Data Collection and Preprocessing Phase

Date	16 June 2025
Team Lead Name	Jayanth Srinivas Bommisetty
Project Title	Sloan Digital Sky Survey (SDSS) galaxy classification using machine learning
Maximum Marks	2 Marks

Data Quality Report Template

The images were initially checked for corruption to ensure the dataset's integrity. All images were then resized uniformly to 256x256 pixels to align with the model's input shape. Lastly, the dataset was examined for class balance to confirm fair representation across all five galaxy categories.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Images loading issues and corrupted images in "Train_images"	High	Identifying the corrupted images and either replace it with another image or delete the corrupted images