1. Goals of Classification

Our primary interest with respect to the classification of radio loud quasars is to identify as many radio loud quasars as possible, with the greatest possible accuracy. In other words we wish to create a model with a large percentage of the radio loud quasars identified alongside a low number of false inclusions. Using such a model will allow a larget set of observations to be properly split into a subset consisting of primarily radio loud observations, a valuable technique for further research. With radio loud being the negative class, we will be primarily focusing on maximizing the negative predictive values (npv),the ratio of true negatives to all negatives, of our models.

2. Construction of Basic Classification Models

We begin with rudimentary models trained on our training data subset. These models include Support Vector Machines (SVMs), Linear Discriminant Analysis (LDA) models, and Neural Nets (NNs) of varying architecture.

Our first step in the construction the majority of our classification models was normalizing our training data to lie in the range (0,1). Not only did this allow for the faster convergence of our models but also slightly increased overall performance.

The first constructed models were SVMs, using all of our main 20 predictors. In this case the non-normalized training data was used as to not interfere with the maximal margin classifier when deciding on the optimal boundary. Various kernels were implemented for the formation of our decision boundary, including linear, polynomial (up to order 5), and radial kernels. All tested SVMs converged to our baseline model of only radio quiet predictions.

Our next class was LDA models. These models were trained on the main predictors excluding the 5 errors to reduce collinearity. With model selection using 10-fold cross validation, the chosen model performed poorly, only achieving an npv of 0.053, classifying a large majority of observations as positive and misidentifying many observations as negative.

The last attempt using basic models was a simple feed-forward neural net with backpropagation. This NN was trained on all main predictors with a single hidden layer of between 4 and 30 neurons. Similarly to the SVMs, even when varying the architecture of the neural net, all models converged to the baseline model.

3. Remedying Baseline Convergence

We noticed a clear issue for any models which trained for any moderate amount of time, they all converged to the baseline model. After trying multiple class imbalance workarounds (over\undersampling and class weighting) we settled on a data imputation method called SMOTE, or Synthetic Minority Over-Sampling Technique. Unlike standard under and oversampling techniques, SMOTE allows for the imputation of new data points in order to boost the representation of the minority class, rather than repetition or reducing the size of the training set. SMOTE imputes new data points by randomly choosing a minority class observation, as well as one if its neighbors within a set range of neighbors. SMOTE then selects a random point in the feature

space along the vector between the randomly chosen point and its randomly chosen neighbor, effectively creating a new observation within the minority class.

With the imputation of approximately 7,000 minority class data points, SMOTE in conjunction with the SVM model using a radial kernel achieved an npv of 0.471, a vast improvement over previous models. As for our LDA models, even with SMOTE imputation the maximum npv achieved was 0.118. Similarly, the best performing neural net had 14 neurons in its hidden layer, achieving an npv of 0.194. With our best performing model only having an npv of 0.471 we decided to increase the complexity of some of our models.

4. Deep Neural Net

Due to the highly customizable nature of neural net architectures we focused on optimizing parameters on a Deep Neural Net (DNN). The main difference between this model and our previous neural net being the number of hidden layers, which we varied between 2 and 3 and consequentially the total number of neurons, which varied between 4 and 24 in each layer. Training this neural net in conjunction with SMOTE yielded an npv of 0.674, meaning that approximately 7 out of 10 predicted radio loud quasars were predicted correctly.

5. Analysis of Classification Models

Constructing classification models as introduced in this paper is a small step forward in terms of dealing with the large amounts of data we are receiving from observatories around the globe. In the future, with the use of deeper and more robust methodologies it is possible that we can accurately subset vast amounts of data into individual pieces to be delved into much more deeply within respective specializations. In this case, our classification models are not quite strong enough to be useful in such ways but the idea shows promise.