MAIS 202 Deliverable 1: Predicting Pulsars

Pulsars are a rare type of Neutron star that produce radio emission detectable here on Earth. They are of considerable scientific interest as probes of space-time, the inter-stellar medium, and states of matter. The following project uses 17898 data points collected during the HTRU survey to predict pulsar stars based on observational features.

1. Dataset:
   HTRU2 Data set https://archive.ics.uci.edu/ml/datasets/HTRU2#
   Pulsar candidates during the High Time Resolution Universe Survey.

2. Methodology:
   a. Data Preprocessing: Divide data into training/test set using the 80/20 rule. Data is already cleaned up to be used, we just need to separate the features and the class. Here we have 8 features (Mean of the integrated profile, Standard deviation of the integrated profile,Excess kurtosis of the integrated profile, Skewness of the integrated profile,Mean of the DM-SNR curve, Standard deviation of the DM-SNR curve, Excess kurtosis of the DM-SNR curve, Skewness of the DM-SNR curve) and 1 class (pulsar/not pulsar)
   b. Machine Learning Model: We hope to build a model able to predict whether an observed astrological object is a pulsar or not. To do so we will try different types of classification algorithms namely Naïve Bayes, Random Forest Classifiers, Neural networks, and others.
   c. Evaluation Metric: Since the pulsars are a minority in the data set (outnumbered 10 to 1) and would like to favor false positives instead of false negatives, we will use confusions matrices as our evaluation metric. This is because pulsars are rare events, so theoretically we prefer to analyze a candidate that isn't a pulsar instead of misclassifying one. This is at the expense of the accuracy of our model since high accuracy across the data set isn't really the purpose of the model, rather high accuracy for true positives.

3. Application:
   We would like to build a webpage where a user can input values for the features used in the model and it will tell them whether the observed object is a pulsar or not. Ideally, it will also tell them the confidence level of the prediction and also give the possibility to choose the type of algorithm to use for the prediction/ and maybe give the possibility to predict using less features. I am unsure if all of this is possible but I'll figure it out as we go.