Project Initialization and Planning Phase

Date	07 July 2024
Team ID	SWTID1720190579
Project Title	Early Prediction of Chronic Kidney Disease Using Machine Learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to address a specific problem. 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.

The proposed work in this project is to create a machine-learning model that can identify CKD patients at the early time of the disease using different patient's characteristics. Thus, by using appropriate machine learning algorithms, it aims to recognize the algorithm with the highest accuracy in predicting CKD. The intended objective is to improve the patient's status, which would be achieved by allowing the provision of timely medical intervention to halt the progress of an illness. The outlined solution includes data accretion and data preparation, modeling, assessment, and model identification for efficient CKD early-stage prediction.

Project Overview	
Objective	The aim of this project is to construct a precise and fast artificial neural network model to predict CKD at its initial stage or possibly even before the patient is aware of his or her condition. From the patients' descriptions, demographic and clinical characteristics, as well as laboratory data, this model would like to predict those who could develop CKD in its early stages. Using the scores of patients early helps in the early and timely interventions which can also be beneficial in improving the outcome of the patient's treatment, and also the resource utilization in the large health care systems. The project will be centered on the employment of numerous machine learning algorithms, fine-tuning of the models, and testing their efficiency with actual data to achieve stability and accuracy in the final result.
Scope	This work suggests using medical data to create an ML system for CKD classification. With a focus on improving diagnostic performance and decision support for CKD, subprocesses encompass

	data collecting, data cleansing and transformation, variable selection, model creation, and application to the healthcare domain.
Problem Statement	
Description	CKD is a gradual degenerative disease and deteriorates over time, and if not treated or DIAGNOSED EARLY is lethal and results in kidney failure. When it comes to early CKD prediction by using machine learning, it entails the analysis of large amounts of patient's data to produce a pattern or risk factors. CKD risk models use this information: demographic, medical history data, laboratory, and other results to predict disease progression. These models help the doctors in precise diagnosis, management of the disease process and its aggressive course being controlled. This approach improves on the quality of patient care delivery and decreases on the costs of health care delivery since the main focus is on the prevention and early management of the disease.
Impact	This project's model for diagnosing chronic kidney diseases will see a positive shift as an improvement in provision of care so as to address an increasing patient base with more accuracy and efficiency as an outcome. Worldwide its mission is to reduce health costs or enhance the quality of patient's life; such cause sets a model example of how AI can be utilized in disease management.
Proposed Solution	
Approach	In the case of the diagnosis project of CKD, specifications are about collecting clinical data of patients with CKD, data pre-processing to include imputing missing values and normalizing data, understanding the features distribution and distribution-free methods, choosing proper machine learning algorithms for CKD diagnosis, benchmarking the indices on clinical utility of the proposed solution.
Key Features	Major Solution Ingredients/main components that defines this work Some of the techniques used include data acquisition from patients' records, incorporating the test results, data pre-processing which may include handling of missing values in a dataset and data normalization, feature selection which may be limited to creatinine level or demography data, model selection which may encompass decision trees and support vector machine (SVM). The specific objective of the solution is to achieve high diagnostic accuracy in relation to the selected image dataset.

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	AMD Ryzen 7 5800H with Radeon Graphics
Memory	RAM specifications	16 GB
Storage	Disk space for data, models, and logs	1 TB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, collections, mathplotlib, seaborn, missingno, pickle
Development Environment	IDE, version control	Jupyter Notebook, Git, SPYDER
Data		
Data	Source, size, format	e.g., Kaggle dataset, 10,000 images