**Name: Date: 16/01/2020**

**Dataset:**

HTRU2 is a data set which describes a sample of pulsar candidates collected during the High Time Resolution Universe Survey .

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.

As pulsars rotate, their emission beam sweeps across the sky, and when this crosses our line of sight, produces a detectable pattern of broadband radio emission. As pulsars rotate rapidly, this pattern repeats periodically. Thus pulsar search involves looking for periodic radio signals with large radio telescopes.

Each pulsar produces a slightly different emission pattern, which varies slightly with each rotation . Thus a potential signal detection known as a 'candidate', is averaged over many rotations of the pulsar, as determined by the length of an observation. In the absence of additional info, each candidate could potentially describe a real pulsar. However in practice almost all detections are caused by radio frequency interference (RFI) and noise, making legitimate signals hard to find.

**Input variables**

1. Mean of the integrated profile.
2. Standard deviation of the integrated profile.
3. Excess kurtosis of the integrated profile.
4. Skewness of the integrated profile.
5. Mean of the DM-SNR curve.
6. Standard deviation of the DM-SNR curve.
7. Excess kurtosis of the DM-SNR curve.
8. Skewness of the DM-SNR curve.

**Output variable**

* Class

**Questions:**

1. **Number of samples in the dataset?**
2. **What type of problem are we solving, regression or classification?**
3. **Show a histogram of the output variable and one input variable?**
4. **What is the input variable that shows better correlation with the output variable?**
5. **What percentage of samples are you using for training?**
6. **What non-neural network model will you be using?**
7. **Show two graphs (one for training data and one for test data) comparing the predicted values from your non-neural network model with the dataset values.**
8. **What activation function will you be using for the last layer of your neural network? Why did you choose that activation function?**
9. **What are the characteristics (neurons, number of layers, …) of the Neural Network you are using? You can show the summary that appears when you create the model.**
10. **What is the error (whichever measure you have chosen to train your model) for the training data and for the test data?**