

Project Initialization and Planning Phase

Date	30 th June 2024
Team ID	SWTID1720090815
Project Title	Early Prediction of Chronic Kidney Disease
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

Project Overview	
Objective	Develop and deploy a high accuracy model that will help predict the risk of chronic kidney disease
Scope	The Project aims to predict the Chronic Kidney Disease (CKD) in patients at the primary level without complex procedures.
Problem Statement	
Description	As someone concerned about my health, I worry about chronic kidney disease (CKD) because it often goes undetected until it's advanced. Current tests are expensive and invasive, making it hard to get regular checkups.
Impact	An early warning system would be ideal, allowing me to monitor my health and proactively manage my risk factors before serious damage occurs. Ideally, this system would use data I already have access to, so I wouldn't need additional tests.
Proposed Solution	
Approach	Acquiring a credible dataset, and performing data pre-processing techniques to address missing data, imbalance data, and messy data. In the case of limited data records, data augmentation techniques can be utilized to create a synthetic data, which is later merged with the real-world data. The model will be trained on this data while undergoing parameter tuning and cross-validation training. Four models will be tested and the model with the best recall and accuracy will be used.
Key Features	The proposed solution follows several techniques to handle model

	overfitting, and optimization techniques such as GridSearchCV.
--	--

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	NVIDIA MX250 GPUs multi-core processor (4 cores)
Memory	RAM specifications	16 GB
Storage	Disk space for data, models, and logs	10GB HDD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, seaborn, matplotlib, imblearn, pickle, copulas, and xgboost.
Development Environment	IDE, version control	VSCode, Jupyter Notebook, GitHub
Data		
Data	Source, size, format	UC Irvine machine learning repository, 400 Data records, .csv file