## 1. Performance estimates of different ML classiers

Use the known classifications in the 3FGL catalog to estimate the performance of different ML classification algorithms: DT, BDT, random forest, ensemble of three, neural nets.

Other algorithms?

1. Calculate the score and loss rates for Galactic and extragalactic sources.

* Galactic vs. Extragalactic
* AGN vs. Pulsars
* Young vs. msp pulsars
* FSRQ vs. Bllacs

1. Study the performance of the algorithms by changing complexities, i.e., the depth of the trees, the number of trees in the random forest etc.
   * Number of trees
   * Depth of trees
   * Weighting
   * Reshuffling the input (expected to not fluctuate with correct weight)
   * For NNs: Number of hidden layers
   * Weighting
   * Activation function (usually softmax)
   * Probabilities
2. Try different weighting of the training samples: uniform weights, boost the weights of the Gal sources (since there are fewer of them) (ties with 2.c).

Features in classification (some are log): spectral index, flux, signif\_curve, var, unc (so far)

To add: flux differences (4 more)

## 2. Application to 3FGL and prediction

1. Calculate the most probable classes for the unidentified sources.
   * 1. Probabilities
     2. Direct classes
2. Compare the predictions for different classifiers.
3. Search for outliers (among sources with known classes).
4. Try to separate AGNs in classes, e.g., FSRQs and BL Lacs.

## 3. Comparison with the FL8Y

1. Find out if some of unidentified sources in 3FGL are identified in FL8Y, compare with the predicted classes.
2. Use FL8Y for training and make predictions for the unidentified sources in FL8Y.
   * 1. Use 3fgl for training and predict on Fl8y as a test first and vice versa.
     2. Then only use fl8y as training, testing, and prediction.

References:

Pablo Saz Parkinson:

https://arxiv.org/abs/1602.00385