Classification of AGNs and Pulsars using Machine Learning techniques

Aakash Bhat and Dmitry Malyshev

FAU Erlangen-Nurnberg

Juli 2019

Table of contents

Outline

- Create machine learning algorithms capable of classifying AGNs and Pulsars
- Use 4-year Fermi LAT catalog to train and test the algorithms
- Apply the algorithms on the newly released 8-year list

Outline

Types of sources and algorithms

- Use the classified AGNs (BL lacs etc) and Pulsars
- Find the appropriate "features" to use
- Algorithms include: Random forests, Logistic regression, Neural Networks
- Estimate performance by focusing on individual models (Number and depth of trees in forest based models, Number of layers in Neural networks)

Testing and Predictions

- Use the classified sources from 8 year list to test algorithms
- Predict for unassociated sources in 8 year list
- Further predictions on sub-classes

Outline

- Total of 1905 sources classified in the 4 year list
- Features: Flux, uncertainty, Significant curvature, spectral index, Hardness ratios, Galactic latitude
- 70% training and 30% testing
- $hr_{ij} = \frac{EnergyFlux_j - EnergyFlux_i}{EnergyFlux_j + EnergyFlux_i}$

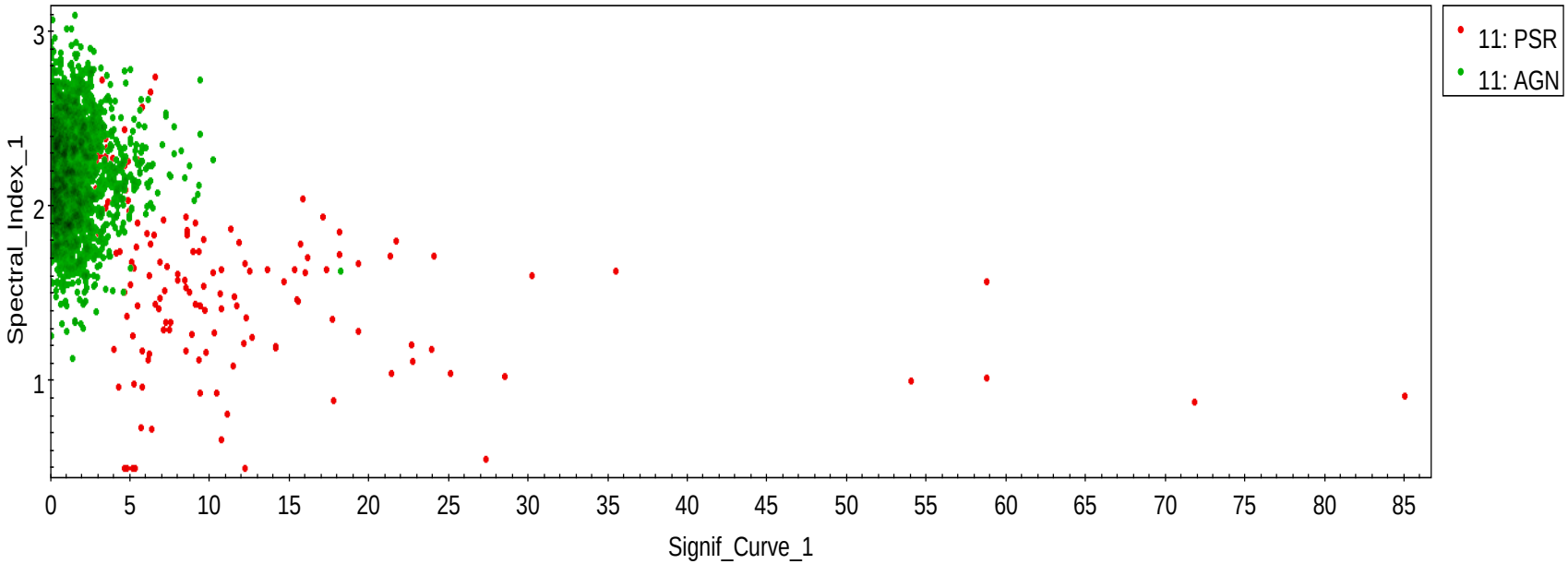


Figure: Motivation for classification

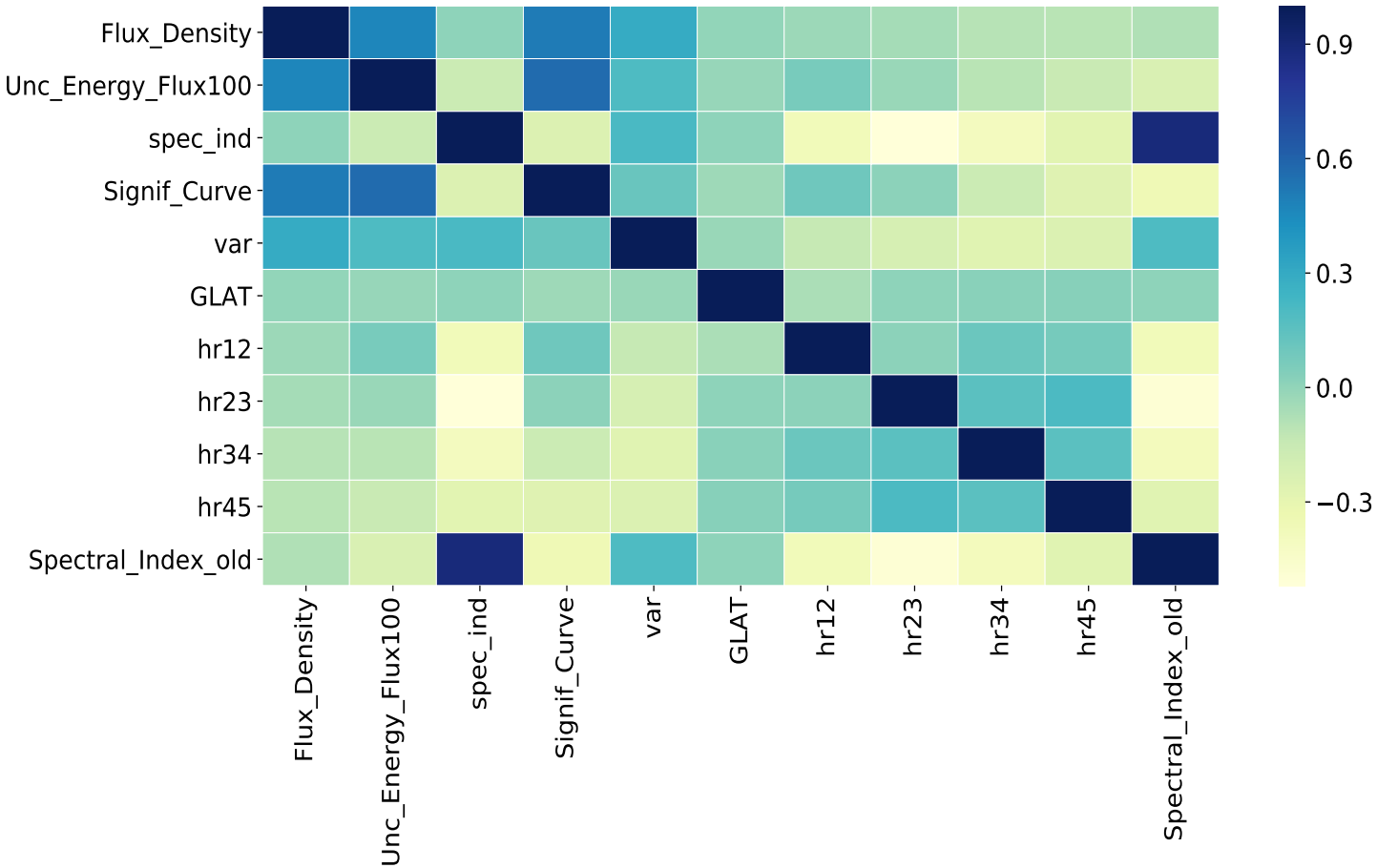


Figure: Correlation Matrix

tweaking

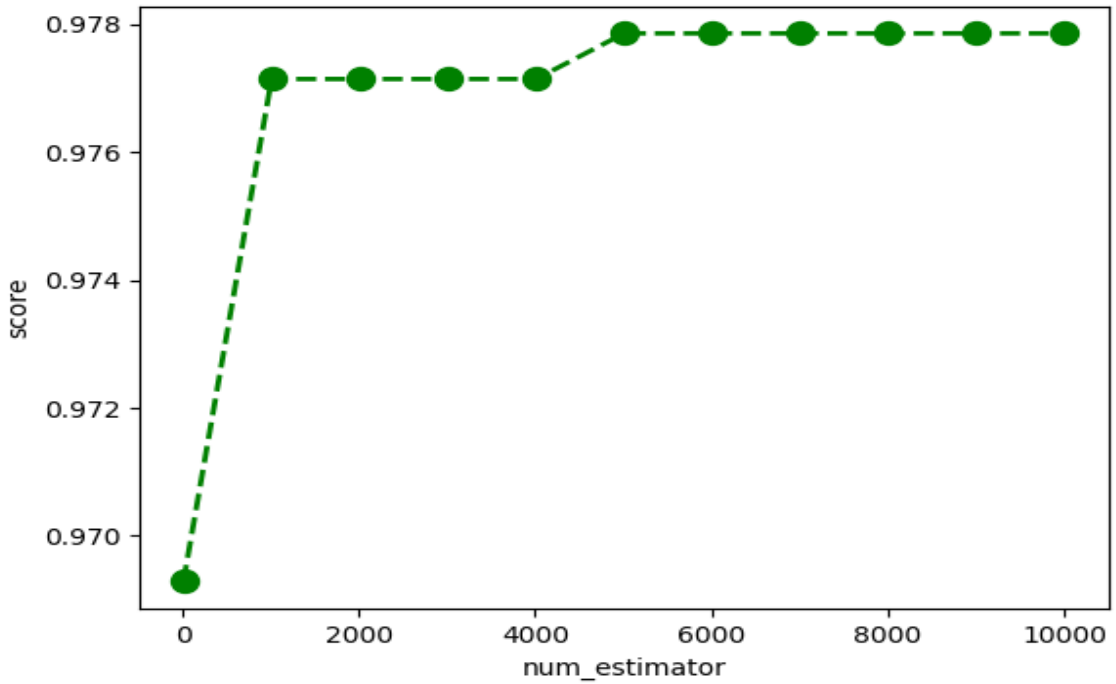


Figure: Score vs. Number of trees

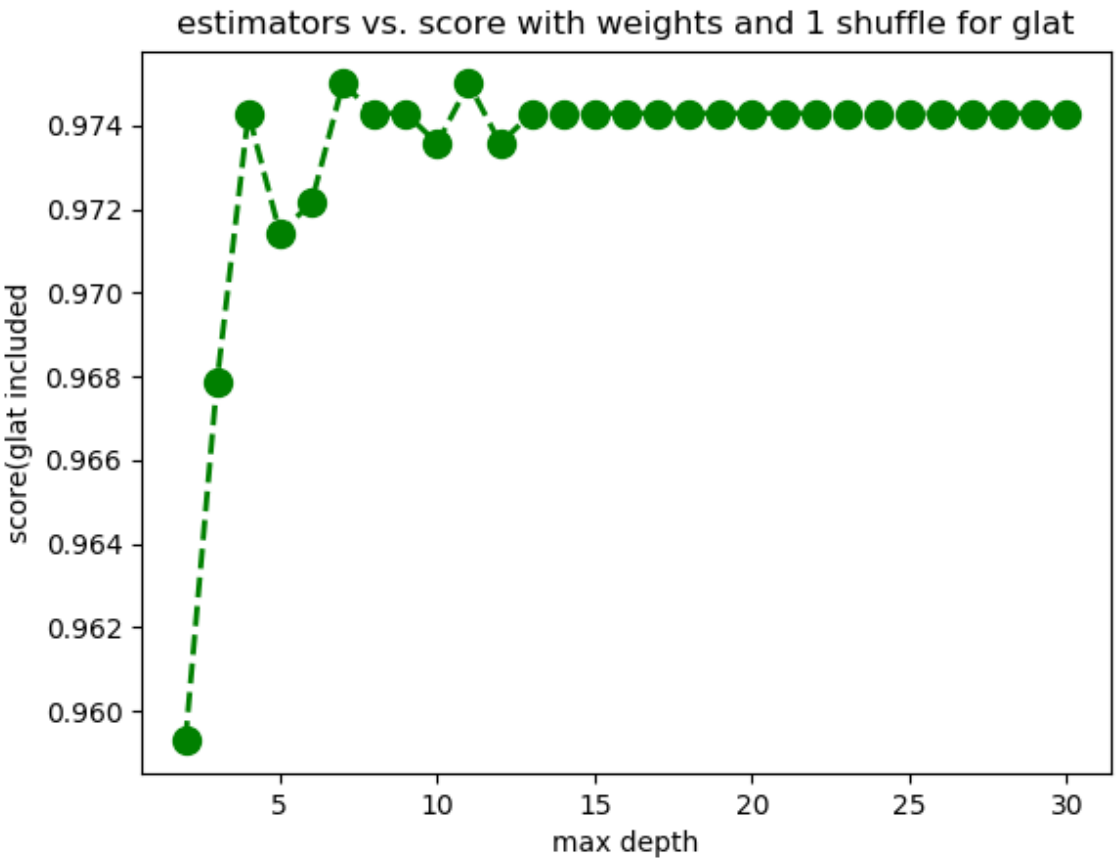


Figure: Score vs. Maximum depth of trees

Algorithm	Percentage
Random Forest	97.03
Extra Trees Classifier	97.03
Gradient Boost	96.79
Ada Boost	97.53
Logistic Regression	95.06

Table: Scores for training data

Algorithm	Fluxden	Uncertainity	Index	Signifcurv	Var	GLAT	hr12	hr23	hr34	hr45
Random Forest	0.	0.	0.106	0.287	0.1006	0.062	0.048	0.058	0.0605	0.275
Extra Trees	0.	0.	0.152	0.340	0.086	0.077	0.064	0.059	0.0747	0.144
Gradient Boost	0.	0.	0.021	0.528	0.308	0.020	0.006	0.034	0.027	0.053
Ada Boost	0.	0.	0.12	0.16	0.12	0.26	0.1	0.06	0.04	0.14

Table: Importance of Features

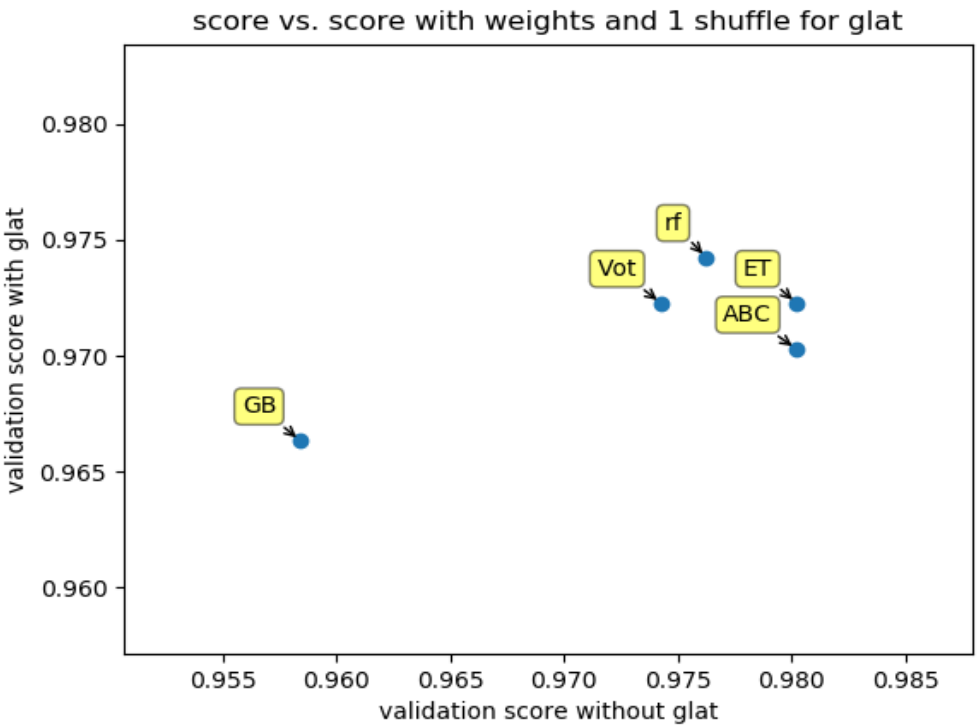


Figure: Scores with and without GLAT

Testing Data

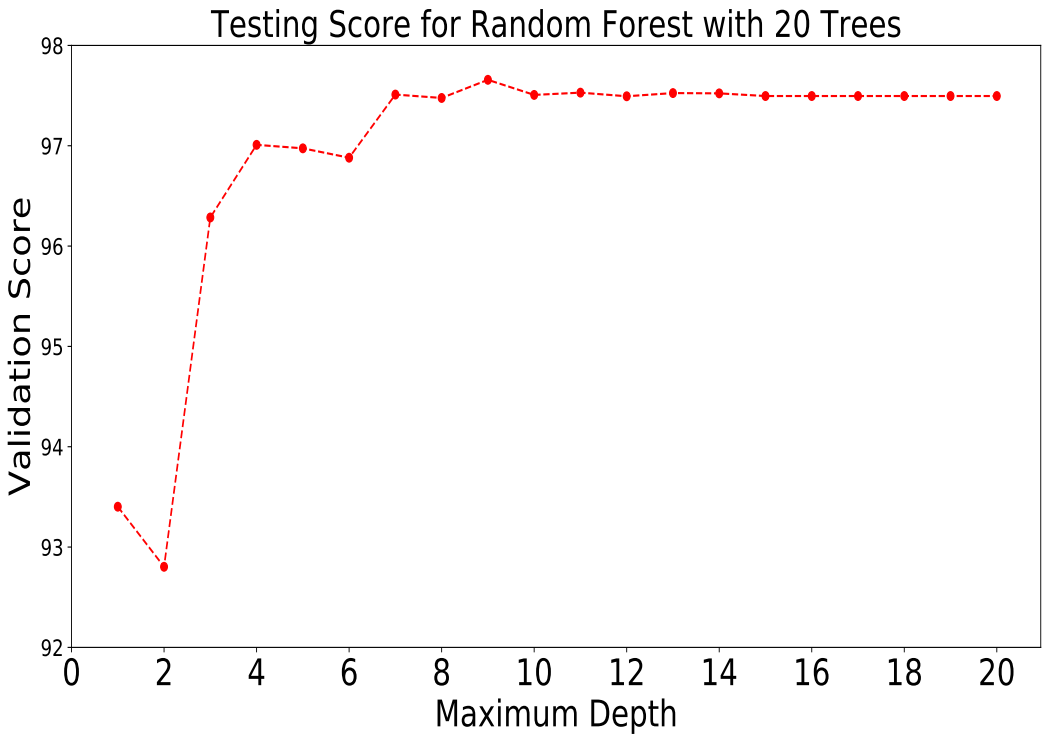


Figure: Dependence of scores on maximum depth of Random Forests

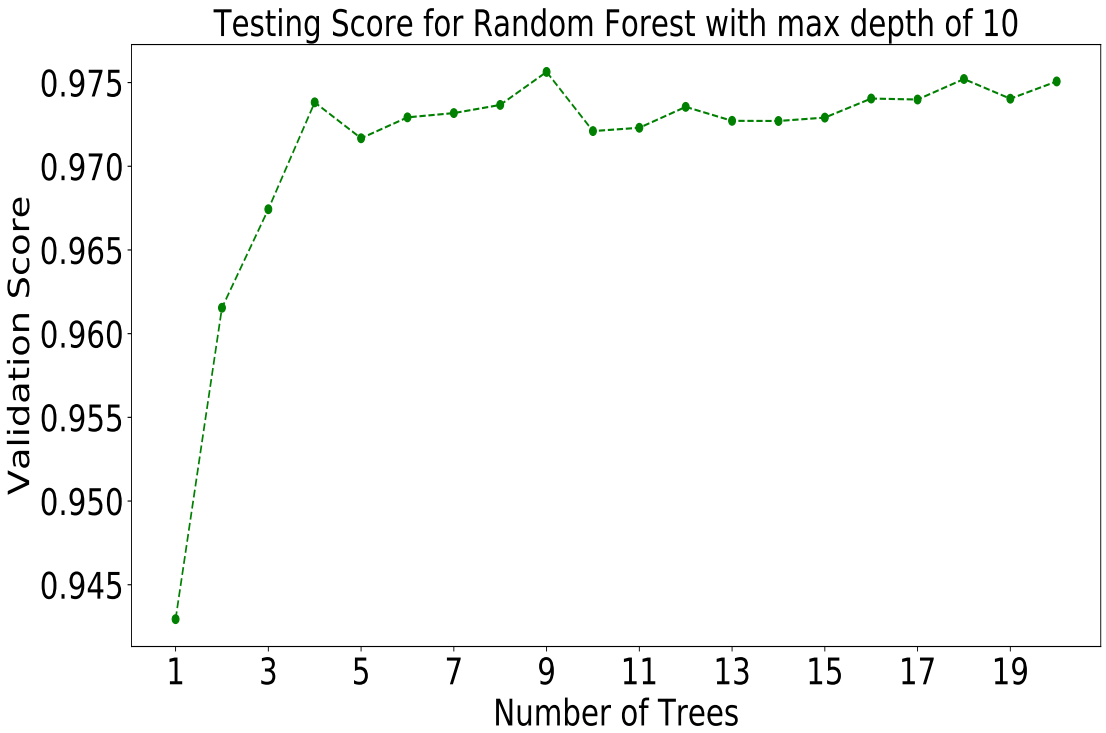
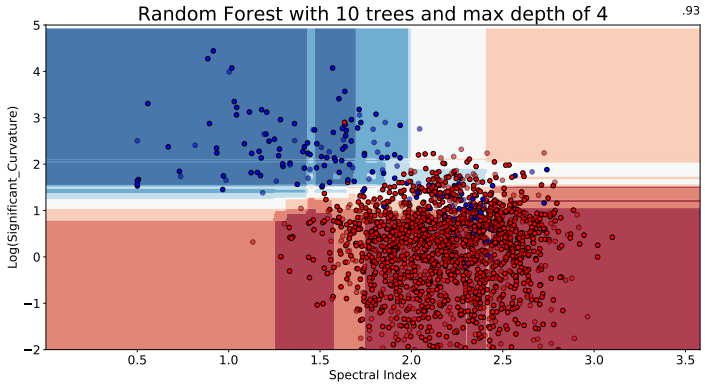


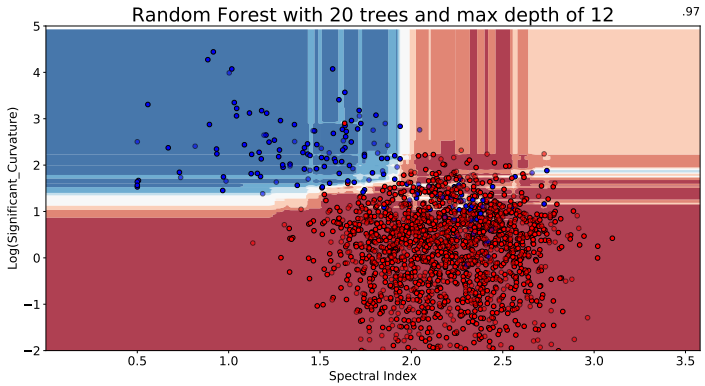
Figure: Dependence of scores on number of trees for Random Forests

Similar Studies for two other Classifiers

Classification Domains



(a) Classification Domains for RF with 10 trees and 4 Max Depth



(b) Classification Domains for RF with 20 trees and 12 Max Depth

- Testing on unassociated data with classes from FL8Y

Algorithm	Percentage
Random Forest	95.45
Extra Trees Classifier	96.5
Gradient Boost	96.15
Ada Boost	96.15
Logistic Regression	88.11

Table: Scores for actual testing data

Neural Networks

- Neural Networks with 1-2 hidden layers
- Hyperbolic Tan as activation function and Sigmoid as the final layer
- Give accuracy above 90% (Max accuracy of 98%).

- Predicting on Unassociated Data

Algorithm	Accuracy	Expected Pulsars (1056 sources)
RF (50,12)	98.30	148
BDT(50,15)	96.86	184
NN(20 N, 200 E)	98.09	213

Table: Performance and Prediction of three types of classifiers

Future Work

- Predict for fl8y unassociated sources and match with other catalogs
- Predict for further sub-classes (BLLACS and FSRQs, yng and ms PSR)