**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| **Date** | 15 June 2024 |
| **Team Id** | 739820 |
| **Project Title** | Predicting the unpredictable: A Look into the World of Powerlifting |
| **Max marks** | 2marks |



**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 loan approval based on applicant information. Using a dataset with features such as gender, marital status, income, and credit history, the objective is to build a model that accurately classifies loan status (approved or denied), facilitating efficient and informed decision-making in the lending process. |
| Data Collection Plan |  **Search for datasets**: related to loan approvals, financial information, and applicant details.   **Prioritize datasets**: with diverse demographic information. |
| 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 collected encompassed variables such as player id , Age, Name, Equipment, Body weight in kgs ,Best squat in kgs, Best deadlift in kgs, Best bench press in kgs details for machine learning analysis. |

**Raw Data Sources Report:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source**  **Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle  Dataset | The dataset comprises applicant details (Player id, Name, Age, Equipment, Best bench press in kgs, Body weight in kgs, Best squat in kgs ) | <https://www.kaggle.com/open-powerlifting/powerlifting-database> | CSV | 623.3kB | Public |