flask import Flask, request, redirect, render_template
app = Flask(__name__)
@app.route("/",methods=['GET', 'POST'])
def index():
    return render_template('index.html')
@app.route("/val",methods=['POST'])

def val():
    test=[]
    if request.method == 'POST':
        test.append(request.form.get("age"))
        test.append(request.form.get("bp"))
        test.append(request.form.get("sg"))
        test.append(request.form.get("al"))
        test.append(request.form.get("su"))
        rb=request.form.get("rbc")
        if rb=='abnormal':
            test.append(1)
        else:
            test.append(0)
        pc=request.form.get("pc")
        if pc=='abnormal':
            test.append(1)
        else:
            test.append(0)
        pcc=request.form.get("pcc")
```

```
if pcc=='present':
    test.append(1)
else:
    test.append(0)
ba=request.form.get("ba")
if ba=='present':
    test.append(1)
else:
    test.append(0)
test.append(request.form.get("bgr"))
test.append(request.form.get("bu"))
test.append(request.form.get("sc"))
test.append(request.form.get("sod"))
test.append(request.form.get("pot"))
test.append(request.form.get("hemo"))
test.append(request.form.get("pcv"))
test.append(request.form.get("wc"))
test.append(request.form.get("rc"))
ht=request.form.get("htn")
if ht=='yes':
    test.append(1)
else:
    test.append(0)
d=request.form.get("dm")
if d=='yes':
    test.append(1)
else:
    test.append(0)
ca=request.form.get("cad")
if ca=='yes':
    test.append(1)
else:
    test.append(0)
ap=request.form.get("appet")
if ap=='good':
    test.append(1)
elif ap=='poor':
    test.append(0)
else:
    test.append(np.nan)
p=request.form.get("pe")
```

```
        if p=='yes':
            test.append(1)
        else:
            test.append(0)
        an=request.form.get("ane")
        if an=='yes':
            test.append(1)
        else:
            test.append(0)
    print(test)
    test_df=pd.DataFrame(test)
    test_df=np.array(test_df).reshape(1, -1)

    ans1=loaded_class.predict(test_df)
    ans2=loaded_reg.predict(test_df)
    if int(ans1)==1:
        answer1="Sorry to say!! You have CHRONIC DISEASE!!!"
        return render_template('rename.html',answer1=answer1,answer2=ans2)
    else:
        answer1="Happy to say that you don't have CHRONIC DISEASE"

        return render_template('rename2.html',answer1=answer1,answer2=ans2)

if __name__ == "__main__":
    app.debug=True
    app.run(debug=False)
```

## 7.3 DATABASE SCHEMA

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import roc_curve, auc, confusion_matrix,
classification_report,accuracy_score
from sklearn.ensemble import RandomForestClassifier
import warnings
warnings.filterwarnings('ignore')

%matplotlib inline

def auc_scorer(clf, X, y, model): # Helper function to plot the ROC curve
    if model=='RF':
        fpr, tpr, _ = roc_curve(y, clf.predict_proba(X)[:,1])
    elif model=='SVM':
        fpr, tpr, _ = roc_curve(y, clf.decision_function(X))
    roc_auc = auc(fpr, tpr)

    plt.figure()    # Plot the ROC curve
    plt.plot(fpr, tpr, label='ROC curve from '+model+' model (area = %0.3f)' % roc_auc)
    plt.plot([0, 1], [0, 1], 'k--')
    plt.xlim([0.0, 1.0])
    plt.ylim([0.0, 1.05])
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title('ROC Curve')
    plt.legend(loc="lower right")
    plt.show()

    return fpr,tpr,roc_auc
```

```
 df = pd.read_csv("C:/Users/Sinegalatha/Desktop/2nd year online class/nalaiya
thiran/dataset/kidney_disease.csv")
df.head()
df['wc']

df.info()

df.describe()

df[df.duplicated()]

df.isna().sum()

df2 = df.dropna(axis=0)

df.head()

df2['class'].value_counts()

corr_df = df2.corr()

mask = np.zeros_like(corr_df, dtype=np.bool)
mask[np.triu_indices_from(mask)] = True

# Set up the matplotlib figure
f, ax = plt.subplots(figsize=(11, 9))

# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True)

# Draw the heatmap with the mask and correct aspect ratio
sns.heatmap(corr_df, mask=mask, cmap=cmap, vmax=.3, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5})
plt.title('Correlations between different predictors')
plt.show()

X_train, X_test, y_train, y_test = train_test_split(df2.iloc[:,:-1], df2['class'],
                                                     test_size = 0.33, random_state=44,
                                                    stratify= df2['class'] )
X_train.head()
print(X_train.shape)
print(X_test.shape)
```

```
y_train.value_counts()

tuned_parameters = [{'n_estimators':[7,8,9,10,11,12,13,14,15,16],'max_depth':[2,3,4,5,6,None],
                 'class_weight':[None,{0: 0.33,1:0.67},'balanced'],'random_state':[42]}]
clf = GridSearchCV(RandomForestClassifier(), tuned_parameters, cv=10,scoring='f1')
clf.fit(X_train, y_train)

print("Detailed classification report:")
y_true, lr_pred = y_test, clf.predict(X_test)
print(classification_report(y_true, lr_pred))

confusion = confusion_matrix(y_test, lr_pred)
print('Confusion Matrix:')
print(confusion)
fpr,tpr,roc_auc = auc_scorer(clf, X_test, y_test, 'RF')

print('Best parameters:')
print(clf.best_params_)
clf_best = clf.best_estimator_

plt.figure(figsize=(12,3))
features = X_test.columns.values.tolist()
importance = clf_best.feature_importances_.tolist()
feature_series = pd.Series(data=importance,index=features)
feature_series.plot.bar()
plt.title('Feature Importance')

list_to_fill = X_test.columns[feature_series>0]
print(list_to_fill)

corr_df = pd.isnull(df).corr()

mask = np.zeros_like(corr_df, dtype=np.bool)
mask[np.triu_indices_from(mask)] = True

f, ax = plt.subplots(figsize=(11, 9))

cmap = sns.diverging_palette(220, 10, as_cmap=True)

sns.heatmap(corr_df, mask=mask, cmap=cmap, vmax=.3, center=0,
        square=True, linewidths=.5, cbar_kws={"shrink": .5})
```

```
plt.show()

df2 = df.dropna(axis=0)
no_na = df2.index.tolist()
some_na = df.drop(no_na).apply(lambda x: pd.to_numeric(x,errors='coerce'))
some_na = some_na.fillna(0) # Fill up all Nan by zero.

X_test = some_na.iloc[:,:-1]
y_test = some_na['class']
y_true = y_test
lr_pred = clf_best.predict(X_test)
print(classification_report(y_true, lr_pred))

confusion = confusion_matrix(y_test, lr_pred)
print('Confusion Matrix:')
print(confusion)

print('Accuracy: %3f' % accuracy_score(y_true, lr_pred))
# Determine the false positive and true positive rates
fpr,tpr,roc_auc = auc_scorer(clf_best, X_test, y_test, 'RF')

import pickle
pickle. dump(clf_best, open('randomclass_chronic', 'wb')
from sklearn.ensemble import RandomForestRegressor
reg=RandomForestRegressor()
reg.fit(X_train,y_train)
RandomForestRegressor()
y_pred=reg.predict(X_test)
pickle. dump(reg, open('randomreg_chronic', 'wb'))
y_pred
l_pred=list(y_pred)
l_test=list(y_test)
d={'prob':l_pred,'out':y_test}
df_i=pd.DataFrame(d)
df_i.head()
df_i.to_csv('C:/Users/Sinegalatha/Desktop/2nd year online class/nalaiya
thiran/output/file1.csv')
```

# CHAPTER - 8

# TESTING

## 8.1 TEST CASES

| TEST CASE NO | TEST CASE SCENARIO | RESULT |
|---|---|---|
| 1 | Verify user is able to fill id and password and login to the website | Pass |
| 2 | Verify user has to fill the information | Pass |
| 3 | Verify user is able to navigate from homepage to other pages? | Pass |
| 4 | Verify user is able to get the expected results accurately | Pass |
| 5 | Verify information collected page elements | Pass |

### 8.2 USER ACCEPTANCE TESTING

| Section | Total Cases | Not Tested | Fail | Pass |
|---|---|---|---|---|
| Print Engine | 7 | 0 | 0 | 7 |
| Client Application | 51 | 0 | 0 | 51 |
| Security | 2 | 0 | 0 | 2 |
| Outsource Shipping | 3 | 0 | 0 | 3 |
| Exception Reporting | 9 | 0 | 0 | 9 |
| Final Report Output | 4 | 0 | 0 | 4 |
| Version Control | 2 | 0 | 0 | 2 |

# CHAPTER - 9

# RESULT

### 9.1 PERFORMANCE METRICES

Here we will be evaluating the model built. We will be using the test set for evaluation. The test set is given to the model for prediction and prediction values are stored in another variable called y_pred. The score of the model is calculated and its performance isestimated.

| Learning Algorithm | Test Data | Training Data | Training Time |
|---|---|---|---|
| **Linear Regression** | **0.84** | **0.85** | **6 Minutes** |

# CHAPTER - 10

# ADVANTAGES & DISADVANTAGES

## ADVANTAGES

- In early detection of CKD should be beneficial because it enables clinicians to initiate effective treatment of mild disease, preventing loss of kidney function and delaying or avoiding progression to kidney failure.

- Higher preexisting rate of kidney testing in at-risk groups

- A lower prevalence of disease in people with and without such risk factors

- A higher prevalence of other diseases that require similar treatment among people with CKD,and which could be detected instead

- A lower baseline likelihood of developing kidney failure in the reservoir of cases (*e.g.*, slower progression or shorter life expectancy in the absence of early detection)

- Lower capacity to treat newly identified cases of CKD

- Overtesting in the target population upon initiation of early detection (*e.g.*, testing more frequently than necessary, duplicate testing from multiple providers)

## DISADVANTAGES

- Poverty related factors such as infectious diseases secondary to poor sanitation, inadequate supply of safe water, environmental pollutants and high concentrations of disease-transmitting vectors continue to play an important role in the development of CKD in low-income countries.

- Although rates of diabetic nephropathy are rising, chronic glomerulonephritis and interstitial nephritis are among the principal causes of CKD in many countries. Of note is the emergence of HIV-associated nephropathy as the major cause of CKD in Sub-Saharan Africa.

- A high prevalence of CKD of unknown etiology has been reported in rural agricultural communities from Central America, Egypt, India and Sri Lanka. Male farmworkers are affected disproportionately, and the clinical presentation is suggestive of interstitial nephritis, confirmed on renal biopsies.

- The strong association with farm work has led to suggestions that exposure to agrochemicals, dehydration, and consumption of contaminated water might be responsible.Additionally, the use of traditional herbal medications is common and frequently associated with CKD among the poor.

- In Mexico, CKD prevalence among the poor is 2-3fold higher than the general population, and the etiology is unknown in 30% of ESRD patients.

# CHAPTER -11

# CONCLUSION

This study developed an algorithm for predicting CKD at an early stage. The dataset contains input parameters obtained from CKD patients, and the models are trained and validated using the valid parameters. To diagnose CKD, decision tree, random forest, and support vector machine learning models are built. The accuracy of prediction is used to assess the performance of the models. The study's findings revealed that the Random Forest Classifier model outperforms Decision Trees and Support Vector Machines in predicting CKD. As an extension of this research, the comparison may also be done depending on the duration of execution and feature set selection.

# CHAPTER - 12

# FUTURE SCOPE

The National Kidney Foundation's Kidney Disease Outcomes Quality Initiative (K/DOQI) Clinical Practice Guidelines on Chronic Kidney Disease: Evaluation, Classification and Stratification of Risk published in 2002 provided the first definition of CKD independent of cause, and classification of severity based on GFR level. The guidelines have been widely disseminated and generally accepted.However, concerns have been expressed about the definition and classification, methods to estimate GFR, and ascertainment of proteinuria.

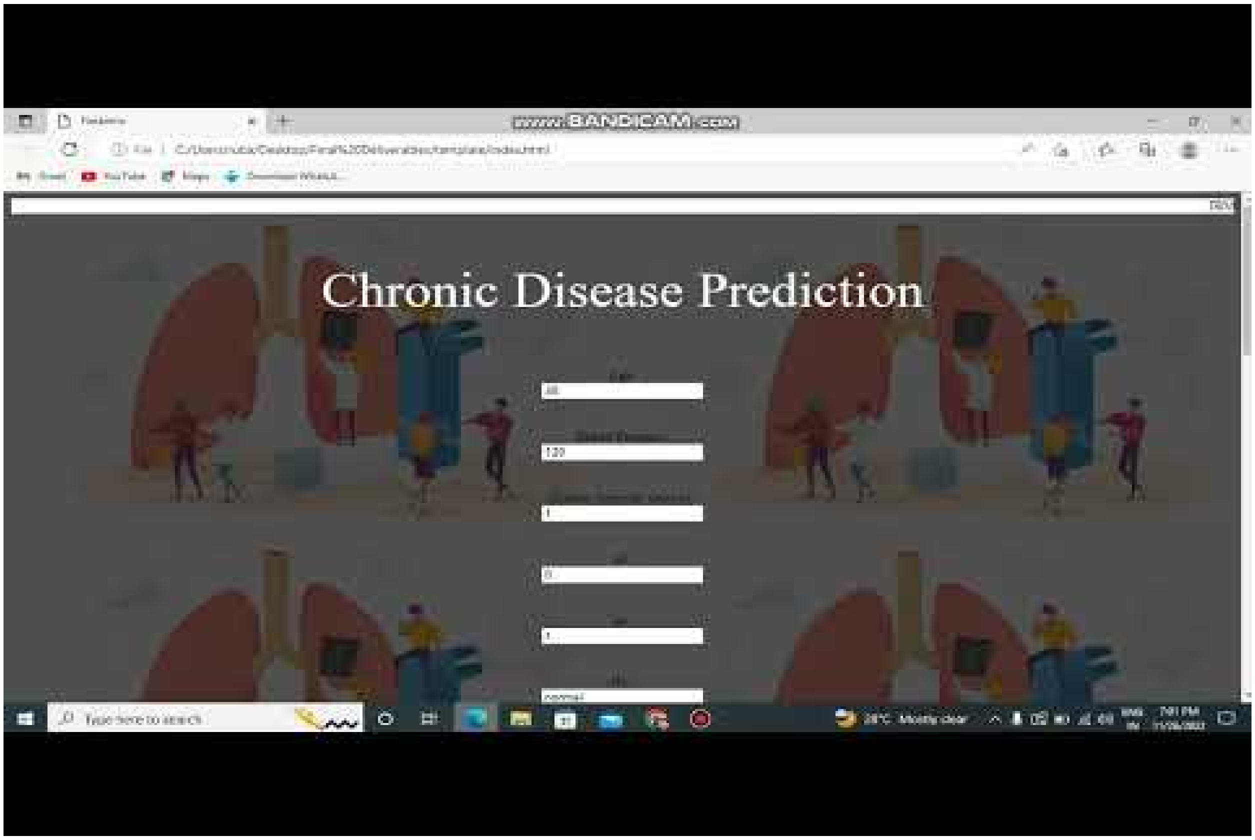
The goals for the KDIGO Controversies Conference were  to provide a clear understanding to both the nephrology  and non nephrology communities of the evidence base for the K/DOQI definition and classification of severity of CKD; to develop global consensus for the adoption of a simple definition and classification system for CKD, clarifications and modifications to current guidelines to facilitate more widespread implementation of initiatives for patient care and physician and public education worldwide; and  to identify a collaborative research agenda and plan that would improve the evidence base and facilitate the implementation of the definition and classification of CKD

# CHAPTER - 13

# APPENDIX

## GitHub & Project Demo Link

https://github.com/IBM-EPBL/IBM-Project-41116-166063955
https://drive.google.com/file/d/1o_hUlkIA765i-7JuOjQ_r8UC4t-AjAlX/view?usp=share_link


www.BANDICAM.com
Chronic Disease Prediction
Type here to search