**PREDICTING A PULSAR STAR**

**- Problem statement**

**Overview**

A pulsar is a highly magnetized rotating neutron stars that emits beams of electromagnetic radiations out of its magnetic poles.[]](https://en.wikipedia.org/wiki/Pulsar#cite_note-2) This radiation can be observed only when a beam of emission is pointing toward Earth, and is responsible for the pulsed appearance of emission. Neutron stars are very [dense](https://en.wikipedia.org/wiki/Density), and have short, regular rotational periods. This produces a very precise interval between pulses that ranges from milliseconds to seconds for an individual pulsar. Pulsars are one of the candidates for the source of ultra high energy cosmic rays

One way to think of a pulsar is like a lighthouse. At night, a lighthouse emits a beam of light that sweeps across the sky. Even though the light is constantly shining, you only see the beam when it is pointing directly in your direction. The video below is an animation of a neutron star showing the magnetic field rotating with the star. Partway through, the point-of-view changes so that we can see the beams of light sweeping across our line of sight – this is how a pulsar pulses.

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.

Machine learning tools are now being used to automatically label pulsar candidates to facilitate rapid analysis. Classification systems in particular are being widely adopted,

which treat the candidate data sets as binary classification problems. Here the legitimate pulsar examples are a minority positive class, and spurious examples the majority negative class.

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

The data set shared here contains 16,259 spurious examples caused by RFI/noise, and 1,639 real pulsar examples. These examples have all been checked by human annotators.

Each row lists the variables first, and the class label is the final entry. The class labels used are 0 (negative) and 1 (positive).

**Attribute Information:**

Each candidate is described by 8 continuous variables, and a single class variable. The first four are simple statistics obtained from the integrated pulse profile (folded profile). This is an array of continuous variables that describe a longitude-resolved version of the signal that has been averaged in both time and frequency . The remaining four variables are similarly obtained from the DM-SNR curve . These are summarised below:

*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.*

*Class*

**ACCESSING AND ANALYZING THE DATA :**

<https://drive.google.com/drive/folders/1-Qqy5Kdlymu25elCRb0S7z2lcfkY2oiN>

**Prepare a classification machine learning algorithm to get the best ROC score for the given dataset. Submission can be done using .ipynb file .**