

Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	739962
Project Title	One Year Life Expectancy post on Thoracic Surgery using Machine Learning
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	Predicting one-year life expectancy after thoracic surgery is crucial for patient well-being and clinical decision-making. Machine learning (ML) offers advanced capabilities to analyze complex datasets and provide accurate predictions, aiding in better management and outcomes for patients undergoing thoracic surgery.

Data Collection Plan	<ul style="list-style-type: none"> ● Search for datasets related to patient data, demographic information, and decisions regarding details. ● Prioritize datasets with diverse demographic information.
Raw Data Sources Identified	<p>The raw data sources for this project include datasets obtained from Patients, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as cough,fev,fvc,performance,Haemoptysis,pain ,weakness tumor_size,diabetes_mellitus, MI_6mo, pad, smoking,asthma,age.</p>

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset	<p>The data consist of</p> <ul style="list-style-type: none"> • Pain • FVC • FEV1 • Performance • Haemoptysis • Dyspnoea • Cough • Weakness • Tumor Size • Diabetes Mellitus • MI_6mo • PAD • Smoking • Asthma • Age • Death_1year 	ThoracicSurgery.csv - Google Drive	CSV	18 KB	Public