

Early Prediction of Chronic Disease Using Machine Learning:

Milestone 1: Project Initialization and Planning Phase

Activity 1: Define Problem Statement

Chronic kidney disease (CKD) is a significant health issue that affects a substantial portion of the global population. Early and accurate detection of CKD is crucial for minimizing health complications and improving patient outcomes. Machine learning techniques have shown promise in predicting and diagnosing CKD, enabling timely interventions and personalized treatment plans. The problem statement is to develop a machine learning model that can accurately predict the risk of developing CKD at an early stage. The model should analyse patient data, including demographic information, medical history, and laboratory test results, to identify individuals at high risk of developing CKD. The goal is to provide healthcare providers with a tool that can aid in early intervention, personalized treatment planning, and improved management of CKD patients. By addressing this problem, we aim to:

- Enable healthcare providers to identify individuals at risk of developing CKD.
- Facilitate early detection of CKD to initiate timely treatment.
- Personalize treatment plans based on individual risk factors.
- Improve patient outcomes and reduce the burden of CKD on healthcare systems

Refer template: [click here](#)

Activity 2: Project Proposal (Proposed Solution)

The proposed project, "Early Prediction of Chronic Disease Using Machine Learning," aims to leverage machine learning for more accurate and early prediction of chronic diseases. We plan to create machine learning classifier model that will be able to predict whether a person has high chances of getting a CKD or not. The relevant features like blood sugar, albumin, creatine etc.. will be fed to the model based on which it will take the decision. Then we will deploy this model on a user-friendly webpage using Flask framework of python. Refer template: [click here](#)

Activity 3: Initial Project Planning

Initial Project Planning involves outlining key objectives, defining scope, and identifying stakeholders for a chronic disease prediction system. The plan is pretty straight forward. First we will collect the dataset and then clean and preprocess it, feed it to models, select the best model ,fine tune it and deploy it using flask.

Refer template: [click here](#)

Milestone 2: Data Collection and Preprocessing Phase

Reference notebooks: [NB_v1](#), [NB_V2](#)

Activity 1: Data Collection Plan

The dataset for "Early Prediction of Chronic Disease Using Machine Learning" was provided by The SmartInternz project workspace.

Refer template: [Click here](#)

Activity 2: Data Quality Report

The dataset for "Early Prediction of Chronic Disease Using Machine Learning" was thoroughly checked for quality, completeness, and relevance. All features were relevant and related to the actual problem statement except ID. It had a lot of missing values, some corrupted values and had non-uniform datatypes , which is taken care of while preprocessing.

Refer template: [Click here](#)

Activity 3: Data Exploration and Preprocessing

Data Exploration involves analyzing the chronic disease dataset to understand patterns, distributions, and outliers. Distribution plots, box plots, data description, correlation map were made to understand the nature of data and identify any outliers, missing values and interdependencies. First the garbage values were corrected, missing values filled and outliers replaced. This data was later scaled just before training.

Refer template: [Click here](#)

Milestone 3: Model Development Phase

Activity 1: Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features for the chronic disease prediction model. Used ch2 method to get top ten features that are highly correlated to target variable. The model built on this didn't give promising output. Later include most features and built a model and the model predictions were accurate and true.

Refer template: [Click here](#)

Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing specific machine learning models for chronic disease prediction. The data was trained and tested on Logistic regression, SVM, Random Forest and Decision tree. Random forest resulted in overfitting, even after using cross-validation. Decision tree's accuracy was not satisfying. Same with SVM. Logistic regression gave promising output. So, Logistic regression was chosen.

Refer template: [Click here](#)

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code employs selected algorithms on the chronic disease dataset. The logistic regression model was trained and tested.

Refer template: [click here](#)

Refer notebook: [click here](#)

Milestone 4: Model Optimization and Tuning Phase

Activity 1: Hyperparameter Tuning Documentation

The selected model will be fine-tuned for optimal performance. Used gridSearchCV to get the best parameters and retrained the model with those parameters. Then used GridSearchCv again with prime attribute to be tested as accuracy, and used the new model to make sure it doesn't overfit. The results were satisfactory.

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models. The performance metrics are attached to the link.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification articulates the rationale for choosing the ultimate model. We did finalise the logistic regression model that gave about 97% accuracy with cv=10. We manually tested it and it gave satisfying output.

Refer notebook: [Click here](#)

REFER TEMPLATE :[CLICK HERE](#)

Milestone 5: Project Files Submission and Documentation

For project file submission in Github, Kindly click the link

[GitHub Repo](#)

Milestone 6: Project Demonstration

Here is the link for the Demo Video.: [Demo link](#)

Flask Deployment: [Deploy in your own local server](#)

Team Members: (VIT)

- Abhinav Bansal -22BAI1074
- Ratish R A -22BAI1106
- Manoj G -22BCE3839
- Mayank Raj -22BAI1118