### Introduction

- The Fermi Gamma-Ray Space Telescope is used to detect high-energy photons produced by astronomical objects, among which are gamma-ray-emitting BL Lacertae objects, or BL Lacs. BL Lacs have beamed jets of matter pointed directly towards the Earth, which are streamed from black holes.
- However, BL Lacs are difficult to identify due to their spectra’s similarities to other classes of astronomical objects.
- The goal of this analysis is to accurately classify BL Lacs, as well as determine the predictive variables most useful for said classification.

### Analysis

- We attempt to learn a number of classifiers.
  - The methods we use are Logistic Regression, Forward Subset Selection, Classification Tree, Random Forest, Gradient Boosting, K Nearest Neighbors (KNN), and Naive Bayes.
  - We use Area Under Curve (AUC) as our metric of model quality.
  - AUC is a measure of how well a model performs over a range of classification thresholds.
  - Possible AUC values range from 0 to 1, and higher values indicate better predictive ability.
  - The highest AUCs are from logistic regression, logistic regression with forward selection, and random forest.
  - Logistic regression models the probability that we would sample data of one of the classes, given specific values of the predictor variables.
  - Forward Selection chooses the optimal predictor variables by adding them one at a time and performing logistic regression at each step.
  - Random forest aggregates decision tree constructed using subsets of predictor variables and bootstrapped samples of the dataset.
- We select the Random Forest for our final predictions because it has the highest AUC value.
  - Optimal threshold for random forest: 0.709.
  - MCR: 0.1988

### References

- Freeman, P. E. 2021, online at https://github.com/pefreeman/36-290/tree/master/PROJECT_DATASETS/FERMI

### Conclusion

We were successfully able to construct a classifier using the Random Forest model that classifies objects in both classes relatively well, with an MCR 10% lower than if we were to simply classify all objects as non-BLLs.