

# Capstone Three: Project Proposal

## Problem Identification

### Problem Statement

A major healthcare facility is wanting to predict fetal health outcomes accurately using Cardiotocograms (CTGs) data to prevent or decrease child and maternal mortality.

### Context

A decrease in the child mortality rate is one of the United Nations' Sustainable Development Goals. Parallel to the notion of child mortality is of course maternal mortality, which accounts for 295 000 deaths during and following pregnancy and childbirth (as of 2017). The vast majority of these deaths (94%) occurred in low-resource settings, and most could have been prevented.

Cardiotocogram (CTG) is the most widely used in the clinical routine evaluation to detect fetal state and has enabled clinical practitioners to detect signs of fetal compromise at an early stage. It provides information on uterine respiration and fetal heart rate, which can be used to determine whether the fetus is healthy, suspect, or pathological.

Having a predictive model would be a valuable tool for the healthcare industry. This healthcare facility wants to make an impact by being able to predict fetal health outcomes which would in the end prevent or decrease child and maternal mortality.

### Criteria for Success

A healthcare facility wants to predict fetal health outcomes using the best features of CTG data. Accurate classification of fetal outcomes would lead to timely intervention, especially for high-risk pregnancies.

### Scope of Solution Space

The healthcare facility wants to predict fetal health outcomes accurately. This predictive model not only would classify fetal health outcomes but also prevent or decrease child and maternal mortality rates.

### Constraints within Solution Space

The dataset is from Kaggle and contains 2,126 rows of 22 features extracted from Cardiotocogram (CTG) exams, which were then classified by three expert obstetricians into 3 classes: Normal, Suspect, and Pathological.

The model that will be created has to work within this dataset, as no additional data is obtainable.

### **Stakeholders to provide Key Insight**

Healthcare facility upper management and healthcare providers.

### **Key data sources**

The source of data is Kaggle which has 1 data file and a usability score of 10.

This dataset has 22 feature variables and 2129 records.

### **Problem Method and Solution**

- Identification of the problem:
  - Write a problem statement based on the client's needs.
- Data wrangling:
  - Collect, organize, and clean the dataset: data acquisition, data shaping, and data quality.
  - Tidy data, reshaping data by sorting, reindexing, renaming, subsets rows, and columns, using a query, summarizing data, handling missing data, making new columns, combining data sets, grouping data, and plotting data.
- Perform exploratory data analysis:
  - Assess data quality, plot different combinations of variables, data visualization (point plots, scatter plots, bar plots, histogram, or line plots) depending on what needs to be communicated in the plot, profile the data, and explore the data across many dimensions.
- Preprocess the data by standardizing and training the dataset:
  - Imputing missing values, transforming the values, encoding categorical variables, train-test split method, measuring metric performance, cross-validation of data
- Modeling the dataset by training the model to make predictive insights:
  - Selection of the best algorithm for the problem, performing inference for modeling behavior.
- Documentation of the work done:
  - Storytelling using a slide deck, and written report.
- Deliverables:
  - Writing codes, writing project report summary with results, and presentation to stakeholders using slide deck format.