

Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	740051
Project Title	SDSS galaxy classification using Machine Learning
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 SDSS galaxy classification project applies machine learning to analyze spectral data, identifying galaxy types like spirals or ellipticals, advancing understanding of cosmic evolution and structure.
Data Collection Plan	Kaggle
Raw Data Sources Identified	<p>Raw data sources for SDSS galaxy classification using machine learning include SDSS spectral observations, feature extractions like flux measurements, and metadata such as redshift and positional information for comprehensive astronomical analysis.</p> <p>The raw data sources for this project include datasets obtained from Kaggle, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected informatio</p>

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	The dataset consists of single CSV file. This dataset is mainly concerning Indian climatic conditions. There are seven input features.	"C:\Users\ABHINAY THEDLAA\OneDrive\Desktop\miniproject\miniproject\Flask\sdss_100k_galaxy_form_burst.csv"	CSV	18847KB	Public

	es only one output featur es.				
--	--	--	--	--	--