Project Initialization and Planning Phase

Date	17 June 2024
Team ID	xxxxxx
Project Title	sloan digital sky survey (sdss) galaxy classification using machine learning
Maximum Marks	3 Marks

Project Proposal

This proposal outlines a plan to enhance the Sloan Digital Sky Survey (SDSS) galaxy classification through machine learning, aiming to improve both efficiency and accuracy. It addresses current system limitations, offering improved performance, minimized risks, and increased user satisfaction. Notable elements include a machine learning-driven classification model and support for real-time decision-making.

Project Overview	
Objective	The primary Objective is to transform the galaxy classification process in the Sloan Digital Sky Survey (SDSS) through the use of advanced machine learning methods, leading to quicker and more accurate evaluations.
Scope	This project thoroughly evaluates and improves the SDSS classification system by integrating machine learning to create a more reliable and efficient framework.
Problem Statement	
Scenario – 1 Galaxy Morphology Classification	
Description	Understanding galaxy morphology is crucial for astronomers studying how galaxies form and evolve. Machine learning offers a powerful approach, allowing researchers to develop models that classify galaxies into various morphological categories like spiral, elliptical, or irregular.
Impact	Automating the classification process allows astronomers to efficiently examine vast collections of galaxy images and uncover patterns or trends linked to galaxy morphology.
Scenario – 2 Galaxy Redshift Estimation	

Description	Redshift, reflecting how much a galaxy's light has been stretched to longer wavelengths by the universe's expansion, is a key parameter in exploring cosmic distances and understanding cosmological events.
Impact	Machine learning algorithms can be trained to predict galaxy redshifts using features derived from their spectral data or photometric measurements provided by SDSS. Precise redshift predictions help astronomers map the three-dimensional layout of galaxies across the universe and study large-scale cosmic structures like galaxy clusters and filaments.
Scenario – 3 Active Galactic Nuclei (AGN) Identification	
Description	Galaxies containing active galactic nuclei (AGN) show strong emission from a compact central region, driven by matter accreting onto supermassive black holes. Detecting AGN candidates within SDSS data is crucial for analyzing their characteristics and exploring their role in galaxy evolution.
Impact	Machine learning techniques can be developed to detect distinctive AGN signatures in galaxy spectra or multi-wavelength photometric data, enabling the automated identification of AGN-hosting galaxies in extensive surveys like SDSS. This approach allows astronomers to perform large-scale statistical studies of AGN characteristics and examine their influence on galaxy formation and evolution.
Proposed Solution	

Approach	• Use supervised machine learning models:
	  ○ Random Forest Classifier for morphology classification
	  ○ Random Forest Regressor for redshift prediction
	• Preprocess data: handle missing values, encode categorical data, and normalize features.
	• Evaluate models using accuracy, R^2 score, RMSE, and MAE.
	• Deploy using Flask web framework with Bootstrap UI.

Key Features	• End-to-end ML pipeline (data cleaning to deployment) • Two-in-one solution: classification and regression • Web interface for user interaction • Supports fast, real-time prediction • Easy to expand for future astronomical datasets

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	T4 GPU
Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	512 GB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, matplotlib, seaborn
Development Environment	IDE, version control	Google collab, Git
Data		
Data	Source, size, format	Kaggle dataset