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 Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	Missing Values	Moderate	The presence of missing values in the dataset, particularly in fields such as laboratory results, medications, and comorbidity indicators, may introduce bias and affect the accuracy of the predictive model. Strategies for imputation or handling missing data will be implemented to mitigate this issue.

Dataset	Data Inconsistencies	Moderate	Inconsistencies or discrepancies in the data, including coding errors, duplicate records, and inconsistencies in data formats or units, may compromise the integrity of the dataset and lead to erroneous predictions. Data cleaning and validation procedures will be conducted to identify and rectify such inconsistencies
Dataset	Class Imbalance	Low	Class imbalance, where one class (e.g., readmitted patients) is significantly underrepresented compared to the other class (e.g., non-readmitted patients), may skew the model's predictions and reduce its effectiveness. Techniques such as oversampling, undersampling, or using weighted loss functions will be employed to address this issue and ensure balanced model training.

Dataset	Temporal Variability	Low	Temporal variability in patient data, such as changes in disease progression, treatment regimens, or healthcare practices over time, may influence readmission risk and impact the model's predictive performance. Time-series analysis and dynamic modeling approaches will be explored to capture and adapt to temporal trends in the data.