

SmartBridge -Forecasting Economic Prosperity: SDSS GALAXY CLASSIFICATION USING MACHINE LEARNING

Milestone 1: Project Initialization and Planning Phase

Activity 1: Define Problem Statement

Description: The project aims to develop and deploy machine learning models to accurately classify galaxies in the SDSS dataset. The goal is to enhance astronomical research by providing reliable classifications that account for complex patterns in the data, facilitating better understanding of the universe.

Activity 2: Project Proposal (Proposed Solution)

Description: The proposed solution involves applying advanced machine learning techniques to analyze the SDSS dataset. By selecting and training robust models, the project aims to improve the accuracy of galaxy classifications, aiding in astrophysical discoveries and research.

Activity 3: Initial Project Planning

Description: This involves defining the objective of classifying galaxies, acquiring the SDSS dataset, forming a multidisciplinary team, establishing a structured timeline with milestones, allocating resources, assessing risks, and implementing a communication plan for effective collaboration and progress tracking.

Milestone 2: Data Collection and Preprocessing Phase

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

Description: Gather relevant SDSS data, ensuring data quality through verification and addressing missing values. The dataset includes features like redshift, magnitudes, and spectral lines.

Activity 2: Data Quality Report

Description: The collected data from the SDSS is comprehensive and reliable. Data preprocessing techniques such as normalization, handling missing values, and outlier detection further enhance dataset integrity for accurate machine learning model training

Activity 3: Data Exploration and Preprocessing

Description: Initial exploratory data analysis reveals patterns and correlations within the SDSS dataset. Missing values are handled through imputation methods, and categorical variables are encoded appropriately. Feature scaling is applied to ensure uniformity in data distribution for machine learning model training.

Milestone 3: Model Development Phase

Activity 1: Feature Selection Report

Description: Feature selection is performed using techniques such as correlation analysis, recursive feature elimination, and feature importance from ensemble models. Key variables influencing galaxy classification are identified for model development.

Activity 2: Model Selection Report

Description: After evaluating various machine learning algorithms including decision trees, random forests, and support vector machines, the best performing models are selected based on metrics such as accuracy, precision, and recall. Hyperparameter tuning further optimizes model accuracy and robustness.

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

Description: The selected model is trained on the SDSS dataset, achieving high accuracy and robust performance in classifying galaxies based on the selected features.

Milestone 4: Model Optimization and Tuning Phase

Activity 1: Hyperparameter Tuning Documentation

Description: The selected model's hyperparameters are optimized using techniques like Research, focusing on parameters such as `adept`, `min_samples_split`, and `max_features`. Cross-validation helps fine-tune the model, achieving improved accuracy and reducing errors. **Deliverable:**

Activity 2: Performance Metrics Comparison Report

Description: This report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the tuned model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning. **Deliverable:**

Activity 3: Final Model Selection Justification

Description: This report articulates the rationale for choosing the final model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives. **Deliverable:**

Milestone 5: Project Files Submission and Documentation

Activity: Submission and Documentation

Description: For project file submission on GitHub, refer to the provided flow and documentation links. **Deliverables:**

Milestone 6: Project Demonstration

Activity: Project Demonstration

Description: Record a video explaining the project and demonstrating its execution during the presentation. **Deliverable**