STATEMENT OF WORK (V1)

MARCOS BITTENCOURT

NOVEMBER 6, 2020

PARTH CHOTAI - 100802846

Table of Contents

1.1 Problem Definition

Reduction of child mortality is reflected in several of the United Nations' Sustainable Development Goals and is a key indicator of human progress. The UN expects that by 2030, countries end preventable deaths of new-borns and children under 5 years of age, with all countries aiming to reduce under-5 mortality to at least as low as 25 per 1,000 live births.

Deaths during and following pregnancy and childbirth is 295,000 (as of 2017). Most of these deaths (94%) occurred in low-resource settings, and most could have been prevented.

1.2 Objective / Rationale Statement

Cardiotocograms (CTGs) are a simple and cost accessible option to assess fetal health. This allows healthcare professionals to act in order to prevent child and maternal mortality.

We will predict **fetal_health** from CTGs data. The goal is to be able to respond to the risk of death in advance.

1.3 Dataset Information

The dataset (which is taken from Kaggle) contains 2126 records of features extracted from Cardiotocogram exams, the CTGs were also classified by three expert obstetricians and a **consensus classification label** assigned to each of them. Classification was both with respect to a **morphologic pattern (A, B, C. ...)** and to a **fetal class (N, S, P)**. Therefore, the dataset can be used either for 10-class or 3-class experiments.

Inputs

The dataset contains a total of 21 inputs described below:

1. baseline value
2. accelerations
3. fetal_movement
4. uterine_contractions
5. light_decelerations
6. severe_decelerations
7. prolongued_decelerations

8. abnormal_short_term_variability
9. mean_value_of_short_term_variability
10. percentage_of_time_with_abnormal_long_term_variability
11. mean_value_of_long_term_variability
12. histogram_width
13. histogram_min
14. histogram_max
15. histogram_number_of_peaks
16. histogram_number_of_zeroes
17. histogram_mode
18. histogram_mean
19. histogram_median
20. histogram_variance
21. histogram_tendency

Target Variable

It uses the **fetal_health** as the **target variable**. As mentioned above, fetal class is classified according to 3 situations (**N** — Normal, **S** — Suspect or **P** — Pathological).

1.4 Dataset Constraints

After initial EDA, it has been found that **target variable** distribution is imbalanced. In order to solve the problem related to the imbalance of the dataset, we will need to use sampling to equalize the number of samples for each of the classes **N**, **S** and **P**.

1.5 Testing Process

- Invariance Tests
- Directional Expectation Tests
- Minimum Functionality Tests (aka data unit tests)

1.6 Project Plan

The table below contains the project tasks and their estimated completion dates.

Tasks	Details	Delivery Date
Statement of Work (V1) **(Business Understanding & Problem Discovery)**	Problem Definition, Data Requirements and Assumptions	6th November, 2020
Statement of Work (V2) **(Data Acquisition & Understanding)**	Exploratory Data Analysis, Data Cleaning and Feature Engineering	23rd November, 2020
Modeling & Prototyping	Data Manipulation, Preliminary Model Building and Evaluation	23rd November, 2020
Deployment	Final Delivery	18th December, 2020