

Project Initialization and Planning Phase

Date	11 June 2024
Team ID	SWTID1749620488
Project Title	Chronic Kidney Disease Detection
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal outlines a solution to detect Chronic Kidney Disease (CKD) using machine learning techniques. 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 detects chronic kidney disease from clinical data and assists early diagnosis.
Scope	The project involves data cleaning, visualization, model training, evaluation, and deployment. It will focus on supervised learning techniques using structured health records to predict CKD presence.
Problem Statement	
Description	Chronic Kidney Disease is a silent, progressive condition that often goes undiagnosed until advanced stages. Early prediction is crucial for effective treatment. However, manual diagnosis is time-consuming and subject to human error.
Impact	An automated ML-based system can help doctors and clinics in early screening, saving lives and reducing treatment costs.
Proposed Solution	
Approach	The project uses data preprocessing, label encoding, and model training with Logistic Regression, Gradient Boosting, Decision Tree, and Random Forest. The best-performing models are evaluated using classification metrics and then saved using pickle.

Key Features	<p>Robust Preprocessing: Handles missing values using mean/mode, corrects inconsistent labels, and converts datatypes.</p> <p>Multi-Model Training: Trains and compares Logistic Regression, Gradient Boosting, Decision Tree, and Random Forest classifiers.</p> <p>Performance Evaluation: Uses accuracy score, confusion matrix, and classification report for model assessment.</p> <p>Model Persistence: Saves trained models as .pkl files using pickle for reuse without retraining.</p>
--------------	---

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications	1 x NVIDIA T4 GPU or CPU (Colab)
Memory	RAM	8 GB
Storage	Disk space for data, models, and logs	1 GB (approx. for dataset + models)
Software		
Frameworks	Python frameworks	Jupyter Notebook (Google Colab)
Libraries	Additional libraries	scikit-learn, pandas, numpy, seaborn, matplotlib, pickle, missingno
Development Environment	IDE, version control	Google Colab, Git
Data		
Data	Source, size, format	Kaggle / UCI dataset, CSV, ~400 records