Data Collection and Preprocessing Phase

Date	17 July 2025
Team ID	739816
Project Title	Galactic Gazetteer :A comprehensive Dataset of Planet Classification Images
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

For this project, the data will be collected from publicly available astronomical image datasets such as NASA's Planetary Data System (PDS) and the European Space Agency (ESA) archives. These datasets include labeled images of various planet types such as terrestrial planets, gas giants, ice giants, and dwarf planets, which will serve as the primary source for training the planet image classification model. The dataset can be downloaded in JPEG or PNG image formats along with metadata files (CSV or JSON) containing the corresponding planet category labels for each image.

Data Collection Plan Template

Section	Description
Project Overview	This project aims to develop a deep learning-based system for classifying planetary images into categories like terrestrial planets, gas giants, ice giants, and dwarf planets using a CNN model. The system will be trained on a dataset of labeled planetary images from NASA and ESA. Key stages include data collection, preprocessing, model training, and evaluation. A Flask web app will be built to allow users to upload images and get real-time classification results. The project aims to streamline planetary classification for educational, research, and astronomical data analysis.
Data Collection Plan	The data for this project will be collected from publicly available astronomical image datasets, including sources like Kaggle,NASA's Planetary Data System (PDS) and ESA archives. The dataset will include labeled images of various planets, categorized into types such as terrestrial planets, gas giants, ice giants, and dwarf planets. These images will serve as the primary data source for training and evaluating the deep learning model for planet image classification. The data will be supplemented with metadata files, including planet

	type labels, to ensure accurate training and evaluation of the model.
Raw Data Sources Identified	The raw data for this project will be sourced from publicly available astronomical image repositories such as NASA's Planetary Data System (PDS) and the European Space Agency (ESA). These datasets contain a diverse collection of labeled images of various planets, including terrestrial planets, gas giants, ice giants, and dwarf planets, captured from different telescopes and space missions. The images will be accompanied by metadata, including the planet type, image resolution, and the observation details (such as the spacecraft or telescope used and the date of capture), which will help train a robust planet classification model.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Description of the data in this source.	https://www.kaggle.com/datasets/emirhanai/planets-and-moons-dataset-ai-in-space/code	CSV	**1.2** GB	Public