Fetal Health Classification Using Machine Learning on Cardiotocogram Data

1. Team Members:

Dipen Patel	2025ICS-008
Harsh Wadhwani	2025MPCS-003
Kushagra Singh	2025ICS-007

2. Dataset Details:

Name: Fetal Health Classification Dataset
Source: https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification
Description:
The dataset contains **2,126 records** from **Cardiotocogram (CTG)** exams, each classified by expert obstetricians into **three categories** *Normal*, *Suspect*, and *Pathological*. It includes **21 features** related to fetal heart rate (FHR) and uterine contractions that help assess fetal well-being.

3. Problem Statement:

This project aims to **develop machine learning models** capable of accurately classifying fetal health into three categories **Normal**, **Suspect**, and **Pathological** using CTG-derived features. Automating this classification can assist healthcare professionals in **early risk detection**, improving decision-making and contributing to the global goal of **reducing maternal and child mortality** as outlined in the **United Nations Sustainable Development Goals (SDGs)**.

4. Reason for Choosing the Dataset/Problem:

The project addresses an important healthcare challenge — reducing child and maternal mortality rates. Using CTG data, healthcare professionals can monitor fetal well-being more accurately and prevent complications during pregnancy. The dataset is reliable, well-documented, and highly suitable for supervised learning tasks in medical research.

Fetal health monitoring is a critical aspect of maternal care, as early detection of abnormalities can prevent complications during pregnancy and childbirth. Despite medical advancements, **child and maternal mortality** remain major global health challenges, especially in low-resource settings.

Cardiotocography (CTG), a non-invasive technique that records fetal heart rate (FHR) and uterine contractions, provides valuable insights into fetal well-being. However, interpreting CTG data manually can be subjective and time-consuming, often leading to inconsistent diagnoses.

5. Objectives:

Data Preprocessing: Clean, normalize, and prepare data for model training to ensure consistency and model accuracy.

Exploratory Data Analysis (EDA): Identify trends, correlations using correlation heatmaps and feature selection technique, and Visualize relationships between features and target classes for better interpretability

Model Building: Implement and compare multiple models **Logistic Regression**, **Decision Tree**, **Random Forest**, **K-Nearest Neighbors (KNN)**, **Support Vector Machine (SVM)**, **Multi-Layer Perceptron (MLP)**, **Gradient Boosting**.

Hyperparameter Tuning: Optimize models using Grid Search and Cross-Validation for best accuracy.

Evaluation: Assess models using **Accuracy**, **Precision**, **Recall**, **F1-Score**, and **Confusion Matrix** to identify the most effective classifier.

6. Tools & Libraries:

Python, NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, Jupyter Notebook.