

Data Collection and Preprocessing Phase

|               |  |
|---------------|--|
| Date          | June 25 2024   |
| Team ID       | team-740066  |
| Project Title | Hospital Readmission Prediction 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 Rw 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 hospital readmission based on patient information. Using a dataset with features such as gender,age,number_outpatient,number_inpatient,number_emergency,number_daignoses, the objective is to develop a machine learning model that accurately predicts whether a patient is likely to be readmitted to the hospital within a specified time frame after discharge.                                      |
| Data Collection Plan | <div><ul style="list-style-type: none"><li>Search for datasets related to hospital readmission prediction.</li><li>Prioritize datasets with diverse demographic information.</li></ul></div> <div>By following this data collection plan, healthcare organizations can effectively gather, preprocess, and utilize data to develop accurate predictive models for hospital readmission, thereby enhancing patient care and resource allocation.</div> |
|                      |   |

|                             |   |
|-----------------------------|---|
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle ,the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as gender,age,number_diagnoses,readmitted,diag_1,diag_2,diag_3, |
|-----------------------------|---|

**Raw Data Sources Report:**

| Source Name    | Description   | Location/URL  | Format | Size | Access Permissions |
|----------------|---|---|--------|------|--------------------|
| Kaggle Dataset | The dataset comprises patient details(gender,age)and number inpatient, number outpatient,number daigonsis,number emergency,readmitted | <a href="https://www.kaggle.com/datasets/abhishvekvats95/hospital-readmission-prediction-using-diabetes">https://www.kaggle.com/datasets/abhishvekvats95/hospital-readmission-prediction-using-diabetes</a> | CSV    | 8MB  | Public             |