Data Collection And Preprocessing Phase

Date	10 July 2024
Team ID	740021
Project Title	Sepsis Survival Minimal Clinical Records
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.

Section	Description
Project Overview	The machine learning project aims to predict sepsis survival based on minimal clinical records. Using a dataset with features such as age, gender, episode number, the objective is to build a model that accurately classifies sepsis survival outcomes, aiding in timely and informed clinical decisions.

Data Collection Plan	1. Search for datasets related to sepsis patients survival, and clinical records 2. Prioritize datasets with diverse demographic and clinical information
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from platforms like Kaggle. The provided sample data represents a subset of the collected information, encompassing variables such as age, gender, episode number, hospital outcomes.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permission
Kaggle Dataset	The dataset comprises patient details (age, gender, episode number) and sepsis survival clinical outcomes	https://github.com/Jorukasandeep25/dataset.git	CSV		Public