**Phase-2 Submission Template**

**Student Name:** P.PRAJIN

**Register Number:** 422223243044

**Institution:** SURYA GROUP OF INSTITUTIONS

**Department:** B.Tech/ARTIFICIAL INTELLIGENCE & DATA SCIENCE

**Date of Submission:** 13/05/2025

**Github Repository Link**[**https://github.com/praju1223/prajin-nann-mudhalvan**](https://github.com/praju1223/prajin-nann-mudhalvan)

# Problem Statement:

Many diseases are detected at advanced stages, reducing treatment effectiveness. Manual diagnosis can lead to errors, affecting patient outcomes Traditional methods struggle to predict disease onset and progression. Large amounts of patient data can be difficult to analyze and interpret. Developing tailored treatment plans can be complex and time-consuming**.**

# Project Objectives :

Create accurate AI-powered models to predict disease onset and progression. Enhance disease diagnosis accuracy using machine learning algorithms. Develop tailored treatment plans based on individual patient characteristics. Identify high-risk patients and enable early intervention to reduce disease burden. Enhance patient care and outcomes by leveraging predictive analytics and AI insights**.**

1. **Flowchart of the Project Workflow**

# 

# Data Description.

Age, gender, and other patient demographics . Previous diagnoses, treatments, and health conditions . Vital signs, lab results, and other clinical measurements . Genetic information and biomarkers . Patient lifestyle habits, such as diet and exercise .

# 5. Data Preprocessing

Remove errors, inconsistencies, and missing values. Convert data into suitable formats for analysis. Normalize data to ensure consistent scales. Impute or remove missing values. Encode categorical variables for model compatibility

# 6.Exploratory Data Analysis (EDA)

Calculate means, medians, and standard deviations. Use plots to identify trends and patterns. Examine data distributions and outliers. Identify relationships between variables. Discover insights and trends in patient data.

# 7. Feature Engineering

Calculate means, medians, and standard deviations. Use plots to identify trends and patterns. Examine data distributions and outliers. Identify relationships between variables. Discover insights and trends in patient data.

# 8. Model Building

Choose suitable machine learning algorithms. Train models using patient data. Optimize model parameters for best performance. Assess model accuracy and effectiveness. Refine models based on evaluation results.

# 9. Visualization of Results & Model Insights

Visualize model accuracy, precision, and recall. Display predicted disease risks and probabilities. Show which features contribute most to predictions . Visualize individual patient data and predictions. Extract actionable insights from model results.

# 10. Tools and Technologies Used

TensorFlow, PyTorch, or scikit-learn. Pandas, NumPy, and Matplotlib. Relational databases (e.g., MySQL) or NoSQL databases (e.g., MongoDB). Python, R, or SQL. Tableau, Power BI, or D3.js.

# 11. Team Members and Contributions

Data cleaning :**P.PRAJIN**

EDA : **S.NIRMAL KUMAR**

Feature engineering :**A.NAVIN**

Documentation and reporting : **S.NIRMAL KUMAR**