Classification Project Proposal

Predicting Space Objects (Star, Galaxy, Quasar)

Hayat & Juri

10/20/21

أكاديمية سدايا
SDAIA Academy

Overview

This project presents a classification model for predicting Stars, Galaxies and Quasars. The model will be useful for astrophysicists, researchers, space observers and people interested in identifying observed data from the space and classify them into either Starts, Galaxy or Quasars. The data is originally sourced from Sloan Digital Sky Survey and imported from (Kaggle.com). The Sloan Digital Sky Survey (SDSS) offers accessible open-sourced data of space observations with dataset available since 1988.

As for the model development, we will be following the presented workflow for data analysis.

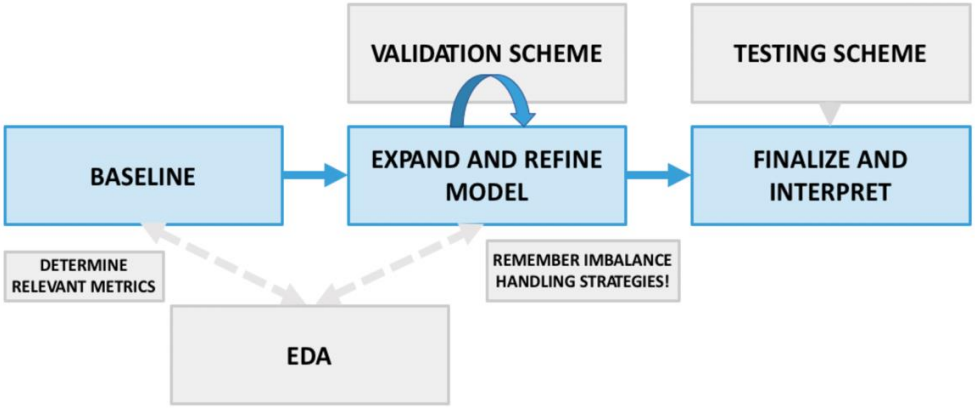


Figure 1: Classification Model Workflow

Data Description

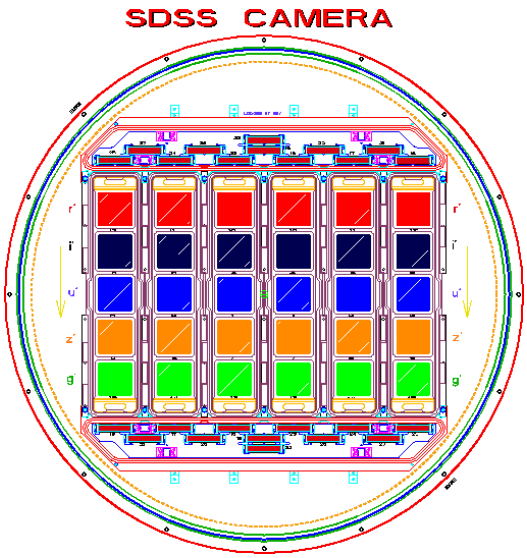
The dataset have 10,000 observations of space provided by the SDSS. In each observation there are 18 feature columns.

Table 1: Data Description

Features	Description	Remarks
objid	object Identifier	uniquely identify every object in data
Ra	J2000 Right Ascension (r-band), angular distance	
dec	J2000 Declination (r-band), angular distance	
u	better of DeV/Exp magnitude fit	Thuan-Gunn astronomic magnitude system. u, g, r, i, z represent the response of the 5 bands of the telescope.
g	better of DeV/Exp magnitude fit	
r	better of DeV/Exp magnitude fit	
i	better of DeV/Exp magnitude fit	
z	better of DeV/Exp magnitude fit	
Run	Run Number	identifies the specific scan
Rereun	Rerun Number	Each rerun consists only in a change to the photometric pipeline, not to the underlying data
Camcol	Camera column	identifying the scanline within the run

Field	Field number	The field is an integer uniquely identifying a detection in the photo catalog
Specobjid	Object Identifier	uniquely identify every space object
Class	Space object class (galaxy, star, or quasar object)	
Redshift	Final Redshift	
Plate	plate number	
Mjd	MJD of observation	modified Julian date (gives the number of days since midnight on November 17, 1858)
Fiberid	fiber ID	

The SDSS camera scans the sky in strips along great circles. Each strip consists of six parallel *scanlines*. Each scanline includes data in all five filters, *ugriz*.



The SDSS camera worked in drift scan mode, opening its shutter for extended periods and imaging a continuous strip of the sky. In the figure shown bellow the sky drifts downwards. Each continuous drift scan is referred to as a run and there is a unique integer identifying the run.

The SDSS camera had six parallel camera columns, meaning that each run is divided into six parallel scanlines, one for each camera column. These images are known as camcols, and are numbered 1 through 6.

The SDSS spectrograph uses optical fibers to direct the light at the focal plane from individual objects to the slithead. Each object is assigned a corresponding fiberID. Each fiber is surrounded by a large sheath which prevents any pair of fibers from being placed too close on the same

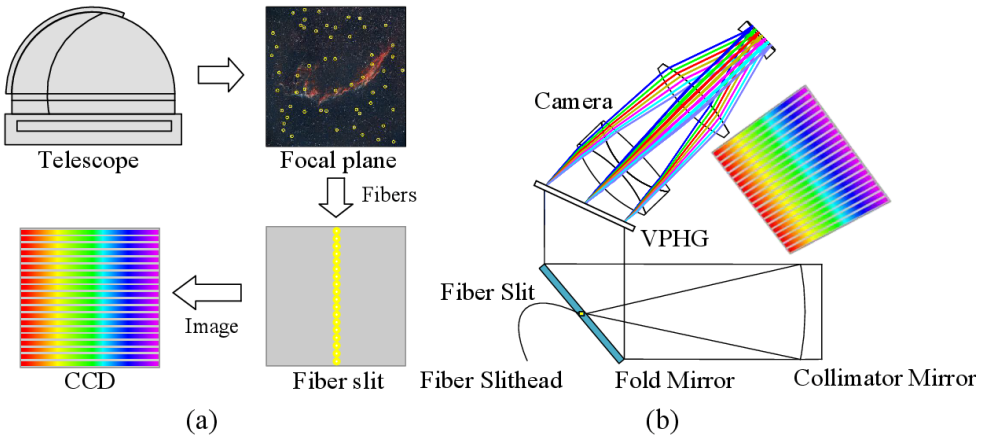


plate. When two targets are too close to each other the highest-priority target is observed. The other target may be observed if there is an overlapping plate covering this region.

Tools

- Pandas and NumPy packages to data manipulation.
- Matplotlib library for data visualization.
- LogisticRegression model from sklearn.linear_model class for classification.
- train_test_split function in Sklearn model selection for data splitting.
- KNeighborsClassifier model from sklearn.neighbors
- DecisionTreeClassifier model from sklearn.tree
- RandomForestClassifier model from sklearn.ensemble for classification model.
- Jupyter notebook that hosts the code.
- Prezi for presentation.

MVP Goal:

The aimed output of the prediction model would have a throughout report with expected or predicted classification of the type of space object whether it's a Star, Galaxy, or Quasar, where this information will be useful for our clients (astrophysicists, researchers, space observers) where they are going to use those findings to focus on actual research rather than wasting time on identifying space objects. This would be achieved by Building classification models then selection and evaluation the best model using proper validation and testing methods on the SDSS data.

Conclusion

Huge space observation data are available from SDSS, which can provide great insights for astrophysicists, researchers, space observers and other people interested in the field. A predication model to classify the space data into Stars, galaxy, or quasar will provide great assistance for space observers.