DataCollectionandPreprocessingPhase

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| TeamID | 740686 |
| ProjectTitle | Lymphography Classification Using ML |
| Maximum Marks | 2 Marks |

# DataCollectionPlan &Raw DataSources IdentificationReport:

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 analysisanddecision-making endeavor.

# Data Collection Plan:

|  |  |
| --- | --- |
| **Section** | **Descripion** |
| Project Overview | The machine learning project aims to predict the condition of lymph nodes. It’s so significant that it is used to accurately sense the stage in cancer progression,which decides the treatment to be adopted. Using a dataset with features such as class,lymphatics,by pass the objective is to build amodel that accurately sense the stage in cancer progression,facilitating efficient and informed decision-making in the lendingprocess. |
| Data Collection Plan | This detailed description outlines a comprehensive approach to collecting and preparing data for lymphography classification, ensuring transparency, compliance, and readiness for subsequent stages of machine learning model development |
| Raw Data SourcesIdentified | The raw data sources for this project include datasets obtained fromKaggle & UCI, the popular platforms for data science competitionsand repositories. The provided sample data represents a subset ofthecollectedinformation,encompassingvariablessuchasclass, |

|  |  |
| --- | --- |
|  | by pass, block of affere, extravasates and changes in node details for machine learninganalysis |

**RawDataSourcesReport:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SourceName** | **Description** | **Location/URL** | **Format** | **Size** | **AccessPermissions** |
| KaggleDataset | The datasetcomprisesannotated medical images aimed at classifying conditions of the lymphatic system. It includes diverse lymphography images labeled with categories such as Normal, Metastasis, and Primary Tumor, crucial for automated diagnosis. | <https://www.kaggle.com/datasets/naifislam/lymphography-dataset> | CSV | 5.62kB | Public |
| UCI | It consists of lymphography data essential for medical imaging research. It includes features detailing lymphatic system status and imaging findings related to lymph node conditions. | <https://archive.ics.uci.edu/dataset/63/lymphography> | CSV | 2.98kB | Public |