

## DT and RF: Astronomical Object Detection

Dataset: astronomy\_2k.csv

### Context

The detection of celestial objects observed through telescopes as being either a star, a galaxy or a quasar, is an important classification scheme in astronomy. A star is an astronomical object consisting of a luminous plasma spheroid held together by the force of its own gravity. Galaxies are gravitationally bound groupings or systems of stars that additionally contain other matter such as stellar remnants, interstellar gas, cosmic dust and even dark matter. Quasars, also called Quasi-stellar objects (abbrev. QSO) are a kind of highly luminous "Active Galactic Nucleus".

### Objectives

- The objective of the problem is to use the tabular features available to us about every astronomical object, to **predict whether the object is a star, a galaxy or a quasar**, through the use of supervised machine learning methods.
- We have to use simple non-linear methods such as **k-Nearest Neighbors, Decision Trees and Random Forest** to perform this classification.
- Perform hyper parameter tuning and compare the accuracy and other performance measures of different models

### Data Description

The following dataset consists of 250,000 celestial object observations taken by SDSS. Each observation is described by 17 feature columns and 1 class column that identifies the real object to be one of **a star, a galaxy or a quasar**.

- **objid** = Object Identifier, the unique value that identifies the object in the image catalog used by the CAS
- **u** = Ultraviolet filter in the photometric system
- **ra** = Right Ascension angle (at J2000 epoch)
- **dec** = Declination angle (at J2000 epoch)
- **g** = Green filter in the photometric system
- **r** = Red filter in the photometric system
- **i** = Near-Infrared filter in the photometric system
- **z** = Infrared filter in the photometric system
- **run** = Run Number used to identify the specific scan
- **rerun** = Rerun Number to specify how the image was processed
- **camcol** = Camera column to identify the scanline within the run
- **field** = Field number to identify each field
- **specobjid** = Unique ID used for optical spectroscopic objects (this means that 2 different observations with the same spec\_obj\_ID must share the output class)
- **class** = object class (galaxy, star, or quasar object)
- **redshift** = redshift value based on the increase in wavelength

- **plate** = plate ID, identifies each plate in SDSS
- **mjd** = Modified Julian Date used to indicate when a given piece of SDSS data was taken
- **fiberid** = fiber ID that identifies the fiber that pointed the light at the focal plane in each observation