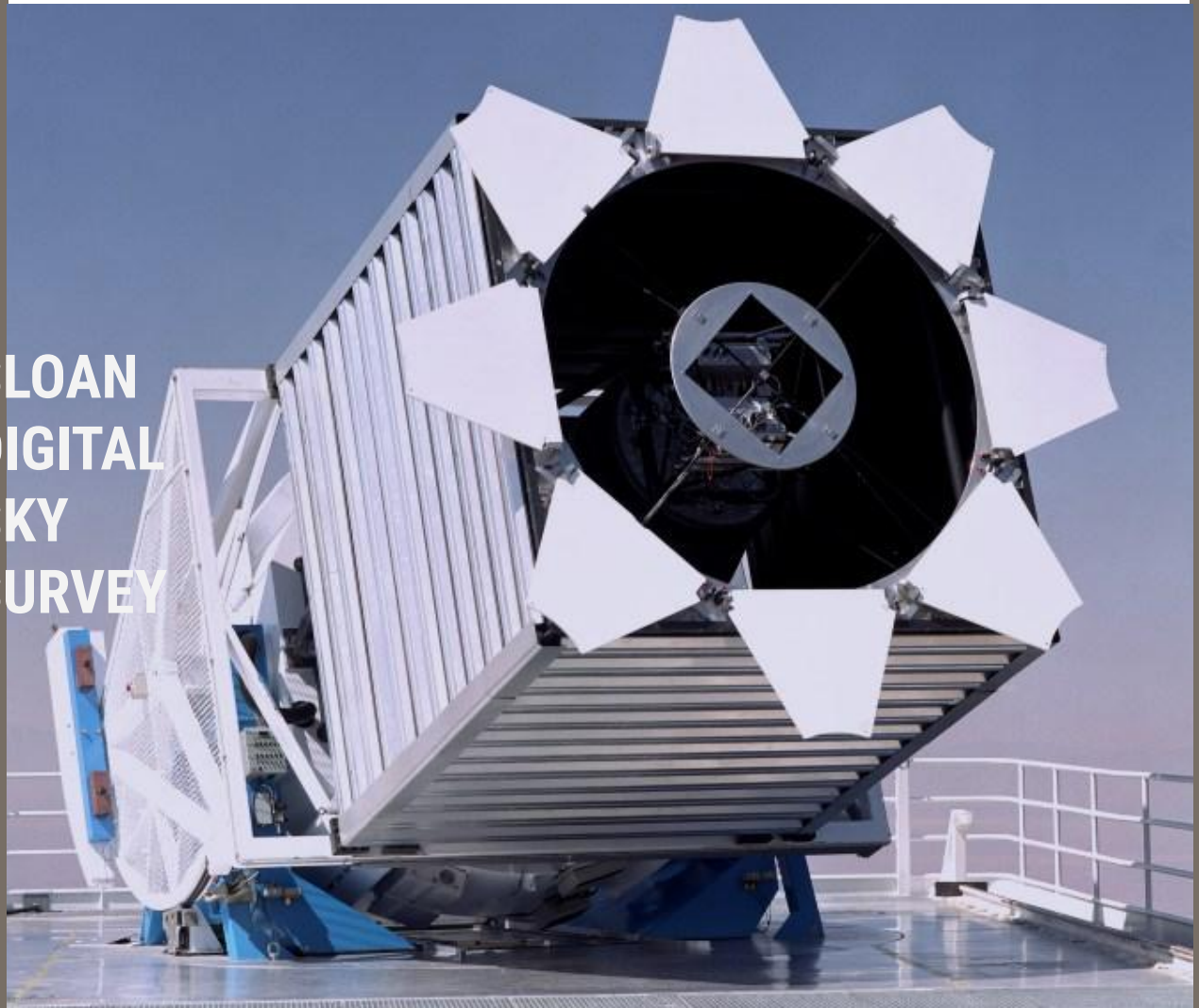
STELLAR CLASSIFICATION

By: Marco Jimenez

What is Stellar Classification?

In Astronomy, stellar classification is the classification of stars based on their spectral characteristics. The classification scheme of galaxies, quasars, and stars is one of the most fundamental in astronomy. The early cataloguing of stars and their distribution in the sky has led to the understanding that they make up our own galaxy and, following the distinction that Andromeda was a separate galaxy to our own, numerous galaxies begun to be surveyed as more powerful telescopes were built. This project aims to classify stars, galaxies, and quasars based on their spectral characteristics

SLOAN DIGITAL SKY SURVEY



This dataset consists of 100,000 observations of space taken by the SDSS (Sloan Digital Sky Survey). Every observation in this dataset is described by 17 feature columns and 1 class column which identifies it to be either a star, galaxy, or quasar

UNDERSTANDING

How do I analyze data that
I know nothing about?

Angular Positions

Right ascension angles(alpha), Declination angles(delta)

Spectral Photometric Characteristics

Ultraviolet(u), Infrared(z), Near Infrared(i)

Ultraviolet(u), Green(g), red(r) Near Infrared(i), Infrared(z)

Observational Identifying Information

Julian Dates(MJD), SDSS plate ID's(plate), Object Identifiers(obj_ID)

Optical Spectroscopic ID's(spec_obj_ID),

CLASS

Galaxies, Stars, Quasars

ANGULAR FEATURES

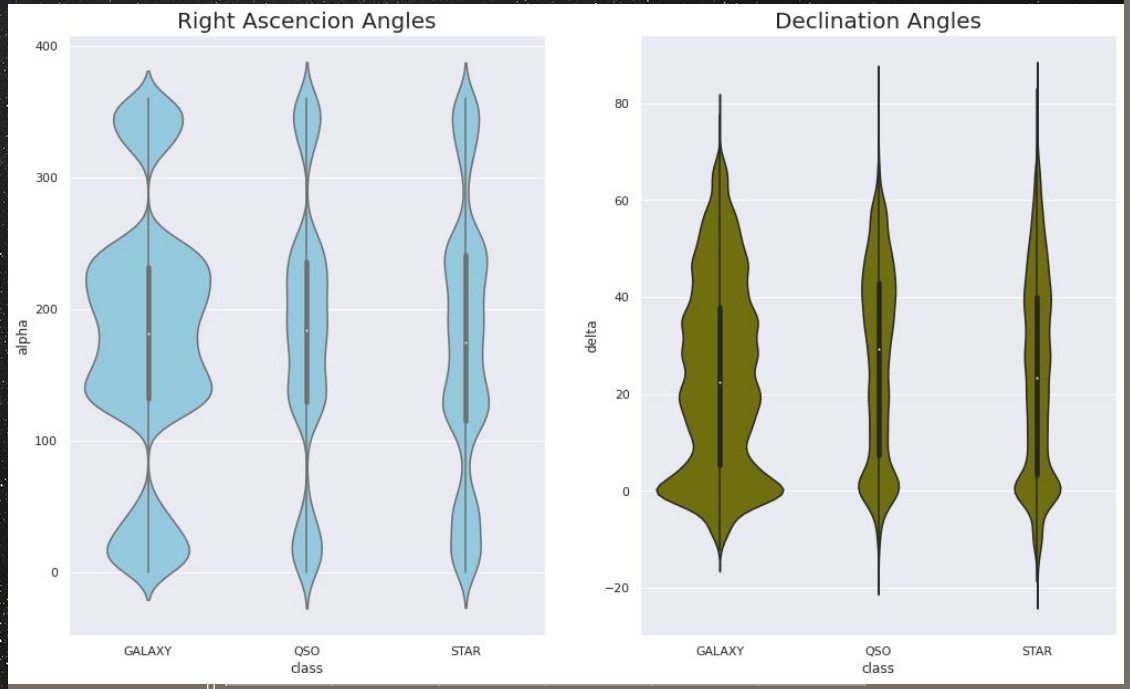
Alpha: Right Ascension Angle

Right Ascension is the astronomical equivalent of longitude. Right Ascension measures the positions of celestial objects in an east-west direction, like longitude, but unlike longitude right ascension is a time-based coordinate.

Delta: Declination Angle

Declination is the astronomical equivalent of latitude. Declination is an angular distance of a point north or south of the Celestial Equator, essentially a projection of the Earth’s equator into space

They contribute to locating an existing celestial object, but not to predicting its stellar classification.



These features together essentially represent the stellar coordinates of celestial objects from our POV on Earth. While useful in coordinating the location of a given object, they do not contribute to its stellar classification. This is supported by the violin plots above as the shape of distribution is very similar across all 3 classes in both plots.

SPECTRAL FEATURES

U
Ultraviolet filter in the photometric system.

g
Green filter in the photometric system.

r
Red filter in the photometric system.

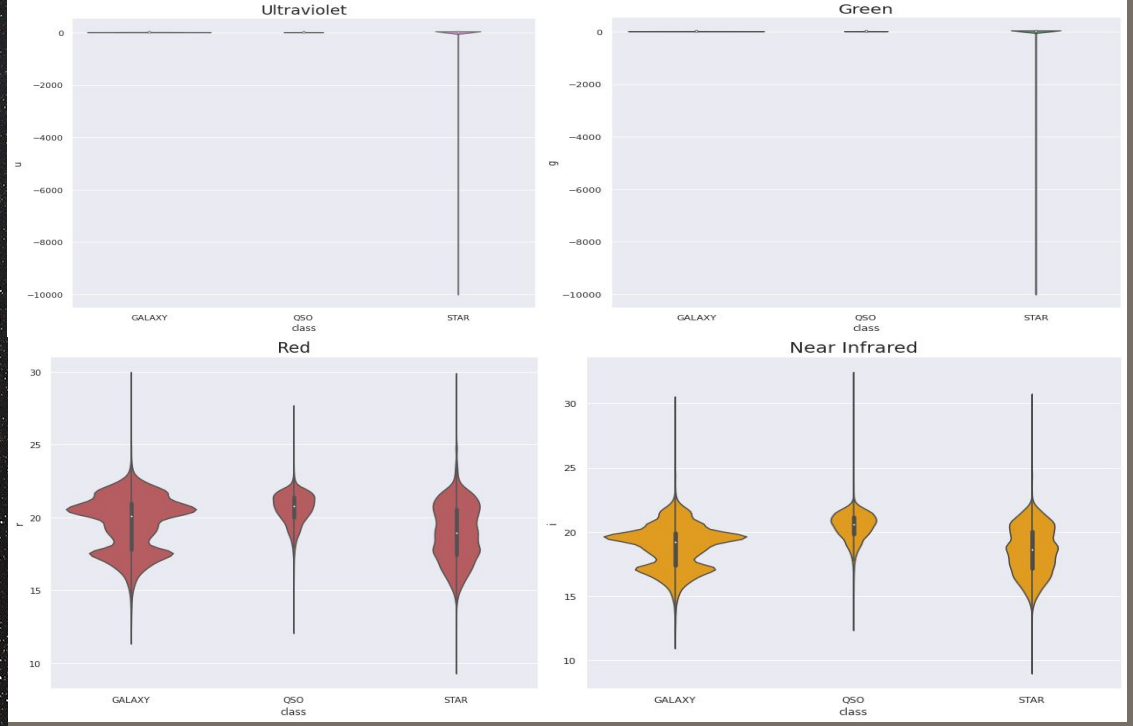
i
Near Infrared in the photometric system.

z
Infrared filter in the photometric system.

About Spectral Characteristics

Spectral characteristics are the most significant factor in classifying a stellar object. This is done through analyzing electromagnetic radiation from a stellar object by splitting it with a prism into a spectrum exhibiting the rainbow of colors interspersed with spectral lines. The strength of different spectral lines vary mainly due to the temperature of the photosphere. Based on this method and more, stellar objects can be classified into their separate classes.

These features are essential in classifying stellar bodies.



As you can see from the bottom two plots that are actually displaying information,

OBSERVATIONAL IDENTIFYING INFORMATION

MJD

Modified Julian Date. Used to indicate when a given piece of SDSS data was taken.

Spec obj_ID

Unique ID used for optical spectroscopic objects.

Run_ID

Run number used to identify the specific scan.

Fiber_ID

Fiber ID that identifies the fiber that pointed the light at the focal plane in each observation

Plate

Plate ID identifies each plate in SDSS

Clean Dataset!

Beyond my observations, this data is ready for machine learning. It is a clean dataset with no missing values or inconsistent features. The only obstacles I foresee are handling the outliers and imbalance in the dataset when it comes to machine learning. It may even benefit me to remove the Quasar class altogether to make it a simple binary classification problem. Either way I’m ready for the challenge!

THANKS!

Marco Jimenez

Email
Linked In
Github