DATA699: Capstone Research Project

Fetal Health Classification

Aina Ahanmisi

Table of Contents

Introduction

Every year about one million fetal deaths occur in the United States and almost three million in the world, according to the Center for Disease Control and Prevention (DeSisto et al). The fetal mortality rate has declined some after 1990 but from 2003 to present time, rates have not decreased since (Centers for Disease Control and Prevention 2015). The goal in reducing fetal deaths is important because the reduction of fetal deaths has a detrimental effect on the world and mankind. The amount of fetal deaths in the United States have been at a steady rate for years now and has actually been increasing in some states on it's own. Preventing child and maternal mortality can also help with discovering findings on what new health conditions and risk factors disrupt the pregnancy term. Reducing the amount of fetal deaths will in fact, help in human progression by advancing scientifically and offer proper world evolution (The Global Goals For Sustainable Development 2021).

Problem Identification

The objective of this project is to assist in the goal of reducing the deaths of fetuses by properly identifying the status of a fetus early enough to reverse the causes leading to death through data exploration and classification. The target audience of this project are professionals in the Center for Disease Control and Prevent, World Health Organization, medical practitioners and scientists. These target audiences are groups of people who are able to make decisions and budget for what studies deserve funds to continue with further studies. This audience would care for fetal health because the research aids the global goal to help with human progression. This project can also target regular civilians who are mothers and fathers, more specifically, people with children. I believe mothers who have experienced fetal deaths would like to know this information or find comfort in research being done to help reduce this issue.

Dataset Description

The data set that will be used for this project is from Kaggle. There are statistics on a total of 2126 women in this Fetal Health Classification data set. All the women have already been classified into a specified category of either normal, suspect or pathological each for the status of their fetus. Normal (1) is considered to be a healthy fetus because it aligns with the normal statistics of a healthy fetus. The suspect (2) category is considered to be on the borderline of an unhealthy and healthy fetus and has a possible risk of death. The pathological (3) category is considered to be an extreme case, with the fetus being severely unhealthy and at high risk of death.

The statistics of the 2126 women are medical measurements extracted from cardiotocograms and have been classified by expert obstetricians. There are a total of 22 attributes in the Fetal Health Classification data set. The chart below condenses the 22 attributes into categories below to better briefly describe what each represents from the data set. All attributes are measurements of the fetus and not the mother.

Features	Description
Heart Rate	Measurement of the fetal heart rate where 135 beats per minute is considered to be the normal baseline fetal health rate
Heart Acceleration	Measurement of sudden acceleration of the fetal heart rate. Abrupt acceleration is a good sign of a normal fetus, because this is heavily associated with healthy fetal movement.
Movement	Measure of fetal movements per second, all fetal movement is good. There is nothing bad with an active baby because such movement promotes healthy bone and joint development (Bogle, 2020).
Uterine Contractions	Measure of uterine contractions per second. This measurement can help with infection possibilities or overall well-being of the fetus.
Decelerations	Broken into three attributes which are light, severe and prolonged. Decelerations are sudden but very temporary drops in the fetal heart rate. Light decelerations can be normal but severe or prolonged decelerations are not a good sign. These attributes are also correlated to uterine contractions.
Variability Calculations	Consist of four attributes that are all either measurements or calculations of the variations of the fetal heart rate from beat to beat (Bailey 2009). Calculations include abnormality and mean value of the short term. And percentage of abnormality and mean value of long term.

Histogram	Consist of ten attributes that all are used to form a histogram. These attributes are calculated using all values from the cardiotocography report. The calculations include the width, minimum, maximum, number of peaks and zeroes, mode, mean, median, variance and tendency.
Fetal Health	Each woman out of the 2126 are classified into the fetal health categories based on their report (Campos et al 2000)

Methodology

Overview introduction here.

Data Wrangling/Cleansing

Before moving forward with anything, the data will need to be checked for any missing values in the data set. "Missing data are values that are not recorded in a dataset" (Alam 2020). Missing data can potentially skew the result and provide results that could be inaccurate. It is best to work on data with no missing values. Thankfully this dataset does not have any missing values. For this data set zeros will not be taken out because they are actually important.

```
baseline value                                                    0
accelerations                                                     0
fetal_movement                                                    0
uterine_contractions                                              0
light_decelerations                                               0
severe_decelerations                                              0
prolongued_decelerations                                          0
abnormal_short_term_variability                                   0
mean_value_of_short_term_variability                              0
percentage_of_time_with_abnormal_long_term_variability            0
mean_value_of_long_term_variability                               0
histogram_width                                                   0
histogram_min                                                     0
histogram_max                                                     0
histogram_number_of_peaks                                         0
histogram_number_of_zeroes                                        0
histogram_mode                                                    0
histogram_mean                                                    0
histogram_median                                                  0
histogram_variance                                                0
histogram_tendency                                                0
fetal_health                                                      0
```

Data Exploration

Reference

Alam, Mahbubul (12 June 2020). *Dealing with missing data in data science projects*. Towards Data Science.
https://towardsdatascience.com/dealing-with-missing-data-in-data-science-projects-e8ac7a4efdff

Ayres de Campos et al. (2000) SisPorto 2.0 *A Program for Automated Analysis of Cardiotocograms*. J Matern Fetal Med 5:311-318 (link).

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

Centers for Disease Control and Prevention. 2015. *The Challenge of Fetal Mortality*.
https://www.cdc.gov/nchs/products/databriefs/db16.htm

DeSisto, Carla; Martin, Joyce; Gregory, Elizabeth; Hoyert, Donna; Valenzuela, Claudia. *Fetal Deaths*. Centers for Disease Control and Prevention.
https://www.cdc.gov/nchs/data/nvss/webinar/fetal-death-webinar.pdf

https://www.projectpro.io/recipes/deal-with-imbalance-classes-with-upsampling-in-python