

Predicting a pulsar star – MAIS 202 final project

Project Idea:

I would like to create a web app that the astronomers can use to predict the existence of a pulsar star given the 8 features. They could use the prediction to further investigate and confirm the discovery of the star.

Dataset:

UCI Machine Learning Repo - <https://archive.ics.uci.edu/ml/datasets/HTRU2> [1]

Reasons for choosing this:

- It contains 17,898 instances and 1,639 are real pulsar examples. They have been checked by human annotators. I plan to do a 90-10 split between training and testing.
- The data is consistent in terms of the data type used and does not contain any invalid entries.
- The data is real and the dataset is quite popular making it reliable.
- The dataset contains no unnecessary information, which reduces the redundancy of the data.
- The data contains 9 variables (8 features and 1 label).

Methodology:

1. Data Preprocessing

I would search for null values in the dataset and replace them with their respective mean values. In addition, the data is pretty consistent, so that wouldn't need to be paid heed too.

2. Machine Learning Model

It seems like KNN would be a suitable model for it as the variables could be used to see what stars similar to it are labelled as. It also seems like a good idea to use logistic regression.

3. Evaluation Metrics

I would use a confusion matrix along with since most of the cases aren't pulsar stars. I would find the percentage of correct predictions over total number of cases to find the accuracy of the predictions. I plan on having around 95% accuracy with the results.

4. Final conceptualization

I would build a web app where astronomers can input their query data into the interface and get a prediction.

[1] R. J. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, J. D. Knowles, Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach, Monthly Notices of the Royal Astronomical Society 459 (1), 1104-1123, DOI: 10.1093/mnras/stw656

[1] R. J. Lyon, HTRU2, DOI: 10.6084/m9.figshare.3080389.v1.