**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 06 March 2024 |
| Team ID | 739915 |
| Project Title | BlueBerry Yield Prediction |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

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 Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | The machine learning project aims to predict Blueberry yield based on applicant information. Using a dataset with features such Raining days, Temperature, Pollinate factors .The objective is to build a model that accurately predict the yield, facilitating efficient and informed decision-making in the prediction. |
| Data Collection Plan | Search for datasets related to blueberry yield , crop prediction, and applicant details. |
| Raw Data Sources Identified | The raw data sources for this project include datasets obtained from Kaggle ,the popular platform for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such rainfall, temperature. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle Dataset | The dataset comprises factors, such as weather conditions, soil quality, and pest and disease pressure.. | <https://www.kaggle.com/datasets/saurabhshahane/wild-blueberry-yield-prediction> | CSV | 84KB | Public |