

## Project Initialization and Planning Phase

Date	16 July 2024
Team ID	SWTID1720075199
Project Title	Early Prediction of Chronic Kidney Disease Using Machine Learning
Maximum Marks	3 Marks

### Project Proposal

This project proposal outlines a solution to address early detection of CKD(Chronic Kidney Disease). With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	To develop a machine learning model that accurately predicts the likelihood of chronic kidney disease (CKD) in patients at an early stage
Scope	The project scope is limited to providing diagnostic assistance for enhancing early intervention and detection of chronic kidney disease (CKD) in patients.
Problem Statement	
Description	Chronic kidney disease (CKD) occurs when the kidneys become damaged over a prolonged period (at least three months) and struggle to perform their vital functions. Patients with CKD, depending on the severity of their condition, often require dialysis until they can receive a kidney transplant. The task at hand is to identify methods for detecting the onset of CKD at an early stage.
Impact	<p>Detecting chronic kidney disease (CKD) at an early stage offers numerous benefits, including:</p> <ol style="list-style-type: none"> <li><b>Improved Patient Outcomes:</b> <ul style="list-style-type: none"> <li>Early detection allows for timely medical intervention, which can slow the progression of the disease and improve the overall prognosis.</li> </ul> </li> <li><b>Enhanced Quality of Life:</b></li> </ol>

	<ul style="list-style-type: none"> <li>○ Early treatment can help manage symptoms more effectively, enabling patients to maintain a better quality of life.</li> </ul> <p>3. <b>Prevention of Complications:</b></p> <ul style="list-style-type: none"> <li>○ Early diagnosis helps in preventing or mitigating complications associated with CKD, such as cardiovascular disease, anemia, and bone disorders.</li> </ul> <p>4. <b>Cost Savings:</b></p> <ul style="list-style-type: none"> <li>○ Early treatment can reduce the need for expensive treatments like dialysis and kidney transplants, leading to significant cost savings for both patients and healthcare systems.</li> </ul> <p>5. <b>Better Treatment Options:</b></p> <ul style="list-style-type: none"> <li>○ Patients diagnosed early have more treatment options available, including lifestyle changes, medications, and dietary adjustments, which can help manage the disease more effectively.</li> </ul> <p>6. <b>Increased Survival Rates:</b></p> <ul style="list-style-type: none"> <li>○ Early intervention can significantly improve survival rates by preventing the rapid decline of kidney function and reducing the risk of related health issues.</li> </ul>
<b>Proposed Solution</b>	
Approach	<p>1. <b>Preprocessing:</b></p> <ul style="list-style-type: none"> <li>• <b>Data Collection:</b> Gather a comprehensive dataset comprising clinical and demographic information from reliable sources.</li> <li>• <b>Data Transformation:</b> Normalize and standardize the data to ensure consistency and improve the performance of the machine learning models. Encode categorical variables and create derived features if necessary.</li> </ul> <p>2. <b>Data Cleaning:</b></p> <ul style="list-style-type: none"> <li>• <b>Handling Missing Values:</b> Identify and address missing values in the dataset through imputation techniques or by removing records with excessive missing data.</li> <li>• <b>Outlier Detection:</b> Detect and manage outliers to minimize their impact on the model's performance.</li> </ul>

	<ul style="list-style-type: none"> <li>• <b>Data Consistency:</b> Ensure consistency across the dataset by correcting any discrepancies and ensuring uniformity in data formats.</li> </ul> <p><b>3. Classification Problem:</b></p> <ul style="list-style-type: none"> <li>• During the analysis, we realized that the task at hand is a classification problem, where the goal is to predict whether a patient has CKD or not based on the input features.</li> </ul> <p><b>4. Model Development:</b></p> <ul style="list-style-type: none"> <li>• <b>Decision Tree:</b> <ul style="list-style-type: none"> <li>◦ <b>Initial Model:</b> Developed a decision tree model to classify CKD. However, it yielded less accuracy, indicating the need for a more robust model.</li> </ul> </li> <li>• <b>Random Forest:</b> <ul style="list-style-type: none"> <li>◦ <b>Improved Model:</b> Implemented a random forest model, which provided better performance than the decision tree but still did not meet the desired accuracy levels.</li> </ul> </li> <li>• <b>XGBoost:</b> <ul style="list-style-type: none"> <li>◦ <b>Optimal Model:</b> Finally, employed XGBoost, an advanced gradient boosting algorithm, which achieved the highest accuracy among the models tested. XGBoost's ability to handle missing data, prevent overfitting, and manage complex interactions between features made it the most effective model for this project.</li> </ul> </li> </ul>
Key Features	<ul style="list-style-type: none"> <li>• <b>User Interface:</b> <ul style="list-style-type: none"> <li>• A simple and intuitive interface for healthcare providers to input patient data, including demographic information, medical history, lab test results, and lifestyle factors.</li> </ul> </li> <li>• <b>Predictive Model:</b> <ul style="list-style-type: none"> <li>• The app integrates the XGBoost model, which is trained on a comprehensive dataset to predict the risk of CKD with high accuracy.</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>• <b>Results Interpretation:</b> <ul style="list-style-type: none"> <li>• Provides clear and interpretable results</li> </ul> </li> </ul>
--	--

## Resource Requirements

Resource Type	Description	Specification/Allocation
<b>Hardware</b>		
Computing Resources	CPU/GPU specifications, number of cores	2 x NVIDIA V100 GPUs
Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	1 TB SSD
<b>Software</b>		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, matplotlib, seaborn, xgboost
Development Environment	IDE, version control	Google Collab, Git
<b>Data</b>		
Data	Source, size, format	SmartInternz Dataset, 400 data entries, csv file