# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 15 July 2024 |
| Team ID | 740116 |
| 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.

**Data Collection Plan:**

| Section | Description |
|---|---|
| Project Overview | The machine learning project aims to predict survival outcomes based on clinical records of sepsis patients. Using a dataset with features such as age, sex, episode number, and hospital outcome, the objective is to build a model that accurately classifies survival status (alive or dead), facilitating efficient and informed decision-making in clinical settings. |
| Data Collection Plan | <ul><li>Search for datasets related to sepsis survival clinical records.</li><li>Prioritize datasets with comprehensive clinical details.</li></ul> |

| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from publicly available medical records and repositories. The provided sample data represents a subset of the collected information encompassing variables such as age, sex, episode number, and hospital outcome for machine learning analysis. |
|---|---|

**Raw Data Sources Report:**

| Source Name | Description | Location/URL | Format | Size | Access Permissions |
|---|---|---|---|---|---|
| Kaggle Dataset (Sepsis Survival Primary Cohort) | The dataset comprises clinical details (age, sex, episode number, hospital outcome) for sepsis patients. | https://www.kaggle.com/datasets/joebeachcapital/sepsis-survival-minimalclinical-records | CSV | 1.3MB | Public |