# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 15 March 2024 |
| Team ID | 738193 |
| Project Title | Hospital Readmission Prediction Using Machine Learning |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

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 Template**

| Section | Description |
|---|---|
| Project Overview | The Hospital Readmission Prediction dataset is a collection of data gathered from hospital records, containing information about patients' demographics, medical history, treatments, and outcomes. The primary objective of this dataset is to predict whether a patient will be readmitted to the hospital within a certain period after their initial discharge.<br><br>This dataset typically includes various features such as patient demographics (age, gender, race), medical history (diagnoses, procedures), medications prescribed, laboratory test results, and information about the hospital stay (length of stay, admission type, discharge disposition).<br><br>The goal of analyzing this dataset is to develop predictive models that can accurately identify patients at risk of readmission. By understanding the |

| | |
|---|---|
| | factors contributing to readmission, healthcare providers can implement targeted interventions and improve patient care strategies to reduce the likelihood of unnecessary hospital readmissions. |
| Data Collection Plan | This project uses information from hospitals to predict if patients will need to come back to the hospital soon after they've been discharged. The data we use includes details like a patient's age, medical history, the medicines they were given, and how long they stayed in the hospital. We want to figure out which factors make it more likely for someone to need to come back to the hospital after they've left. By understanding this, we can try to improve how hospitals take care of their patients and reduce the number of people who have to return to the hospital unexpectedly. |
| Raw Data Sources Identified | The raw data for this dataset comes from hospitals. Hospitals keep records of their patients, including information like their age, medical history, the treatments they received, and whether they had to come back to the hospital after being discharged. This data is collected from hospital databases or electronic health records systems. We use this information to study patterns and try to predict if certain factors make it more likely for someone to be readmitted to the hospital. |

**Raw Data Sources Template**

| Source Name | Description | Location/URL | Format | Size | Access Permissions |
|---|---|---|---|---|---|
| Dataset 1 | A hospital readmission dataset is a collection of information about patients who are admitted to a hospital, then return for another stay within a certain period (usually 30 days). This data is like a goldmine for researchers and healthcare providers because it helps them understand why patients come back and how to prevent it. | https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008 | CSV | 18.2 MB | Public |
| Dataset 2 | Description of the data in this source. | https://www.kaggle.com/datasets/saurabhtayal/diabeti | Excel | 4MB | Private (with access) |

| | | c-patients-readmission-prediction | | | |
|---|---|---|---|---|---|
| | | | | | |