Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 20 April 2024 |
| Team ID | Team-738178 |
| Project Title | Envisioning Success : Predicting University Scores With 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 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 create a machine learning model that predicts university scores based on specific characteristics such as quality of education, faculty, research output, and impact metrics. This tool will help prospective students make informed decisions about higher education options while providing universities with valuable feedback for performance improvement. |
| Data Collection Plan | Gather data on university characteristics such as education quality,  faculty quality, publications, influence, citations, patents, and alumni  employment from reputable sources. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from government databases, Academic databases, Kaggle & UCI, the popular platforms for data science competitions and repositories. |

**Raw Data Sources Report:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle Dataset | The dataset comprises university characteristics such as education quality, faculty, research output, influence, citations, patents, and alumni employment | <https://www.kaggle.com/datasets/mylesoneill/world-university-rankings> | CSV | 477 kB | Public |