Data Collection and Preprocessing Phase

Date	04-07-2024
Team ID	740070
Project Title	FetalAI: Using Machine Learning To Predict And Monitor Fetal Health
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan:

Section	Description
Project Overview	"Fetal mortality is a major, but often overlooked, public health problem." According to the National Center for Health Statistics, about 1 million fetal deaths occur in the United States each year, with almost 26,000 being over the age of 20 weeks gestation. After decades of decline, the US fetal mortality rate has plateaued at this shockingly high number. With fetal mortality comes greater risk of adverse maternal health outcomes, as well as maternal mortality risk increase. Cardiotocograms (CTGs) measure values such as fetal heart rate, fetal movement, and uterine contractions. "CTGs are a simple and cost-accessible option to assess fetal health, allowing healthcare professionals to take action in order to prevent child and maternal mortality." Using data from actual patients' CTG exams and their accompanying fetal health outcomes assigned by expert obstetricians, I have determined that automated assessment of fetal health is possible using CTG data.

Data Collection Plan	• Search for datasets related to Fetal Health.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle , the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details. (baseline value,accelerations, fetal_movement,uterine_contractions, light_decelerations,severe_decelerations, prolongued_decelerations, abnormal_short_term_variability, mean_value_of_short_term_variability, percentage_of_time_with_abnormal_long_term_variability, histogram_min, histogram_min , histogram_max, histogram_number_of_peaks, histogram_number_of_zeroes, histogram_mode, histogram_mean, histogram_median, histogram_variance, histogram_tendency, fetal_health	https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification	CSV	224 KB	public