A pulsar star is a high mass star that collapses into a neutron star and spins increasingly rapidly, emitting intense electromagnetic radiation. The beams of energy emitted by the pulsar star are detected as radio waves by our telescopes. Their unique integrated electromagnetic radiation profile and strongly periodic received signal allows for classification and flagging of pulsars from a set of noise and Radio Frequency Interference (RFI) signals.

**Data Collection**

A pulsar is a high-mass star that collapses into a dense magnetized neutron star and spins increasingly rapidly, emitting intense electromagnetic radiation from its magnetic poles. As the star and its poles rotate, the beams of energy radiated from the poles pass over us, resulting in a strongly periodic radio signal being received by radio telescopes on Earth. This is sometimes likened to a “lighthouse” effect, because the radiation can only be observed when a beam is pointed towards us, much like the way a lighthouse can be seen only when the light is pointed in the direction of an observer. The beams of energy emitted by the pulsar star are detected as extremely regular pulses (in the range between milliseconds and seconds) in radio signal (hence the name “pulsar”) by our telescopes. Pulsars are of great scientific interest because of their highly regular signal; once a pulsar has been identified, deviations from its normal period can be used as evidence of gravity waves, dark matter, and other cosmic phenomena.

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

The dataset contains 8 continuous variables that describe a longitude-resolved version of the signal that has been averaged in both time and frequency, and 1 class variable. For real pulsars, the class label is 1; otherwise, the class label is 0. The 8 features were obtained by the HTRUS using classical pulsar candidate analysis, which involves creating and analysing the integrated pulse profile and the Dispersion Measure/Signal-to-Noise Ratio (DM/SNR) curve.

**References**

https://archive.ics.uci.edu/ml/datasets/HTRU2