

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | SWTID1720436539 |
| Project Title | SportSpecs: Unraveling Athletic Prowess with Advanced Transfer Learning for Sports |
| Maximum Marks | 2 Marks |

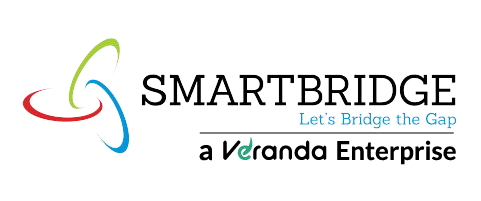
**Data Collection and Preprocessing Phase**

**Data Quality Report Template**

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Kaggle Sports  Classification  Datase | Missing values in the  'sport' column | High | Impute missing values using the mode of the column or by referencing other related columns. |
| Kaggle Sports  Classification  Datase | Inconsistent labeling of sports categories | Moderate | Standardize labels by ensuring consistent formatting and verifying against a predefined list of categories. |
| Kaggle Sports  Classification  Datase | Duplicate entries in the dataset | Moderate | Identify and remove duplicate entries to ensure data integrity. |

.



|  |  |  |  |
| --- | --- | --- | --- |
| Kaggle Sports  Classification  Datase | Imbalanced dataset with respect to different sports | High | Apply resampling techniques like SMOTE (Synthetic Minority Oversampling Technique) or undersample the majority class to balance the dataset. |
| Kaggle Sports  Classification  Datase | Noise in the image data (blurry or low-quality images) | High | Use image enhancement techniques or exclude noisy images from the dataset. |