Team: Shane Kelly

Deliverable 1

Choice of dataset: I chose the Predicting a Pulsar Star dataset (available at https://archive.ics.uci.edu/ml/datasets/HTRU2). I chose this particular dataset due to my personal interest in astronomy. I have the necessary background knowledge to understand the data with relative ease, and I have taken multiple astronomy classes at McGill so I know it is a topic that I enjoy.

Methodology: The dataset is feasible and does not seem to require large amounts of editing in order to work with. It is separated into 9 columns. The first 8 are properties of the observed stars, while the 9th is whether or not that data was actually a pulsar star or not. All 8 columns of physical data should likely be used for maximum accuracy, however the first and fifth columns will likely be the most important, as they are the results of the data while the rest represent the standard deviation and skewness of the curve.

I would like to use logistic regression as the model for this project. Considering the fact that we are classifying data in a binary fashion, as in the data either represents a pulsar or does not, being able to show a percentage certainty of the answer would be a good function. Other classification methods could be used as well, however considering the fact that there is only one class to consider, logistic regression should be the most efficient.

Application: In the case of the pulsar star detection application, I believe it would be difficult to implement in a web application in any way other than allowing the user to input their data through text and then indicating if its a pulsar with the percentage certainty, however also allowing for the uploading of csv files to handles large amounts of data at once would be even better for the practical uses of this application.

As for the outputs, I believe that there are two options. One could simply be a text display telling the user whether or not the input is a pulsar, or formatting a csv with all the output values would be likely even more useful for any automated star checking that astronomers would wish to do. In real world circumstances, I believe that creating an API that can be called to process data within your own code would be the best use of this form of program, however at this point I really do not know how I would go about doing that.