Stellar Objects Classification

Colin McMorrow

**Abstract**

The goal of this project was to leverage data science to build a model that could accurately classify stellar objects from scans of the sky. With instruments becoming more sensitive and measurements more precise, new objects are being discovered every year that have not been seen before. This predictive model could save huge amounts of human energy and resources by automatically identifying objects as they are picked up by scanners. Properly identifying new stars also could point astronomers in the right direction looking for new solar systems and potentially habitable planets.

**Design**

The data for this project was obtained from the Sloan Digital Sky Survey (SDSS), from the 17th data release. The primary task was to explore the features and identify data values that have strong separation by each class, and then construct a predictive algorithm. Once a feature selection was chosen, I trained a few baseline models to get an idea of how the model would perform, and confirm which features were believed to be the strongest predictors. Next, I expanded to more complex models such as random forests, and tested the final models on out-of-sample holdout data.

**Data**

The data was obtained from the Sloan Digital Sky Survey (SDSS) on Data Release 17. There are 100 thousand entries containing both spectral data, equipment data, and scan-related data. U, G, R, I, and Z and all light-absorption features through different wavelength filters. Redshift is a term indicating the speed at which and object is moving away from Earth. Other ID features are related to the equipment used and date of the scan that picked up said object in space. Each row is one individual object recorded.

**Algorithms**

Pandas, numpy, seaborn for data cleaning and visualization

Sklearn for classification models:

* kNN
* Logistic Regression
* Random Forest

Sklearn for scaling and class imbalance

Sklearn for metrics:

* Accuracy
* Confusion matrix
* F1 score

**Communication**



