# EARLY PREDICTION OF CHRONIC KIDNEY DISEASE USING MACHINE LEARNING

Nalaiya Thiran project based learning on Professional readlines for innovation, employment and entrepreneurship

# BACHELOR OF ENGINEERING COMPUTER SCIENCE AND ENGINEERING

#### KARPAGAM INSTITUTE OF TECHNOLOGY

#### **COIMBATORE-641105**

#### A PROJECT BY

**TEAM ID:** PNT2022TMID31803

**TEAM SIZE:** 4

**TEAM LEADER:** GOKUL R

**TEAM MEMBER:** SASIDHARAN M

**TEAM MEMBER:** SUPRAJA S

**TEAM MEMBER:** SURESH T

S.NO	CONTENTS	PAGES
1	INTRODUCTION	1
1.1	Project Overview	1
1.2	Purpose	2
2	LITERATURE SURVEY	3
2.1	Existing problem	3
2.2	References	3
2.3	Problem Statement Definition	5
3	IDEATION & PROPOSED SOLUTION	7
3.1	Empathy Map Canvas	7
3.2	Ideation & Brainstorming	11
3,3	Proposed Solution	16
3.4	Problem Solution fit	18
4	REQUIREMENT ANALYSIS	20
4.1	Functional requirement	20
4.2	Non-Functional requirements	21
5	PROJECT DESIGN	22

5.1	Data Flow Diagrams	22
5.2	User Stories	23
5.3	Solution & Technical Architecture	24
6	PROJECT PLANNING & SCHEDULING	25
6.1	Sprint Planning & Estimation	25
6.2	Sprint Delivery Schedule	28
7	CODING & SOLUTIONING (Explain the features added in the project along with code)	30
8	TESTING	31
8.1	User Acceptance Testing	32
9	RESULTS	33
9.1	Performance Metrics	33
10	ADVANTAGES & DISADVANTAGES	35
11	CONCLUSION	35
12	FUTURE SCOPE	36
13	APPENDIX	37
13.1	Source Code	37

#### 1. INTRODUCTION

#### 1.1 PROJECT OVERVIEW

In today's era everyone is trying to be conscious about health although due to workload and busy schedule one gives attention to the health when it shows any symptoms of some kind. But CKD is a disease which doesn't shows symptoms at all or in some cases it doesn't show any disease specific symptoms it is hard to predict, detect and prevent such a disease and this could be lead to permanently health damage, but machine learning can be hope in this problem it is best in prediction and analysis. By using data of CKD patients with 14 attributes and 400 record we are going to use various machine learning techniques like Decision Tree, SVM, etc. To build a model with maximum accuracy of predicting whether CKD or not and if yes then its severity. Chronic Kidney Disease is one of the most critical illness nowadays and proper diagnosis is required as soon as possible. Machine learning technique has become reliable for medical treatment. With the help of a machine learning classifier algorithms, the doctor can detect the disease on time. For this perspective, Chronic Kidney Disease prediction has been discussed in this article. Chronic Kidney Disease dataset has been taken from the UCI repository. Seven classifier algorithms have been applied in this research such as artificial neural network, C5.0, Chi-square Automatic interaction detector, logistic regression, linear support vector machine with penalty L1 & with penalty L2 and random tree. The important feature selection technique was also applied to the dataset. For each classifier, the results have been computed based of (i) full features, (ii) correlation-based feature selection, (iii) Wrapper method feature selection, (iv) Least absolute shrinkage and selection operator regression, (v) synthetic minority over-sampling technique with least absolute shrinkage and selection operator regression selected features, (vi) synthetic minority oversampling technique with full features. From the results, it is marked that LSVM with penalty L2 is giving the highest accuracy of 98.86% in synthetic minority over-sampling technique with full features. Along with accuracy, precision, recall, F-measure, area under the curve and GINI

coefficient have been computed and compared results of various algorithms have been shown in the graph. Least absolute shrinkage and selection operator regression selected features with synthetic minority over-sampling technique gave the best after synthetic minority over-sampling technique with full features. In the synthetic minority over-sampling technique with least absolute shrinkage and selection operator selected features, again linear support vector machine gave the highest accuracy of 98.46%. Along with machine learning models one deep neural network has been applied on the same dataset and it has been noted that deep neural network achieved the highest accuracy of 99.6%.

#### 1.2 PURPOSE

Your kidneys, each just the size of a computer mouse, filter all the blood in your body every 30 minutes. They work hard to remove wastes, toxins, and excess fluid. They also help control blood pressure, stimulate production of red blood cells, keep your bones healthy, and regulate blood chemicals that are essential to life. Kidneys that function properly are critical for maintaining good health, however, more than one in seven American adults are estimated to have chronic kidney disease (CKD).some other health consequences of CKD include, anemia or low number of red blood cells, increased occurrence of infections, low calcium levels, high potassium levels, and high phosphorus levels in the blood, loss of appetite or eating less, depression or lower quality of life CKD has varying levels of seriousness. So avoid this disease by using machine learning methods. While we using machine learning method we can early predict the disease with the help of the dataset. The dataset includes age, albumin, rbc, pc, pcc, sc, rbcc, anamiea.

#### 2. LITERATURE SURVEY

#### 2.1 EXISTING PROBLEM

Majority of the time the disease is detected in its final stage and which sometimes leads to kidney failure. The existing system of diagnosis is based on the examination of urine with the help of serum creatinine level. Many medical methods are used for this purpose such as screening, ultrasound method. Chronic kidney disease occurs when a disease or condition impairs kidney function, causing kidney damage to worsen over several months or years. Diseases and conditions that cause chronic kidney disease include: Type 1 or type 2 diabetes. high blood pressure.

#### 2.2 REFERENCES

#### [1] Author Name: Tomas E

The small dataset of 400 records have been collected from Apollo Hospital, India in 2015 taken over a two-month period. ANOVA test, the Pearson's correlation, and the Cramer's V test are applied and removed the redundant features in dataset. By using filter feature selection method, three features (hemoglobin, albumin, and specific gravity) are selected and trained using Logistic regression, support vector machines, random forest, and gradient boosting algorithm and reached an accuracy of 99.1% accuracy. Data used in this research is small. So, in future, need to validate the results by using big dataset and for reducing the prevalence of CKD, planned to predict if a person with CKD risk factors such as diabetes, hypertension, and family history of kidney failure will have CKD in the future or not by using appropriate dataset.

#### [2] Author Name: Ebrahime Mohammed Senan, Nizar Alsharif

The dataset was collected from 400 patients containing 24 features. The dataset was divided into 75% training and 25% testing and validation. The dataset was processed to

remove outliers and replace missing numerical and nominal values using mean and mode statistical measures respectively.

## [3] Author Name: Bidri Deepik, Vasudeva Rao KR, Dharmaj N Rampure, Prajwal P and Devanand Gowda G

Early Prediction of Chronic Kidney Disease by using Naive bayes, K-Nearest neighbor. KNN algorithm takes CKD parameters as input and predicts the disease based on old CKD patient's data. To develop user interface for create easier communication between doctors and patients.

#### [4] Author Name: Sreeji S, Balamurugan Balusamy

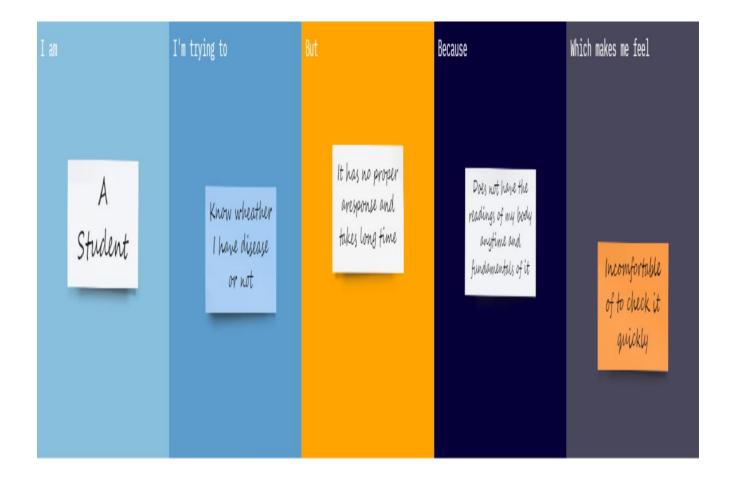
In Design System For Early Detection And Prediction Of Chronic Kidney Disease Using Machine Learning Techniques, The SVM, and Naive Bayes classification algorithms were applied on the processed data. The information assortment contains 400 patient records, for certain qualities missing. It comprises of 24 clinical highlights that show up in the anticipation of ongoing kidney illness, with one class quality demonstrating the event of persistent renal disappointment in the patient. Data used in this research is small. In future by using better algorithms the accuracy can be increased and also need to validate the results by using big dataset.

#### [5] Author Name: Reshma S , Salma Shaji , S R Ajina, Vishnu Priya S R , Janisha A

In this research, The dataset contains 400 samples of two different classes. Out of 25 attributes,11 are numeric and 13 are nominal and one is class attribute. The data set contains number of missing values. Here the information of dataset uses the patient's data like age, blood pressure, specific gravity, albumin, sugar, red blood cells etc. The sample Data used in this research is small. So, in future by using by using better deep learning concepts the accuracy can be increased.

#### 2.3 PROBLEM STATEMENT DEFINITION

Irregular lifestyle of current days leads to many disease chronic diseases. One of them being Chronic Kidney Disease(CKD), it has been of a growing concern, kidney is one of the most important organs in the body required for filtering blood, once a person has lost their kidneys, they could survive only for 18 days without their kidneys, it would take a fortune to just keep the person alive, with treatments like dialysis, transplant etc. 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. People usually don't realize that the medical tests we perform for various purposes can contain valuable information related to kidney disease. Subsequently, the attributes of various medical tests are examined to distinguish which attributes may contain useful information about the disease. The information, they say, helps us gauge the severity of the problem, and we use that information to build a machine learning model that predicts chronic kidney disease. Early detection of kidney disease can help in treatment which could save lives. Analyzing various medical tests, would give us an idea about which attributes help us distinguish the disease. The main aim of this project is to predict whether the patient have chronic kidney disease or not, in more accurate and faster way based on certain diagnostic measurements like Blood Pressure (Bp), Albumin(Al).



- 1. Continuous increase in the number of patients with end-stage renal disease demands early detection of chronic kidney disease (CKD). The goal is to diagnose CKD in its earliest stages in a randomly selected population using a diagnostic algorithm developed by the working group.
- 2.The major problem that we are facing is that lack of capacity to implement huge datasets at the same time and most of the algorithms fail for the real time prediction. Some traditional algorithms were used here, it gets more duration to category and classifies the give data set. Due to this, it becomes not accurate and flexible

#### 3. IDEATION & PROPOSED SOLUTION

#### 3.1 EMPATHY MAP CANVAS

#### 3.1 Empathy Map Canvas

An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment.

#### **Process: How to Build an Empathy Map**

Go through the following steps to create a valid and useful empathy map:

#### 1. Define scope and goals

- **a.** What user or persona will you map? Will you map a persona or an individual user? Always start with a 1:1 mapping (1 user/persona per empathy map). This means that, if you have multiple personas, there should be an empathy map for each.
- **b. Define your primary purpose for empathy mapping.** Is it to align the team on your user? If so, be sure everyone is present during the empathy-mapping activity. Is it to analyse an interview transcript? If so, set a clear scope and timebox your effort to ensure you have time to map multiple user interviews.

#### 2. Gather materials

Your purpose should dictate the medium you use to create an empathy map. If you will be working with an entire team, have a large whiteboard, sticky notes, and markers readily available. (The outcome will look somewhat like the illustration above.) If empathy mapping alone, create a system that works for you. The easier to share out with the rest of the team, the better.

#### 3. Collect research

Gather the research you will be using to fuel your empathy map. Empathy mapping is a qualitative method, so you will need qualitative inputs: user interviews, field studies, diary studies, listening sessions, or qualitative surveys.

#### 4. Individually generate sticky notes for each quadrant

Once you have research inputs, you can proceed to mapping as a team. In the beginning, everybody should read through the research individually. As each team member digests the data, they can fill out sticky notes that align to the four quadrants. Next, team members can add their notes to the map on the whiteboard.

#### 5. Converge to cluster and synthesize

In this step, the team moves through the stickies on the board collaboratively and clusters similar notes that belong to the same quadrant. Name your clusters with themes that represent each group (for example, "validation from others" or "research"). Repeat themes in each quadrant if necessary. The activity of clustering facilitates discussion and alignment — the goal being to arrive at a shared understanding of your user by all team members.

#### 6. Polish and plan

If you feel that you need more detail or you have unique needs, adapt the map by including additional quadrants (like Goals the example below) or by increasing specificity to existing quadrants. Depending on the purpose of your empathy map, polish and digitize the output accordingly. Be sure to include the user, any outstanding questions, the date and version number. Plan to circle back to the empathy map as more research is gathered or to guide UX decisions.

#### **Format**

Traditional empathy maps are split into 4 quadrants (*Says*, *Thinks*, *Does*, and *Feels*), with the user or persona in the middle. Empathy maps provide a glance into who a user is as a whole and are **not** chronological or sequential.

The *Says* quadrant contains what the user says out loud in an interview or some other usability study. Ideally, it contains verbatim and direct quotes from research.

- "I am allegiant to Delta because I never have a bad experience."
- "I want something reliable."
- "I don't understand what to do from here."

The *Thinks* quadrant captures what the user is thinking throughout the experience. Ask yourself (from the qualitative research gathered): what occupies the user's thoughts? What matters to the user? It is possible to have the same content in both *Says* and *Thinks*. However, pay special attention to what users think, but may not be willing to vocalize. Try to understand why they are reluctant to share — are they unsure, self-conscious, polite, or afraid to tell others something?

- "This is really annoying."
- "Am I dumb for not understanding this?"

The *Does* quadrant encloses the actions the user takes. From the research, what does the user physically do? How does the user go about doing it?

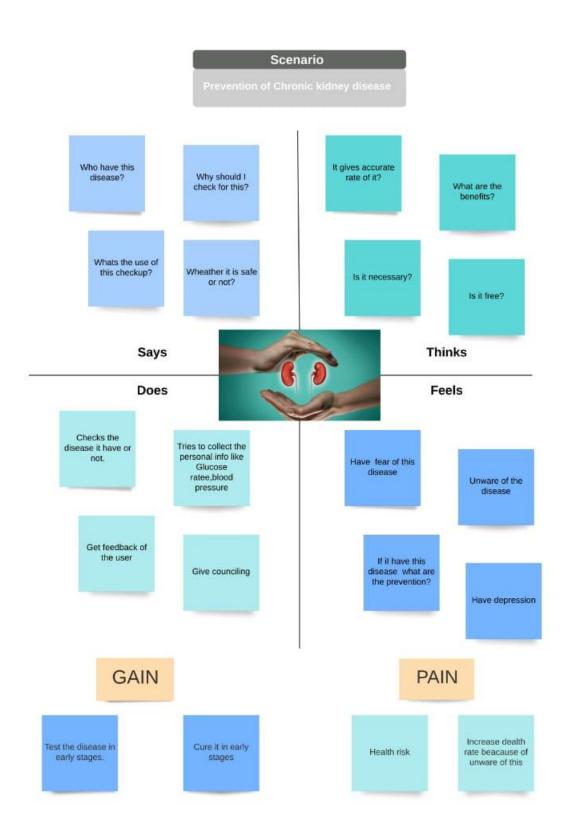
- Refreshes page several times.
- Shops around to compare prices.

The *Feels* quadrant is the user's emotional state, often represented as an adjective plus a short sentence for context. Ask yourself: what worries the user? What does the user get excited about? How does the user feel about the experience?

- Impatient: pages load too slowly
- Confused: too many contradictory prices
- Worried: they are doing something wrong

#### Strengths and limitations of this study

- Throughout this study, we engaged with patient and community partners to address priorities of enhanced self-management and care experience for people living with chronic kidney disease.
- Patient partners contributed their lived experience and methodological expertise through all phases of this research, including study design, collection of rich data through focus group discussions and thematic generation.
- Inclusion of participants with varying stages of advanced, non-dialysis chronic kidney disease, their caregivers, as well as trained peer mentors with kidney disease experience provided a breadth of perspectives and depth to our findings.
- Most participants had not accessed or were unaware of kidney-focused peer support opportunities, which may have influenced the scope of expressed preferences for peer support delivery.
- As most participants were older and primarily English speaking, transferability of findings to younger or more ethnically or culturally diverse individuals with kidney disease may be limited.

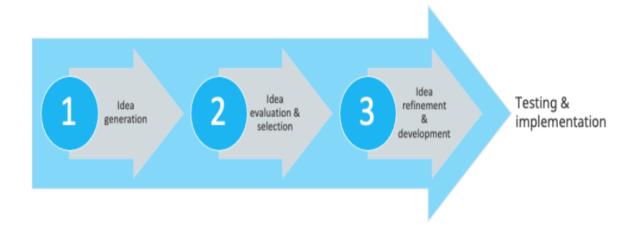


#### 3.2 IDEATION & BRAINSTORMING

Ideation is often closely related to the practice of brainstorming, a specific technique that is utilized to generate new ideas. A principal difference between ideation and brainstorming is that ideation is commonly more thought of as being an individual pursuit, while brainstorming is almost always a group activity. Brainstorming is usually conducted by getting a group of people together to come up with either general new ideas or ideas for solving a specific problem or dealing with a specific situation.

Ideation is often the most exciting stage in a Design Thinking project because almost unrestrained free thinking can occur within the given field. In the Ideation stage, the aim is to generate a large number of ideas — ideas that potentially inspire newer, better ideas — which the team can then filter and narrow down into the best, most practical, or most innovative ones. There are many great methods that can help the design team during the Ideation sessions.





**Step 1:** Team gathering, collaboration and select the problem statement:



# 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.

- 10 minutes to prepare
- 1 hour to collaborate
- 2-8 people recommended



#### Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

#### A Team gathering

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

Set the goal

Think about the problem you'll be focusing on solving in the brainstorming session.

Learn how to use the facilitation tools

Use the Facilitation Superpowers to run a happy and productive session.

Open article →

#### Step: 2 Brainstorm, idea Listing and Grouping:



#### **Brainstorm**

Write down any ideas that come to mind that address your problem statement.

10 minutes

#### Gokul R Suresh T

Provides a portal to donate for CKD children's fund Facility to track treatments, tests and medications

Facility for users to give reviews about a hospital or testing facility Provides virtual interface between doctors and patients for remote consultation

Predicts whether the user has CKD

Suggests nearby test centres and hospitals for treatment Displaying a general overview of CKD including treatments and expenses

Provide a calendar to remind user of future treatments

#### Sasidharan M

Displays the prediction accuracy The lab test results can be viewed and shared by the user

Notifies the user of subsequent tests and medicine purchases Displays the expected date and time of test results Supraja S

Suggests diet chart and physical exercise guide

Suggests precautionary measures for CKD

Facility to talk to a counsellor Analyse whether the user is eligible for CKD children's fund

Facility to order medications online

Facility to join a wellness community and connect with patients who have successfully recovered from CKD Display the government health schemes available

Book appointments in hospitals

#### Step: 3 Grouping idea



#### Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

① 20 minutes

Predicts whether the user has CKD	ANALYSIS	SUGGESTIONS	Suggests precautionary measures for CKD
Analyse whether the user is eligible for CKD children's fund	Displays the prediction accuracy	Suggests diet chart and physical exercise guide	Suggests nearby test centres and hospitals for treatment
Facility to track treatments, tests and medications	REMINDER	WELLNESS	Facility to Join a wellness community and connect with patients who have successfully recovered from CKD
Provide a calendar to remind user of future treatments	Notifies the user of subsequent tests and medicine purchases	Provides a portal to donate for CKD children's fund	Facility to talk to a counsellor
Displays the expected date and time of test results	Displaying a general overview of CKD including treatments and expenses	Book appointments in hospitals	Facility for users to give reviews about a hospital or testing facility
INFO	RMATION	CONNECT WIT	TH HOSPITALS
Display the government health schemes available	The lab test results can be viewed and shared by the	Facility to order medications online	Provides virtual Interface between doctors and patients for remote consultation

#### Step: 4 Idea prioritization

HCERT.



#### Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

1 20 minutes



#### 3.3 PROPOSED SOLUTION

Proposed solution should relate the current situation to a desired result and describe the benefits that will accrue when the desired result is achieved. So, begin your proposed solution by briefly describing this desired result.

#### 3.3.1 Problem Statement (Problem to be solved):

Chronic Kidney Disease (CKD) is a major medical problem and can be cured if treated in the early stages. Usually, people are not aware of it. 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 the problem and it predicts Chronic Kidney Disease it have or not

#### 3.3.2 Idea / Solution description:

We have an certain test readings and the analysis of the readings to predict the person to have a disease or not

#### 3.3.3 Novelty / Uniqueness:

We have the all main readings and it gives the best analysis to it. The renal patient is recognized by undertaking two primary tests. \*A Blood Test to determine Glomerular Filteration Rate(GFR). \*A Urine Test to determine Albumin.

#### **3.3.4 Social Impact / Customer Satisfaction:**

It will surely give the customer satisfying about the result because it have the best result analysis with proper guidelines to cure the disease. As people don't undergo the general test of their health, early detection of CKD is not i dentified. This creates a great social impact of not being aware of CKD. As a result of this many people are getting affected by CKD.

#### 3.3.5 Business Model (Revenue Model):

Strengthening Healthcare Infrastructure.

- Capacity Building of Human Resources.
- Strong Intersectoral co-ordination

#### 3.3.6 Scalability of the Solution:

Strong healthcare solutions Early detection makes us to cure in early stages. Sustained human - to - human transmission. Outbreaks in countries in one region Several outbreaks in other regions. Post peak - levels dropped below peak levels.

#### 3.4 PROBLEM SOLUTION FIT:

#### CS 1. CUSTOMER SEGMENT(S) Define CS, fit into CC Our customer are from all the ages like childhood to adulthood. They are like students or any other age category Focus on J&P, tap into BE, understand RC J&P 2. JOBS-TO-BE-DONE / PROBLEMS Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides. The problem we identify which they have the chronic disease or not. It will help them to take treatment for their general and also personal care for them 3. TRIGGERS TR It will take the customer to have a healthy life. It will check the test of individual for an Identify strong TR & EM better life of other EM 4. EMOTIONS: BEFORE / AFTER Before: They feel insecure of their health. After: They feel better and have a stress free life after the complete checkup

#### **6. CUSTOMER CONSTRAINTS**

CC

RC

#### 5. AVAILABLE SOLUTIONS



- They have spend less time for their quick result of test report.
- Spend less time less time and also less cost
- You can look after your kidneys by eating healthy food, quitting smoking, limiting alcohol consumption, staying active and maintaining a healthy weight.
- Many diseases can affect your kidneys.
- In most cases, early diagnosis and good management can prevent a condition from worsening and reduce the risk of
- If detected early enough, the progress of kidney disease can be slowed and sometimes even prevented. In the early stages, changes to diet and medication can help to increase the life of your kidneys.

# Explore AS, differentiate

BE

#### 9. PROBLEM ROOT CAUSE

Signs and symptoms of chronic kidney disease develop over time if kidney damage progresses slowly.

Loss of kidney function can cause a buildup of fluid or body waste or electrolyte problems

Nausea Vomiting Fatigue and weakness

Sleep problems Signs and symptoms of kidney disease are often nonspecific. This means they can also be caused by other illnesses. Because your kidneys are able to make up for lost function, you might not develop signs and symptoms until irreversible damage has occurred.

#### 7. BEHAVIOUR

You are more at risk of developing chronic kidney disease if you:

- have high blood pressure
- have diabetes
- ve established heart problems (heart failure or past heart attack) or have had a stroke
- are obese (a body mass index of greater than 30)\

Heathy lifestyle choices to keep your kidneys functioning well include:

- Drink plenty of water instead of other drinks. Minimise consumption of sugary soft
- drinks. Maintain a healthy weight.

# Focus on J&P, tap into BE, understand

#### 10. YOUR SOLUTION



- We provide free consultancy to customers for have an better solution via a phone call or any other video call according to their

We provide the better UI to interact with

It will analysis the best test cases to it.

the user and also the doctors.

#### CH

**CHANNELS of BEHAVIOUR** 

- It will analysis the results anywhere from Provide better UI to understand it more detailed any place to the user

- In offline the peoples have to spend time for this checkup. And also waste money from the thread persons so online is better than the offline checkup

Extract online & offline CH of BE

### 4. REQUIREMENT ANALYSIS 4.1 FUNCTIONAL REQUIREMENT

Requirements analysis, also called requirements engineering, is **the process of determining user expectations for a new or modified product**. These features, called requirements, must be quantifiable, relevant and detailed. In software engineering, such requirements are often called functional specifications.

Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describe all the cases where the system uses the functional requirements, these are captured in use cases.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)			
FR-1	User Registration	Registration through Form			
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP			
FR-3	User Requirements	A user friendly application that can detect the chronic kidney disease within a short period of time.			
FR-4	Business Requirements	Must be available and be easily useable by all the people to stay put in the business Should be cost efficient.			
FR-5	User Authentication	Password should be used to validate the credentials and also be maintained in a secure manner.			

## **4.2 NON FUNCTIONAL REQUIREMENTS**

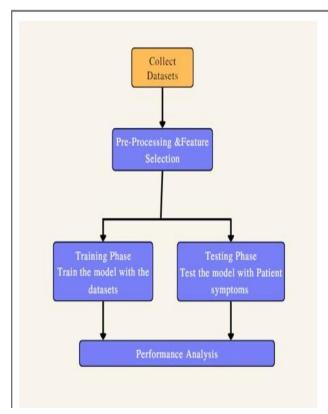
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Proper use ability helps the users to understand the application easily and makes it a better application among the public.
NFR-2	Security	Keeps and maintains the medical records of the patients securely as it involves personal information.
NFR-3	Reliability	Failure can be prevented so that this application becomes more reliable
NFR-4	Performance	Provides user friendly application with good performance
NFR-5	Availability	Easily available to everyone even the people in rural areas

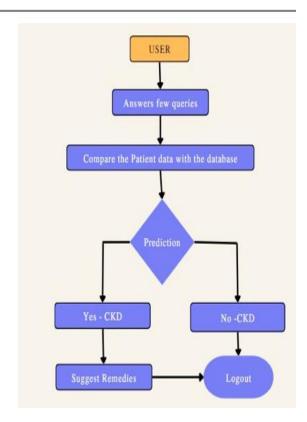
#### 5. PROJECT DESIGN

#### **5.1 DATA FLOW DIAGRAMS**

The thinking that occurs during Project Design, on the other hand, is less concerned with minutiae and more concerned with Project Management software can aid in the organisation of both the high-level strategy and the finer points of a project's design.

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 USER STORIES**

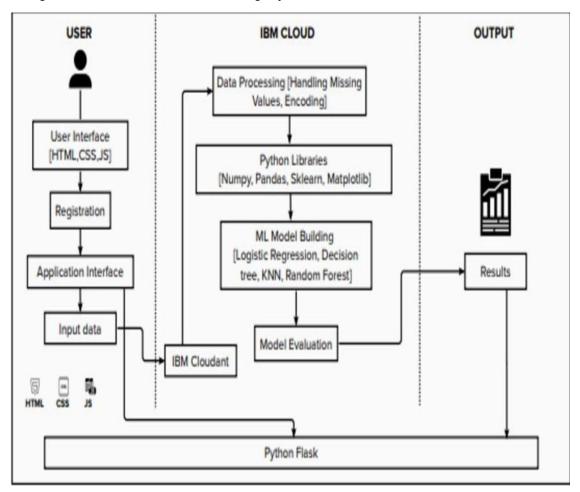
Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email	I can receive confirmation email & click confirm	High	Sprint-1
	Login	USN-3	As a user, I can log into the application by entering email & password	I can create and get into my account	High	Sprint-2
		USN-4	As a user, answers 10- 15 multiple choice queries	I can answer the Quiz	High	Sprint-1
	Provide output to the user	USN-5	As a user, I can view the quiz results	I can view results	High	Sprint-1
Customer Care Executives	Remedies	USN-6	As a customer care executive, I can suggest some remedies	I can suggest remedies	Medium	Sprint-2

Administrator	Predict	USN-7	As the	I deploy	High	Sprint-1
	Disease		admin, I	models		
			build ML	within		
			models to	websites		
			predict			

#### 5.3 SOLUTION & TECHNICAL ARCHITECTURE

Technology architecture associate's application components from application architecture with technology components representing software and hardware components. Its components are generally acquired in the marketplace and can be assembled and configured to constitute the enterprise's technological infrastructure. Technology architecture provides a more concrete view of the way in which application components will be realized and deployed.



#### 6. PROJECT PLANNING & SCHEDULING

#### 6.1 SPRINT PLANNING AND ESTIMATION

Estimation is done by the entire team during Sprint Planning Meeting. The objective of the Estimation would be to consider the User Stories for the Sprint by Priority and by the Ability of the team to deliver during the Time Box of the Sprint Product Owner ensures that the prioritized User Stories are clear, can be subjected to estimation, and they are brought to the beginning of the Product Backlog As the Scrum Team in total is responsible for the delivery of the product increment, care would be taken to select the User Stories for the Sprint based on the size of the Product Increment and the effort required for the same The size of the Product Increment is estimated in terms of User Story Points. Once the size is determined, the effort is estimated by means of the past data, i.e., effort per User Story Point called Productivity.

#### What is Sprint Planning?

Sprint planning is an event in scrum that kicks off the sprint. The purpose of sprint planning is to define what can be delivered in the sprint and how that work will be achieved. Sprint planning is done in collaboration with the whole scrum team.

#### What is Sprint Estimation?

A sprint estimation shows how much effort a series of tasks require. It's based on assumptions, requirements, and dependencies of a project

TITLE	DESCRIPTION	DATE
Literature Survey &	Literature survey on the	27 SEPTEMBER 2022
Information Gathering	selected project & gathering	
	information by referring the	
	technical papers, research	
	publications, journals etc.	
Prepare Empathy Map	Prepare Empathy Map	27 SEPTEMBER 2022
	Canvas to capture the user	
	Pains & Gains, Prepare list of	
	problem Statements that are	
	to be solved by this project.	
Ideation	List the ideas by organizing a	4 OCTOBER 2022
	brainstorming session and	
	prioritize the top 3 ideas	
	based on the feasibility &	
	importance.	
Proposed Solution	Prepare the proposed solution	10 OCTOBER 2022
	document, which includes	
	novelty, feasibility of idea,	
	revenue model, social	
	impact, scalability of	
	solution, etc.	
	Prepare problem - solution fit	15 OCTOBER 2022
Problem Solution Fit	document.	

	Prepare solution architecture	15 OCTOBER 2022
Solution Architecture	document.	
Customan Jaumay	Drange the quetomer journey	21 OCTODED 2022
<b>Customer Journey</b>	Prepare the customer journey	31 OCTOBER 2022
	maps to understand the user	
	interactions & experiences	
	with the application (entry to	
	exit).	
	Prepare the functional	31 OCTOBER 2022
Functional Requirement	requirement document.	
_		
Data Flow Diagrams	Draw the data flow diagrams	31 OCTOBER 2022
	and submit forreview.	
Technology Architecture	Prepare the technology	31 OCTOBER 2022
	architecture diagram.	
Prepare Milestone &	Prepare the milestones	03 NOVEMBER 2022
Activity List	&activity list of the project.	
Project Development -	Develop & submit the	19 NOVEMBER 2022
Delivery of Sprint-1, 2, 3 &	•	
	at . stoped tode of toshing it.	
4		

#### **6.2 SPRINT DELIVERY SCHEDULE**

Since sprints take place over a fixed period of time, it's critical to avoid wasting time during planning and development. And this is precisely where sprint scheduling enters the equation. While there may be multiple project heads collaborating on a sprint, it's ultimately important to have one owner who oversees all aspects of sprint planning. Likewise, there should be one single schedule to avoid confusion and keep projects running according to a set plan. Teams often run into trouble when they create more than one schedule. This can create conflict and derail projects midway through their cycles. To ensure things stay on track, one schedule makes sense.

Sprint	Functional requireme nt (Epic)	User story number	User story andtasks	Story point	priority	Team member
Sprint 1	Data collection	USN 1	Use dataset from Google and clean the dataset	110	High	GOKUL R
Sprint 1	Model	USN 2	Create, test and save the model	10	High	GOKUL R
Sprint2	Display	USN 3	Display user entry form to user	6.7	High	SUPRAJA S, SASIDHARAN M,GOKUL R
Sprint2	Enter data	USN 4	Receive data from user as numeric values	6.7	High	SUPRAJA S, SASIDHARAN M
Sprint2	Enter data	USN 5	Receive data from user as selection from pull down menu	6.7	High	SASIDHARAN M, SURESH T,GOKUL R

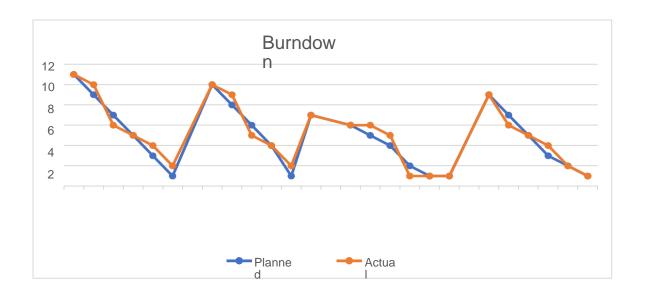
Sprint 3	Select	USN 6	As a user can select prediction	10	Medium	SURESH T,GOKUL R
Sprint 3	View data	USN 7	As a user can view final result	10	Medium	SURESH T,GOKUL R
Sprint 4	Application building for project	USN 8	Deploy into IBM cloud	20	High	GOKUL R

#### Project tracker, velocity:

#### **Velocity:**

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points

AV= SPRINT DURATION /VELOCITY = 20/10 = 2 AV of CKD Project = 20/6 = 3.33 Burndown chart:



#### 7. CODING & SOLUTIONING

models have been constructed using a training data set (280 instances) which is 70% of the original CKD data set . constructed models have been validated using test data which is 30% of original data with respect to the parameter accuracy, here , accuracy has been calculated using a confusion matrix . The best classifier model is the one with highest accuracy.

#### **Accuracy of Decision tree**

Confusion Matrix has been generated by decision tree model for the test data (120 instances) with class (values: CKD, NON CKD) as the target variable is given by table 1. The confusion matrix clearly says that 7 instances are not classified properly and 113 instances have been classified accurately and the accuracy of this classifier model is **94.16%**.

#### **Accuracy of SVM**

Confusion Matrix has been generated by SVM model for the test data(120 instances) with class (values:CKD,NON CKD)as the target variable is given by table 1. The confusion matrix clearly says that 2 instances are not classified properly and 118 instances have been classified accurately and the accuracy of this classifier model is **98.33%** 

#### Accuracy of Random Forest

Confusion Matrix has been generated by SVM model for the test data (120 instances) with class (values:CKD,NONCKD)as the target variable is given by table 1. The confusion matrix clearly says that 1 instances are not classified properly and 119 instances have been classified accuracy.

#### 8. TESTING

#### **UNIT TESTING**

When the testing happens for some individual group or some related units then that type of testing is called as Unit Testing. It is often done by a programmer to test the part of the program he or she has implemented.

Unit Testing is successful means all the modules have been successfully tested and it can proceed further.

#### **FUNCTIONAL TESTING**

This type of testing is tested because to check the functional components or the functionality required from the system is gained or not .It actually falls under the testing of the Black Box testing of Software Engineering. This part includes the feeding of the inputs in the system or the project and to check if that system or the project is getting the same value or not as expected if not then calculate the error as wanted and check for more .Functional Testing of this project mainly involves below things. All of these are tested successfully and errors are also calculated.

#### INTEGRATION TESTING

In a total project or the system, many groups of components are getting added or summed up in the purpose of the project query. Integration testing is about to check the interaction between various modules of the project or the system. This module also includes the hardware and the software requirements of the project. All the individual modules are integrated and tested together.

All the best and extreme cases that the modules are interacting or not are successfully checked and passed, errors are calculated for the machine learning platforms.

#### **SYSTEM TESTING**

This type of testing is actually meant for the system or the project and also the platform and the integrated software and tools, technologies are also tested. The idea or purpose behind the system testing is to check all the requirements that will be provided by the system.

This application of the project along with the tools and technologies has been tested in both Windows and Linux platform and also uncertified online apple MAC platform to check the requirements. It passed successfully.

This is a type of system or software testing where a system has been tested for availability. The purpose of this test is to check the business requirements and assess whether it will be accepted for delivery. In this part ADRIAN of Pyrimag search has been referred to, who worked with the same platform and to check if this project was accepted by the delivery partner or not.

#### 8.2 USER ACCEPTANCE TESTING

Testing is defined as an activity to check whether the actual results match the expected results and to ensure that the software system is defect free. It involves the execution of a software component or system component to evaluate one or more properties of interest. Software testing also helps to identify errors, gaps, or missing requirements in contrast to the actual requirements.

#### 9. RESULT

#### 9.1 PERFORMANCE METRICS

#### 1) CONFUSION MATRIX DESCRIPTION

TP: True Positive means output as positive such that the predicted result is correctly classified.

TN: True Negative means output as negative such that the predicted result is correctly classified.

FP: False Positive means output as positive such that predicted result is incorrectly classified.

FN: False Negative means output as negative such that predicted result is incorrectly classified.

#### 2) CLASSIFICATION ACCURACY

Classification accuracy shows the correct rate of prediction results. It computes the confusion matrix. The classification accuracy is found by equation

$$\mathit{accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} * 100$$

#### 3) CLASSIFICATION ERROR

Classification error shows the incorrect rate of prediction results. It computes the confusion matrix. The classification error is found by equation 3:

$$Error = \frac{FP + FN}{TP + TN + FP + FN} *100$$

### 4) PRECISION

Precision is an important model performance evaluation matrix. It is the fraction of related instances among the total retrieved instances. It is a positive predicted value. The precision is calculated as follows in equation 4:

$$\textit{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} *100$$

### 5) RECALL

Recall is also an important model performance evaluation matrix. It is the fraction of related instances among the total number of retrieved instances. The recall is calculated as follows in equation 5:

$$Recall = \frac{TP}{TP + FN} * 100$$

## 6) F-MEASURE

It is also known as F Score. F-measure is calculated as to measure the accuracy of the test. It is calculated from the precision and recall by equation 6:

$$F - Measure = 2* \frac{Precision * Recall}{Precision + Recall}$$

### 7) ROC AND AUC

The performance of the classification model is measured from the Receiver operating a characteristic curve (ROC).ROC is a graph that is created for true positive rate vs. false positive rate at different classifications threshold. The entire area under the ROC curve is known as area of the curve (AOC). It gives a collective measure of performance across all achievable classifification's threshold.

### 10. ADVANTAGES & DISADVANTAGES

### **ADVANTAGES**

- With using of ANN algorithm, easy for small parameters, multiple processing units are connected to each other very effectively and efficiently errorless.
- With the KNN algorithm, there is no need to separate linear and nonlinear at no cost to learn good for large numbers of records fast to train the model.
- With the use of the DT algorithm, a nonparametric learning model simple to know & visualize .It can produce sets of rules to transfer knowledge .
- With using of the NB algorithm, it can contain numerical & categorical data independent between attributes. It is very fast algorithm.

### **DISADVANTAGES**

- With the ANN algorithm, It is complex and the time is long to train, sometimes overlapping with slow learning.
- With the KNN algorithm, It requires more memory space ,is slow to test , and depends on a number of dimensions sensitive to noise .
- With using of DT algorithm, complex for large dataset, unstable, difficult to learn
- With the use of the NB algorithm, in a small data set to reduce precision, it requires very range.

### 11. CONCLUSION

This paper presented a prediction algorithm to predict CKD at an early stage. The dataset shows input parameters collected from the CKD patients and the models are trained and validated for the given input parameters. Decision tree, Random Forest and Support Vector Machine learning models are constructed to carry out the diagnosis of CKD. The performance of the models are evaluated based on the accuracy of prediction. The results of the research showed

That Random Forest Classifier model better predicts CKD in comparison to Decision trees and Support Vector machines. The comparison can also be done based on the time of execution, feature set selection as the improvisation of this research. The prediction of chronic kidney disease is very important and now-a-days it is the leading cause of death. The performance of the Decision tree method was found to be 99.25% accurate compared to the naive Bayes method. Classification algorithm on chronic kidney disease dataset the performance was obtained as 99.33% Specificity and 99.20% Sensitivity. We are also further working on enhancing the performance of prediction system accuracy in neural networks and deep learning algorithm.

### 12.FUTURE SCOPE

- This would help detect the chances of a person having CKD further on in his life which would be really helpful and cost-effective people.
- This model could be integrated with normal blood report generation, which could automatically flag out if there is a person at risk. Patients would not have to go to a doctor unless they are flagged by the algorithms.
- This would make it cheaper and easier for the modern busy person.

## 13. APPENDIX

## **SOURCE CODE**

# App.py

```
import numpy as np
import pandas as pd
from flask import Flask, request, render_template
import pickle
app = Flask(__name__)
model = pickle.load(open('CKD.pkl', 'rb'))
@app.route("/")
def home():
  return render_template('home.html')
@app.route("/prediction")
```

```
def prediction():
         return render_template("prediction.html")
       @app.route("/home")
       def my_home():
         return render_template('home.html')
       @app.route("/prediction", methods=['POST'])
       def predict():
         #input_features = ([int(x) for x in request.form.values()])
         blood_urea = request.form["blood_urea"]
         blood_glucose_random =
request.form["blood_glucose_random"]
         anemia = request.form["Anemia"]
         if (anemia == "no"):
            anemia = 0
         if (anemia == "yes"):
            anemia = 1
```

```
coronary_artery_disease =
request.form["coronary_artery_disease"]
         if (coronary_artery_disease == "no"):
            coronary_artery_disease = 0
         if(coronary_artery_disease == "yes"):
            coronary_artery_disease = 1
         pus_cell = request.form["pus_cell"]
         if (pus_cell == "no"):
            pus_cell = 0
         if (pus_cell == "yes"):
            pus_cell = 1
         red_blood_cell = request.form["red_blood_cell"]
         if (red_blood_cell == "no"):
            red_blood_cell = 0
         if (red_blood_cell == "yes"):
            red_blood_cell = 1
         diabetics_mellitus = request.form["diabetics_mellitus"]
```

```
if (diabetics_mellitus == "no"):
            diabetics_mellitus = 0
         if (diabetics_mellitus == "yes"):
            diabetics_mellitus = 1
          pedal_edema = request.form["pedal_edema"]
         if (pedal_edema == "no"):
            pedal_edema = 0
         if (pedal_edema == "yes"):
            pedal_edema = 1
         input_features =
[int(blood_urea),int(blood_glucose_random),int(anemia),int(coronary_arte
ry_disease),int(pus_cell),int(red_blood_cell),int(diabetics_mellitus),int(pe
dal_edema)]
          #input_features =
[int(red_blood_cell),int(pus_cell),int(blood_glucose_random),int(blood_ur
ea),int(pedal_edema),int(anemia),int(diabetics_mellitus),int(coronary_arte
ry_disease)]
         print(input_features)
```

app.run(host='localhost', debug=True)

features\_value = [np.array(input\_features)]

### Home.html

```
<html>
<head>
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Early Detection of Chronic Kidney Disease using Machine
Learning</title>
link rel="stylesheet" href="static\style.css">
link
href="https://fonts.googleapis.com/css?family=Poppins:100,200,300,400,600,700
&display=swap" rel="stylesheet">
link rel="stylesheet" href="%link{rel: "stylesheet", href:
```

```
free@6.2.0/css/fontawesome.min.css", integrity: "sha384-
z4tVnCr80ZcL0iufVdGQSUzNvJsKjEtqYZjiQrrYKlpGow+btDHDfQWkFjoaz/Zr",
</head>
</style>
<h1>Early Detection of Chronic Kidney Disease using Machine Learning</h1>
                  <a href="home.html">HOME</a>
                  <a href="about.html">ABOUT</a>
                  <a href="prediction">PREDICTION</a>
                  <a href="result.html">RESULT</a>
                  <a href="contact.html">CONTACT</a>
              </div>
       </nav>
       </div>
   </section>
   <section class="facility">
       <h1>Symptoms</h1>
                  <h3>High BP and Chest pain</h3>
               </div>
                  Swelling of legs occurs because of Diabetes
```

```
the CKD
               </div>
           </div>
   </section>
               <h3>Type 1 or type 2 diabetes</h3>
           </div>
               <h3>Glomerulonephritis</h3>
           </div>
               <h3>Vesicoureteral</h3>
kidneys
           </div>
   </section>
       To care your health with less time and effort spent and to detect
benefited.
</section>
<script>
```

```
</body>
</html>
```

### **Contact.html**

```
<html>
Learning</title>
</head>
h1 {text-align: center;}
</style>
           <a href="index.html"><img src="images/images.png"></a>
                  <a href="home.html">HOME</a>
                  <a href="about.html">ABOUT</a>
                   <a href="result.html">RESULT</a>
                   <a href="contact.html">CONTACT</a>
           </div>
       </nav>
           <h1>Contact Us</h1>
    </section>
    <section class="location">
          <iframe
src="https://www.google.com/maps/embed?pb=!1m18!1m12!1m3!1d3887.368144678623!
```

```
3.1!3m3!1m2!1s0x3bae3d7f3535d71b%3A0x660d25b66e9fb25c!2sWorld+Trade+Centre!5e
0!3m2!1sen!2sin!4v1565255080367!5m2!1sen!2sin" width="720" height="445"
    </section>
    <section class="contact-us">
                    <div>
                         <i class="fa fa-home"></i>
                        </span>
                    </div>
                             <h5>+1 0123456789</h5>
                        </span>
                    </div>
                    <div>
                        </span>
                    </div>
                </div>
                    <textarea rows="8" name="message" placeholder="Message"</pre>
required></textarea>
                    <button type="submit" class="hero-btn red-btn">Send
                    </form>
                </div>
            </div>
    </section>
        To care your health with less time and effort spent and to detect
benefited.
            <i class="fa fa-facebook"></i>
```

#### **Prediction.html**

```
<html>
<head>
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Early Detection of Chronic Kidney Disease using Machine
Learning</title>
link rel="stylesheet" href="static\style.css">
link
href="https://fonts.googleapis.com/css?family=Poppins:100,200,300,400,600,700
&display=swap" rel="stylesheet">
link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/font-awesome/4.7.0/css/font-awesome.min.css">
</head>
</style>
h1 {text-align: center;}
</style>
<h1>Early Detection of Chronic Kidney Disease using Machine Learning</h1>
<br/>
<br/>
<a href="index.html"><i src="images/images.png"></a>
<a href="index.html"><i src="images/images.png"></a>
<a href="index.html"><i lass="fa fa-close" onclick="hideMenu()"></i>

<a href="index.html">href="index.html">REDUTT</a>

<a href="rediction.html">REDUTT</a>

<a href="rediction.html">REDUTT</a>

<a href="rediction.html">REDUTT</a>

<a href="rediction.html">CONTACT</a>

<a href="rediction.html"></a>

<a href="rediction.html"></a>

<a href="rediction.html">REDUTT</a>

<a href="rediction.html">REDUTT</a>

<a href="rediction.html"></a>

<a href="rediction.html"></a>

<a href="rediction.html">REDUTT</a>

<a href="rediction.html"><a href="rediction.html">REDUTT</a>
</a>

<a href="rediction.html"><a href="rediction.html">REDUTT</a>
</a>
</a>
</a>
<a href="rediction.html"><a href="rediction.html"><a href="rediction.html">REDUTT</a>
</a>
</a>
</a>
<a href="rediction.html"><a href="rediction.html"><a href="rediction.html"><a href="rediction.html"><a href="rediction.html"><a href="rediction.html"><a href="rediction.html"><a href="rediction.h
```

```
</nav>
    </section>
        <img src="images/prediction.jpg">
function and specifically slowing down any damage to the kidneys. Newer drugs
We really want to help your health and so losing weight can be a key
component to reducing your risk of progressing with kidney disease. Reducing
calorie intake, which is either smaller portions, less snacking in between
We look to get blood pressure less than 130 systolic, that's the top number.
And less than 80 diastolic, that's the bottom number, on blood pressure
                <span>Stage 1 Normal
                <span>GFR>90mL/min</span>
            </div>
            <div>
                <span>Stage 2 Mild CKD</span>
                <span>GFR=60-89mL/min</span>
            </div>
            <div>
            </div>
                <span>Stage 3B Moderate CKD</span>
                <span>GFR=30-44mL/min</span>
               <span>Stage 4 Severe CKD</span>
```

```
<span>GFR=15-29mL/min</span>
            </div>
             <div>
            </div>
    </div>
    </section>
                                     <h1>Chronic Kidney Disease</h1><h3>Machine
                                </div>
                                <div>
                                     </div>
Random</label>
                                          <input type="number" class="form-</pre>
                                     </div>
Anemia</label>
                                          </select>
                                     </div>
```

```
Artery Disease
                                             <option value = "no">no
</option>
                                        </select>
                                   </div>
                                        <label for="select">Select Pus
                                             <option value = "yes">
yes</option>
                                             <option value = "no">no
</option>
                                        </select>
                                   </div>
                                         <label for="select">Select Red Blood
                                             <option value = "yes">
yes</option>
                                             <option value = "no">no
</option>
                                        </select>
                                   </div>
Mellitus</label>
                                             <option value = "yes">
yes</option>
                                             <option value = "no">no
</option>
                                        </select>
                                   </div>
                                        <label for="select">Select Pedal
Edema</label>
                                             <option value = "yes">
                                        </select>
                                   </div>
href="result.html">PREDICT</a></button>
```

```
</div>
                                </div>
            </form>
          </div>
     </section>
        <h4>About Us</h4>
            <i class="fa fa-linkedin"></i></i>
        </div>
</section>
</script>
</html>
```

### Result.html

```
<html>
<head>
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Early Detection of Chronic Kidney Disease using Machine
Learning</title>
link rel="stylesheet" href="static\style.css">
<link
href="https://fonts.googleapis.com/css?family=Poppins:100,200,300,400,600,700
&display=swap" rel="stylesheet">
<link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/font-awesome/4.7.0/css/font-awesome.min.css">
</head>
</style>
```

```
<h1>Early Detection of Chronic Kidney Disease using Machine Learning</h1>
           <a href="index.html"><img src="images/images.png"></a>
                  <a href="home.html">HOME</a>
                  <a href="about.html">ABOUT</a>
                  <a href="result.html">RESULT</a>
                  <a href="contact.html">CONTACT</a>
              </div>
   </section>
<section>
             <div style="background-color:pink;text-align: center;">
                  </div>
                  <div class="col-md-8 col-sm-7">
                           <h2>Medical Result</h2>
                           Your Condition is normal.
disease.</blockquote>
                      </div>
```

```
</div>
                        </div>
                    </div>
               </div>
          </div>
</style>
    <a href="bookapp.html" class="hero-btn red-btn">BOOKAPPOINTMENT</a>
benefited.
            <i class="fa fa-linkedin"></i></i>
       </div>
        Made <i class="fa fa-heart-o"></i> by PNT2022TMID31803
</section>
<script>
```

## Bookapp.html

```
<title>BOOKING APPOINTMENT</title>
 <fieldset>
    <label>
        <input type="email" name="email" required>
     </label>
   </div>
 </fieldset>
 <fieldset>
     </label>
         <input type="checkbox" name="Morning desired" value="yes">
         <input type="hidden" name="Afternoon desired" value="no">
         <input type="checkbox" name="Afternoon desired" value="yes">
         Afternoon
       </label>
     </div>
    Confirmation requested by
     <label>
     </label>
```

### **OUTPUT SCREENSHOT:**

