PROJECT OBJECTIVE:

Chronic Kidney Disease (CKD) is a major medical problem and can be cured if treated in the early stages. Usually, people are not aware that medical tests we take for different purposes could contain valuable information concerning kidney diseases. Consequently, attributes of various medical tests are investigated to distinguish which attributes may contain helpful information about the disease. The information says that it helps us to measure the severity of the problem and we make use of such information to build a machine learning model that predicts Chronic Kidney Disease

Chronic Kidney Disease prediction is one of the most important issues in healthcare analytics. The most interesting and challenging tasks in day-to-day life is prediction in medical field. 10% of the population worldwide is affected by chronic kidney disease (CKD), and millions die each year because they do not have access to affordable treatment. Chronic kidney Disease can be cured, if treated in the early stages. 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).

LITERATURE SURVEY

S.No	PAPER TITLE	ALGORITHM / TECHNIQUE USED	DESCRIPTION
1.	Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease	Logistic Regression and Classification, Decision tree classifier, Random Forest classifier	This proposed system detects chronic kidney disease using machine learning; They have attained an accuracy of 100% in decision tree classifier, 95.12% in random forest and 98.82% in logistic regression.

2.	Statistical and Data Mining Aspects on Kidney Stones: A Systematic Review and Meta-analysis	Random Forest, Support vector machine, Logistic and NN	They predicted good accuracy with Classification tree and Random Forest (93%) followed by Support Vector Machines (SVM) (91.98%). Logistic and NN has also shown good accuracy results with zero relative absolute error and 100% correctly classified results.
3.	Prediction of chronic kidney disease (CKD) using Data Science	Support Vector Machine, Random Forest, XGBoost, Logistic Regression, Neural networks	This proposed research work is primarily focused on finding the best classification algorithm which can be used for the diagnosis of CKD based on the classification report.
4.	Chronic Kidney Disease Prediction using Machine Learning	Ant Colony Optimization technique and Support Vector Machine (SVM) classifier	This study proposes the use of machine learning techniques for CKD such as Ant Colony Optimization (ACO) technique and Support Vector Machine (SVM) classifier. Final output predicts whether the person is having CKD or not by using minimum number of features.

5.	A Neural Network based Model for Predicting Chronic Kidney Diseases	Artificial Neural Network algorithms	The 14 different properties are analysed and linked to chronic kidney disorder victims and foretold accuracy for a machine learning algorithm named Artificial Neural Network. After analysing the outcomes, it is recognized that the algorithm gives correctness of 96
6.	A Machine Learning Methodology for Diagnosing Chronic Kidney Disease	Logistic regression, Random Forest, Support vector machine, knearest neighbour, Naive Bayes classifier, and Feed Forward Neural Network	A machine learning approach for diagnosing CKD was proposed in this study. An ensemble model that combines logistic regression and random forest with the aid of perceptron was utilized and it was able to attain an average accuracy of 99.83% after ten times of simulation

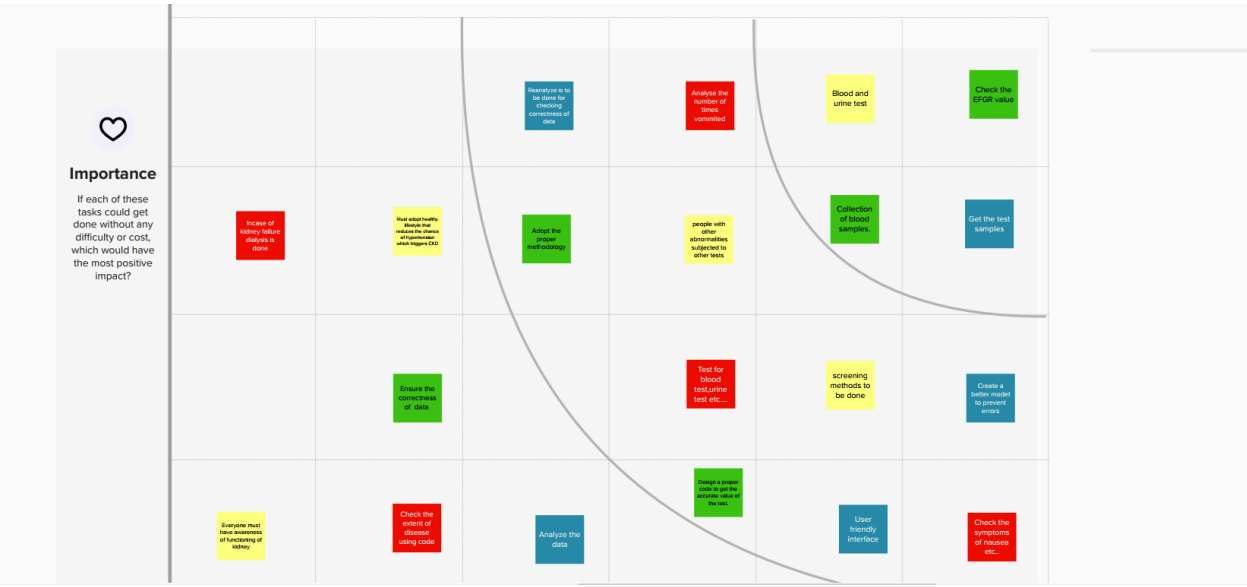
EMPATHY MAP

ild empathy and keep your focus on the user by putting yourself in their shoes.

PROBLEM STATEMENT:

Chronic Kidney Disease is very common, but less than 1 in 10 people with CKD ever require dialysis (artificial kidney treatment) or a kidney transplant.Someone with CKD is at increased risk of heart attack or stroke, especially if they smoke or are overweight.A simple detection technique using machine learning is introduced to detect and predict the disease at an early stage,so that treatment is done at right time without any complexity.

IDEATION

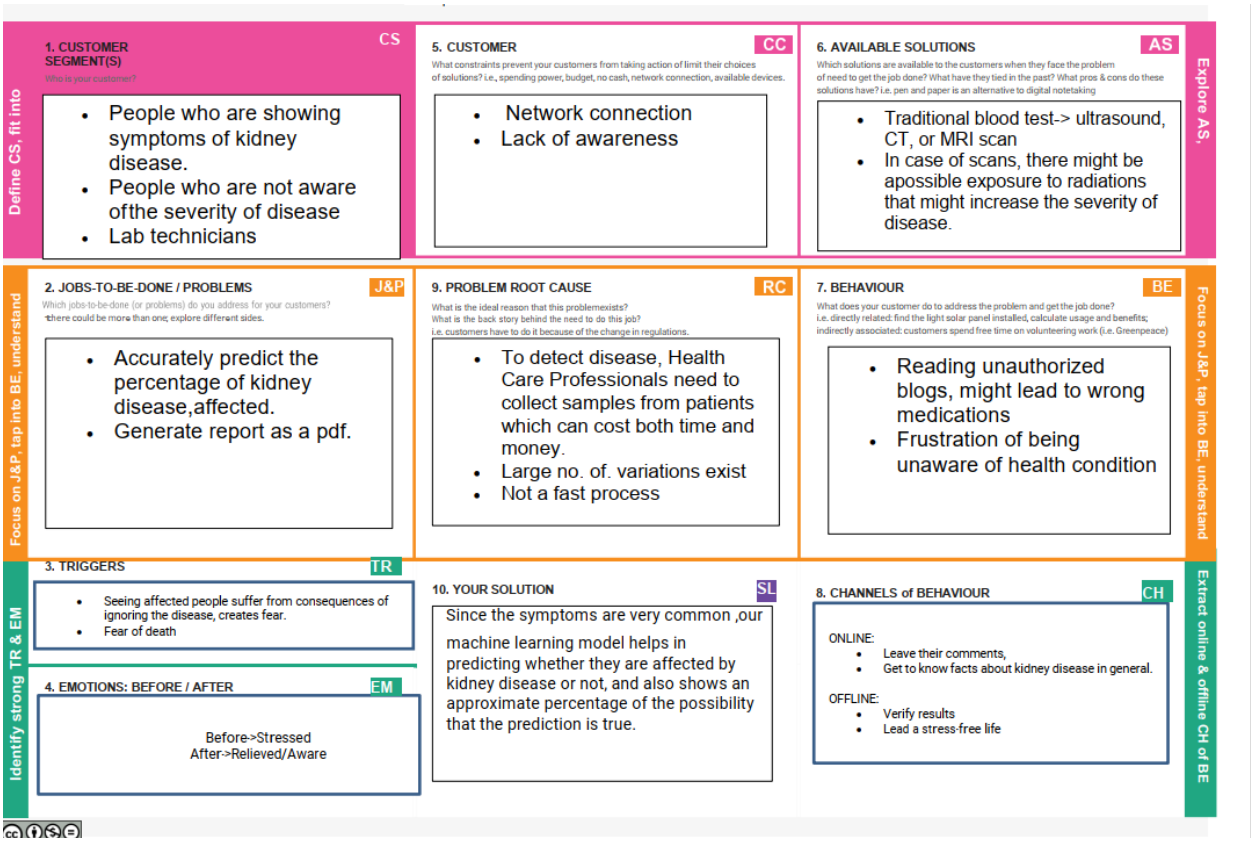


PROPOSED SOLUTION

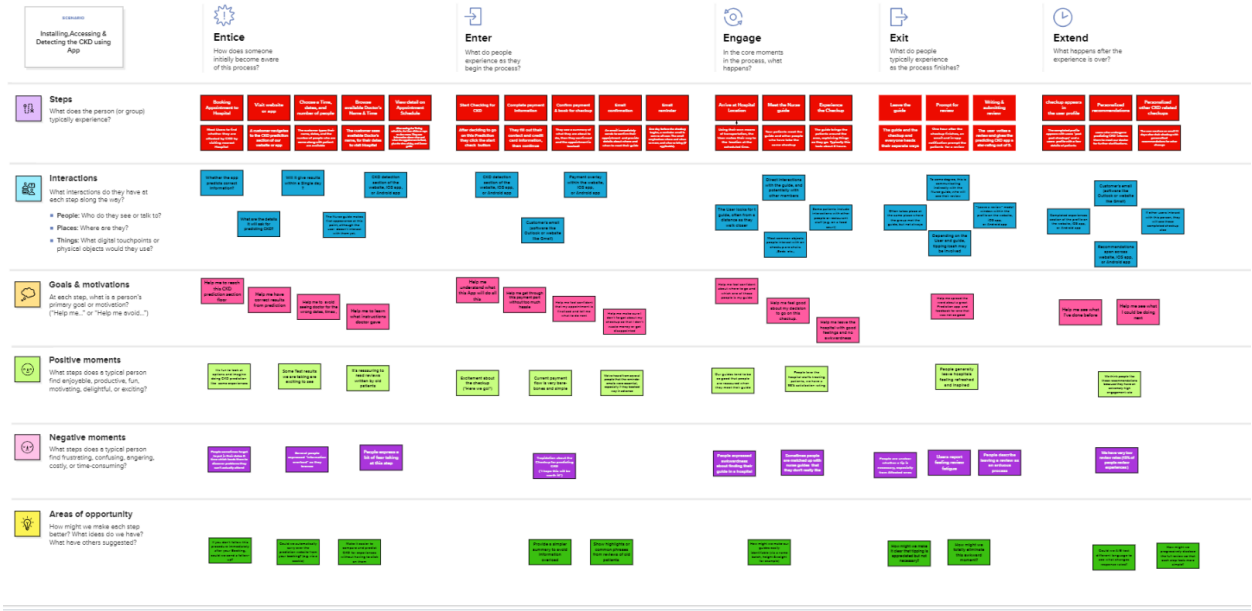
S. No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Early Detection of Chronic Kidney Disease using Machine Learning.
2.	Idea / Solution description	Early detection and cure of Chronic Kidney Disease(CKD) is extremely desirable as it can lead to the prevention of unwanted consequences. Machine learning methods are used to predict the various stages ofCKD using the dataset obtained from the medical records of affected people.
3.	Novelty / Uniqueness	Specifically, we have used the Random Forest and J48 algorithms to obtain a sustainable and practicable model to detect various stages of CKD with comprehensive medical accuracy.

4.	Social Impact / Customer Satisfaction	As the Chronic Kidney Diseases (CKD) is a silent disease, as most sufferers have no symptoms until kidney function drops. Our project can be a huge game changer in a medical field by helping doctors to predict the CKD in a early stage and to the lives of thousands of people.
5.	Business Model (Revenue Model)	This application is recommended to patients in low cost with subscription basis.
6.	Scalability of the Solution	Machine-learning methods can be employed to better analyse nanomaterials and nanoscale biological materials, and aid in the effort to find new materials and the best routes to design nanomaterials optimally.

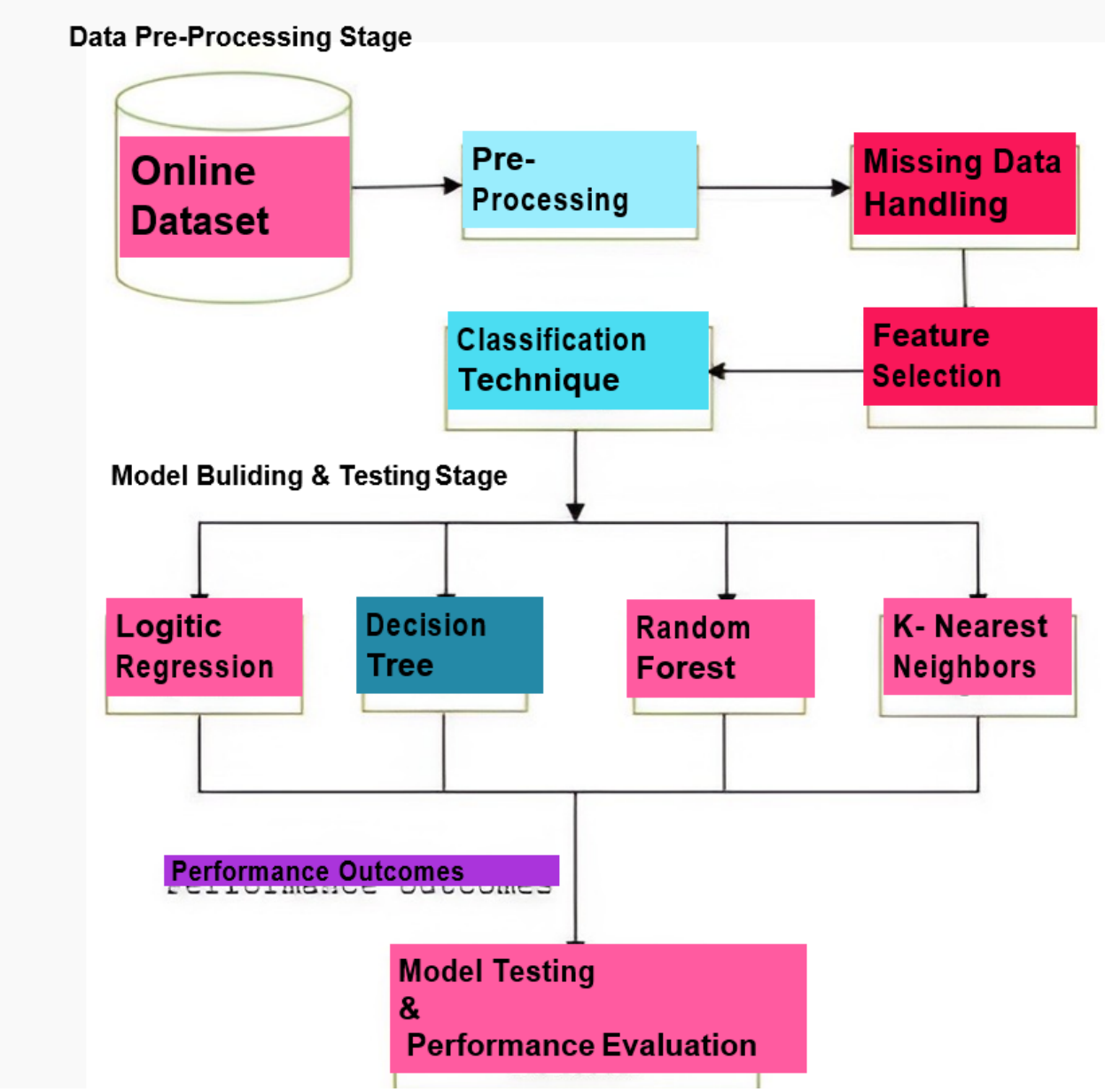
PROBLEM SOLUTION FIT



CUSTOMER JOURNEY MAP



SOLUTION ARCHITECTURE



USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priori
Customer (Web User)	Login	USN-1	As a user, I can login for the application by entering my email	I can access my account / dashboard	High

		USN-2	As a user, I will login, confirmation email once I have registered	I can receive confirmation email	High
		USN-3	As a user, I can login for the application through Mobile number	I can login and access the application	Mediu
	Dashboard	USN-4	As a user, I need to enter my Details	I can get information as per details	High
	Dashboard	USN-5	As a user, I need to enter my Test Details	I can get result based on test details	High
Administrator	Services	USN-6	As a admin I need to provide valid result	I can get a result	High
		USN-7	As a admin I need to provide valid/useful Suggestions	I can get suggestions	Mediu
	Mass Data Process	USN-8	As a admin need to collect all the details and information	I can use it for later period	High
		USN-9	As a admin I need to store all the details and information	I can use it for later period	High
Hospital Administrator	Login	USN-10	As a admin I need to login and access details of customers	I can use for it further next step process	High
	Dashboard	USN-11	As a admin I need to proceed the details with case head	I can use for further next step process	High

FUNCTIONAL REQUIREMENTS

FR NO	Functional Requirement (Epic)	Sub Requirement (Story/Sub-Task)
1	User Registration	User should make registration through the login page
2	User Confirmation	User make sure to confirm the login through sms or mail
3	User Requirements	Create an account,recalculate data using a certain algorithm, and perform other

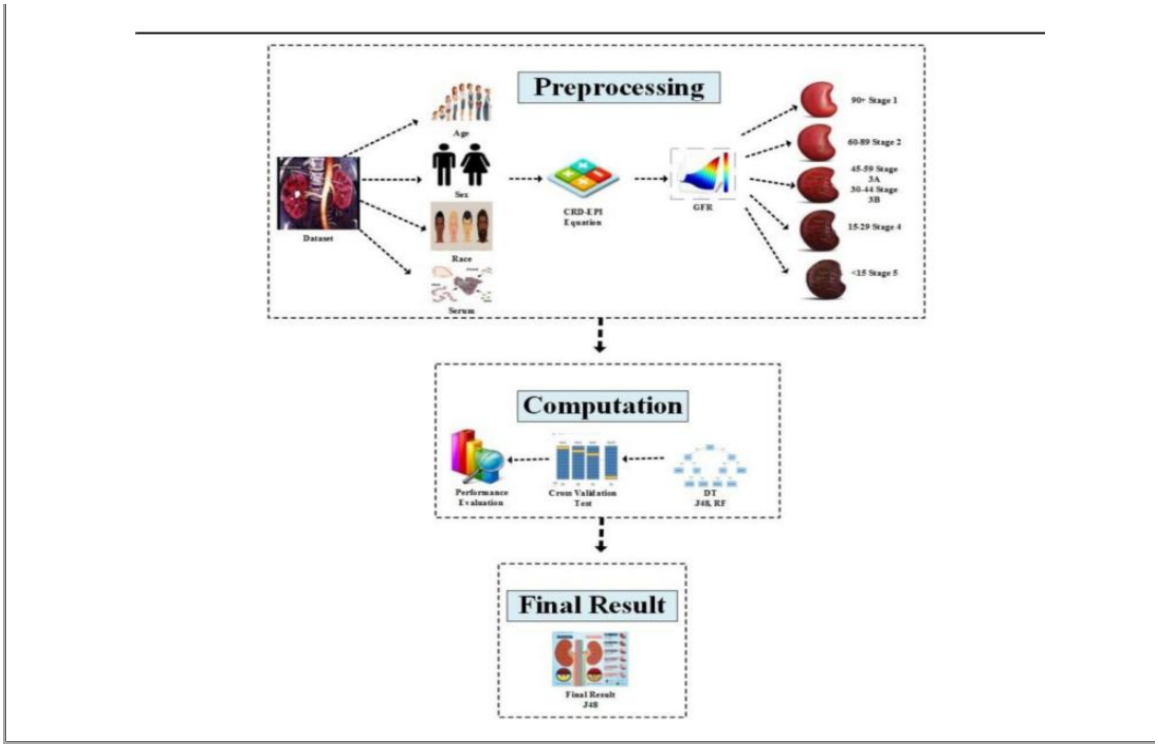
		actions.
4	Business Requirements	An Application allowing patient to identify Chronic kidney disease in around the time of 15 minutes.
5	User Authentication	Make the user to validate credentials through Password
6	User Authorization	Once the server receives a request with authorization, it can verify your credentials and grant youraccess to the resources.

Non-functional Requirements:

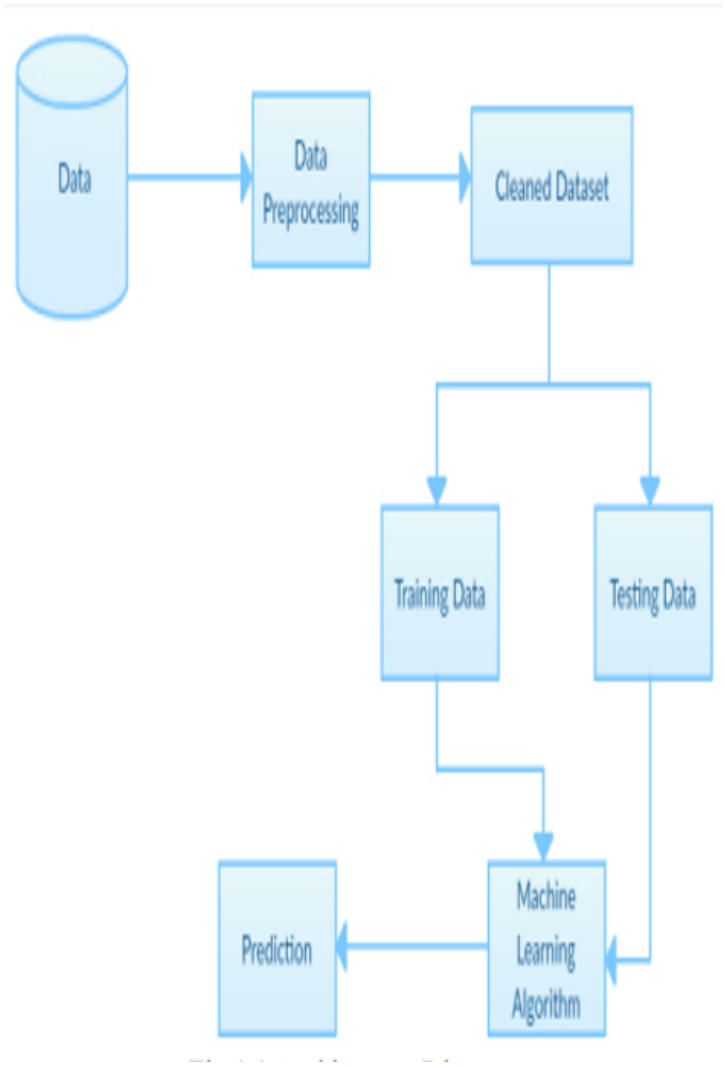
FR No.	Non-Functional Requirement	Description

NFR-1	**Usability**	1. Preventing loss of kidney disease. 2. Delaying oravoiding progression to kidney failure.
NFR-2	**Security**	1. Encrypt your data. 2. Focus the hosting service measure 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. 2. CPU usage. 3. Networkinput/output. 4. Disk input/output.

DATA FLOW DIAGRAM



TECHNOLOGY ARCHITECTURE



MILESTONE ACTIVITY

Activity number	Activity name	Detailed activity description	Assigned to

1	**Preparation Phase**	1. Access the resources (courses) in project dashboard 2. Access the guided project workspace 3. Create GitHub account & collaborate with Project Repository in project workspace 4. Set-up the Laptop / Computers based on the prerequisites for each technology track	
2	**Ideation Phase**		
2.1	Literature survey	Literature survey on the selected project & Information Gathering	G.V.Arun
2.2	Define a problem statement	Prepare the list of problem statements to understand the user needs	M.Gokul shyam
2.3	Empathy Map	Preparation of Empathy Map Canvas to capture the user Pains & Gains	Jeyrenga rajendiran
2.4	Brainstorm & idea prioritization	List the ideas by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance	

PROJECT DESIGN PHASE I

Activity number	Activity name	Detailed activity description	Assigned to
3.1	Proposed Solution	Preparation of proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution	M.Gokul shyam
3.2	Problem Solution Fit	Prepared problem is analysed and make effective solutions for the problem	Muthalagu
3.3	Solution Architecture	Prepare an architecture for solution	G.V.Arun

PROJECT DESIGN PHASE II

Activity number	Activity name	Detailed activity description	Assigned to
4.1	Requirement Analysis	Prepare the Functional Requirement and Non- Functional Document	M.Gokul shyam

4.2	Customer Journey	Preparation of customer journey maps to understand the user interactions & experiences with the application (entry to exit)	Jeyrenga rajendiran

4.3	Data Flow Diagrams	Prepare a Data Flow Diagram for Project use level0 (Industry Standard)	Muthalagu
4.4	Technology Architecture	Prepare Technology Architecture of the solution	G.V.Arun

PROJECT PLANNING PHASE

Activity number	**Activity name**	**Detailed activity description**	**Assigned to**
5.1	Milestones & Tasks	Prepare Milestone & Activity List	M.Gokul shyam
5.2	Sprint Schedules	Prepare Sprint Delivery Plan	M.Gokul shyam

PROJECT DEVELOPMENT PHASE

Activity number	Activity name	Detailed activity description	Assigned to
6.1	Coding & Solutioning	Sprint-1 Delivery: Develop the Code, Test and push it to GitHub.	Muthalagu
6.2	Acceptance Testing	Sprint-2 Delivery: Develop the Code, Test and push it to GitHub. Sprint-3 Delivery: Develop the Code,	M.Gokul shyam G.V.Arun
		Test and push it to GitHub.	
6.3	Performance Testing	Sprint-4 Delivery: Develop the Code, Test and push it to GitHub	Muthalgu Jeyrenga rajendiran

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	New user enters into the System. He/ She can register into the Application by entering user details such as username	2	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran

			and mobile number			
Sprint-2	User Verification	USN-2	The user will receive OTP through SMS.	3	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint-1	Login	USN-3	After Successful registration the user can Log into the application by entering the registered Username and Password	2	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran

Sprint	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Story Points**	**Priority**	**Team Member**
Sprint -1		USN-4	CAPTCHA will be provided to reduce the network traffic.	2	Medium	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -2	Dashboard	USN-5	User can get into the Dashboard only when the Verification Successful. After the user can access the displayed information in the Dashboard	3	Medium	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -3	Data collection	USN-6	Diagnosed result data will be entered by the user.	2	Medium	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran

Sprint -4	Prediction result	USN-7	By the collected data the trained model will predict and display the result.	2	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -4		USN-8	Based on the result the suggestion varies.	2	Low	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -1	Dataset Collection	USN-9	Chronic Kidney Disease dataset identification	2	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran

Sprint	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Story Points**	**Priority**	**Team Members**
Sprint -1	Clean the Dataset	USN-10	The dataset had to be cleaned. Cleaning process includes removing null values, Replacing missing values, segregation of test and train data	3	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -2	Train ML Model in IBM	USN-11	The model will be trained in IBM.	4	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -3	Model Testing	USN-12	The model will be tested using the test data	3	High	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -3	Integration	USN-13	HTML file and python Code Integration	2	Medium	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran

Sprint -4	Deployment	USN-14	The model will be deployed in Cloud	3	Medium	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran
Sprint -4	Further Clarification	USN-15	The problems which are faced by the user while using the application can be clarified	2	Medium	M.Gokul shyam G.V.Arun S.Muthalagu Jeyrengarajendiran

Project Tracker, Velocity & Burndown Chart

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	**11**	6 Days	24 Oct 2022	29 Oct 2022	**11**	19 Nov 2022
Sprint -2	**10**	6 Days	31 Oct 2022	05 Nov 2022	**10**	05 Nov 2022
Sprint -3	**7**	6 Days	07 Nov 2022	12 Nov 2022	**7**	12 Nov 2022
Sprint -4	**9**	6 Days	14 Nov 2022	19 Nov 2022	**9**	19 Nov 2022

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 per day).

Sprint 1 AV = Sprint duration/velocity = 11/6 = 1.83 Sprint 2 AV = Sprint duration/velocity = 10/6 = 1.67 Sprint 3 AV = Sprint duration/velocity = 7/6= 1.16 Sprint 4 AV = Sprint duration/velocity = 9/6 = 1.5

SPRINT 1
https://github.com/IBM-EPBL/IBM-Project-31385-1660199997/blob/36b161209c66ca0b300b7be7af5f9107068834ee/Project%20Devlopment%20Phase/Sprint%201/chronic%20kidney%20disease-%20cleaning%20the%20dataset.ipynb

SPRINT 2
https://github.com/IBM-EPBL/IBM-Project-31385-1660199997/blob/36b161209c66ca0b300b7be7af5f9107068834ee/Project%20Devlopment%20Phase/Sprint%202/Model%20Building.ipynb

SPRINT 3
https://github.com/IBM-EPBL/IBM-Project-31385-1660199997/blob/36b161209c66ca0b300b7be7af5f9107068834ee/Project%20Devlopment%20Phase/Sprint%203/Local%20Deployment.ipynb

SPRINT 4
https://github.com/IBM-EPBL/IBM-Project-31385-1660199997/blob/36b161209c66ca0b300b7be7af5f9107068834ee/Project%20Devlopment%20Phase/Sprint%204/Flask%20Connectivity.ipynb

CONCLUSION

Consequently, the indirect agreement with the objectives of the project, the CRISP-DM methodology was adapted to the context of the problem so that the different logically organized stages were taken; data collection, pre-processing, learning, evaluation, and selection, which allowed the construction of a model capable of classifying the possibility of a diagnosis of CKD with an accuracy of 93%. As evidenced in the results, the decision forests algorithm has obtained quite optimal results, where

predictions of 93% have been obtained. Data preparation is a fundamental step in the process and the absolute precision of the model is directly dependent on this phase. Thanks to the models, we can see how changing the characteristics affects the search for the target value with a simple change of column selection or improvements in the data. The innovation of this work results from the design adjusted to the environment of the health system in Iraq and the pathology of the CKD in our country, with a methodology adapted to the case study and a production architecture proposal for the model with Microsoft Azure tools of form that allows satisfying in the future the scalability of the solution. Furthermore, this methodology could apply to clinical data of other diseases and pathologies inaccurate medical diagnoses. The development of this project allowed the author to acquire more excellent knowledge through both practical and theoretical work about current techniques for the development of machine learning.

GITHUB LINK:https://github.com/IBM-EPBL/IBM-Project-31385-1660199997