

Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWTID1720158677
Project Title	SportSpecs: Unraveling Athletic Prowess With Advanced Transfer Learning For Sports.
Maximum Marks	2 Marks

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 Dataset	Inconsistent labeling of sports categories	Moderate	Standardize labels by ensuring consistent formatting and verifying against a predefined list of categories.
Kaggle Sports Classification Dataset	Blurry or low-resolution images in some sports images.	High	Remove or replace low-quality images with higher resolution ones, or use image enhancement techniques to improve quality.

Kaggle Sports Classification Dataset	Class imbalance with significantly fewer images in the swimming class compared to others	Moderate	Use data augmentation techniques (e.g., rotation, flipping) to artificially increase the number of images in the swimming class.
Kaggle Sports Classification Dataset	Inconsistent image sizes across different sports classes.	Low	Resize all images to a standard size (e.g., 224x224 pixels) during preprocessing to ensure uniformity.
Kaggle Sports Classification Dataset	Duplicate images found in some sport classes	Low	Identify and remove duplicate images to ensure the dataset contains unique samples.