Using the random forest classifier, we can distinguish possible quasars from stars with a 89.83% accuracy.

Using Machine Learning to Improve Efficiency of Quasar Candidates Selection Based On Photometric Data

Qiongwen Mao (Shirley) and Prof. Eilat Glikman

Big Picture

Quasars are extremely luminous astronomical objects that are empowered by supermassive black holes. Studying quasars is crucial for understanding the makeup of the universe.

To identify quasars from other celestial objects, however, requires spectroscopic observations for each objects out of thousands of individual sources, one at a time. Thus, the efficient selection of candidates to be observed is very important.

My projects focuses on using Random Forest Classifier machine learning model to identify quasars prior to pursuing spectroscopic observation, thus improving the efficiency of the selection of quasar candidates.

Current Efforts

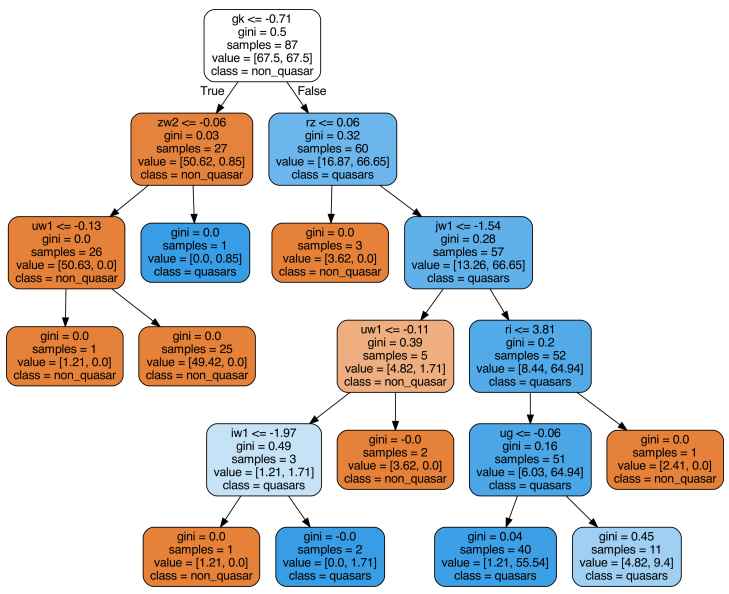
In this project, we collect the luminosity measured at different wavelengths of each celestial object, ranging from ultraviolet to near-infrared and measure the luminosity difference amongst wavelengths.

We use the luminosity difference, termed “color,” as the feature inputs for Random Forest Classifier model to classify quasars and non-quasars objects.

The model is currently trained on 194 labelled data points, consisted of each object’s color features and their Petrosian radius.

Random Forest Classifier is classification model that consists of many decision tree models. It reduces the variance by using different portion of the training data and reduces correlation between trees by randomly select features.

Below is a visualization of one decision tree from my model.



The confusion matrix below shows the performance of my current trained model. “0” indicates non-quasar objects, and ”1” indicates quasars