# Introduction

## Description

Pulsars are kind of Neutron stars and considerably interesting for scientific research. As this exceptional kind of star produces radio emissions detectable here on Earth, machine learning tools can be used to label pulsar candidates to facilitate rapid analysis automatically. Classification systems are widely implemented, considering the candidate data sets as **binary classification problems**.

## Candidate information

Each candidate is described by eight **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.

1. Mean of the integrated profile.
2. The standard deviation of the integrated profile.
3. Excess kurtosis of the integrated profile.
4. The skewness of the integrated profile.
5. Mean of the DM-SNR curve.
6. The standard deviation of the DM-SNR curve.
7. Excess kurtosis of the DM-SNR curve.
8. The skewness of the DM-SNR curve.
9. Class

## Classification task

Given the features above, our final task is to determine whether each of the samples is a Pulsar candidate or not. So, we are faced with a **binary classification problem,** and we need to build different classifiers, analyze them, and compare their performance and cost for various applications.

# Features

1. Histogram of the features.
2. Do we have outliers?
3. Do we need pre-processing?
4. Correlation analysis

# Classification

1. K-fold cross-validation
2. Main application
3. Unbalanced applications
4. Write DCF
5. PCA or No PCA
6. MVG Classifier
7. Linear Logistic Regression
8. Quadratic LR
9. Linear SVM
10. Quad SVM
11. Kernel SVM
12. GMM

# Experimental Validation

# Conclusion