

# Problem Statement Worksheet (Hypothesis Formation)

Can we build a model that intakes features extracted from Cardiotocograms (CTGs) and classifies fetal health with an accuracy of 80% or higher by the end of the month?

H

## 1 Context

Child mortality rate is a key indicator of human progress that is closely monitored by the UN and other world health organizations. The majority of deaths in children under 5 are preventable. CTGs are a simple and cost effective way to assess fetal health. A model that could accurately classify CTG results would greatly aid medical professionals in identifying at risk fetuses, and help improving the under-5 mortality rate.

## 2 Criteria for success

Build a model that demonstrates at least 80% accuracy on test data

Identify which features were most useful in classification.

## 3 Scope of solution space

**Deliverables:** Final model with metrics, Report detailing the process of building the model as well as any interesting findings from EDA, and a presentation.

**Timeline:** Complete by the end of the month

## 4 Constraints within solution space

The dataset contains 2126 observations each with 21 features.

There are 3 classes: Normal, Suspect, and Pathological

The classes are imbalanced: 78% of the observations are normal, 14% are suspect, and 8% are pathological.

## 5 Stakeholders to provide key insight

The hospital administrator

The medical director

The head of Labor and Delivery

## 6 Key data sources

The dataset can be downloaded at:

<https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification>

H

D

E

I

P