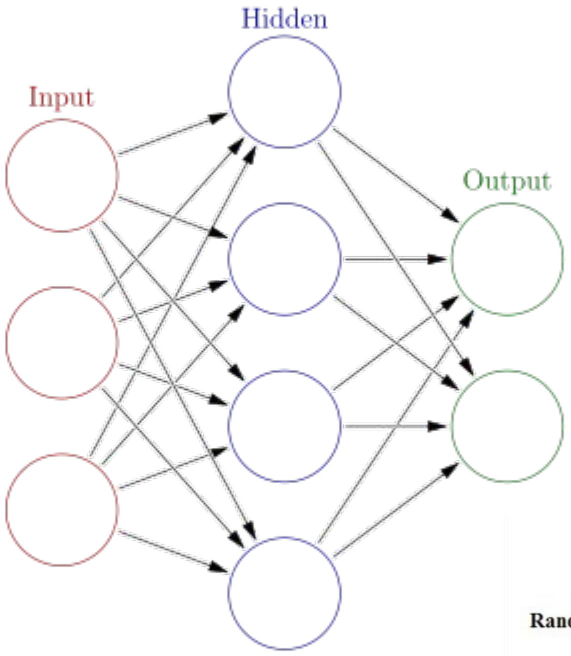
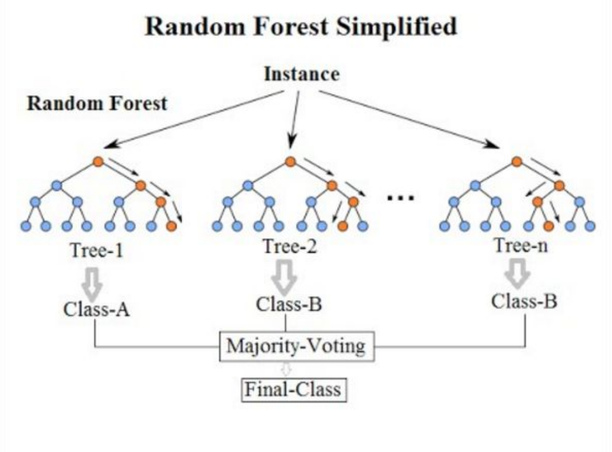
Pulsar Star Classification

by Jonathan Bardey
Flatiron School Data Science

Objective:

Classify Pulsar Candidates using Neural Networks and Machine Learning Models





The Data

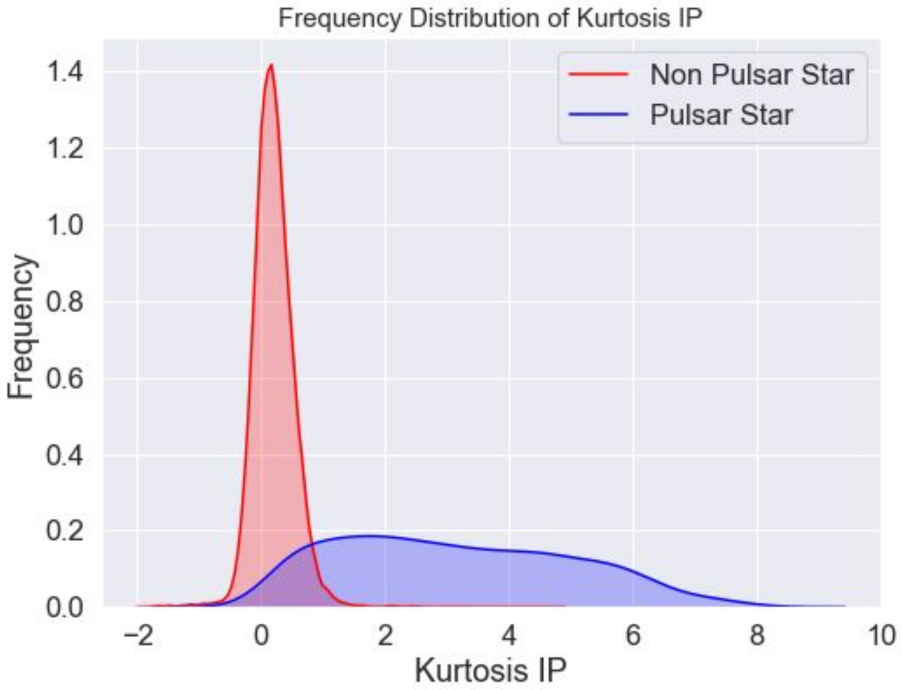
17,898 Pulsar Star Candidates

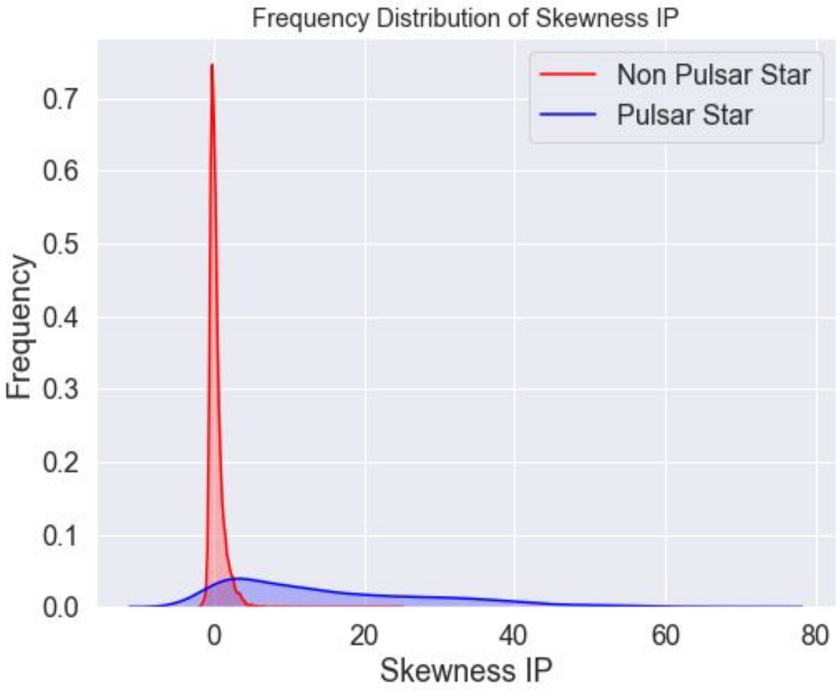
1. Mean of the IP
2. Standard Deviation of the IP
3. Excess Kurtosis of the IP
4. Skewness of the IP
5. Mean of the DM-SNR Curve
6. Standard Deviation of the DM-SNR Curve
7. Excess Kurtosis of the DM-SNR Curve
8. Skewness of the DM-SNR Curve
9. Target (1: pulsar, 0: non-pulsar)

IP: Integrated Profile
DM: Dispersion Measure
SNR: Signal-to-Noise Ratio

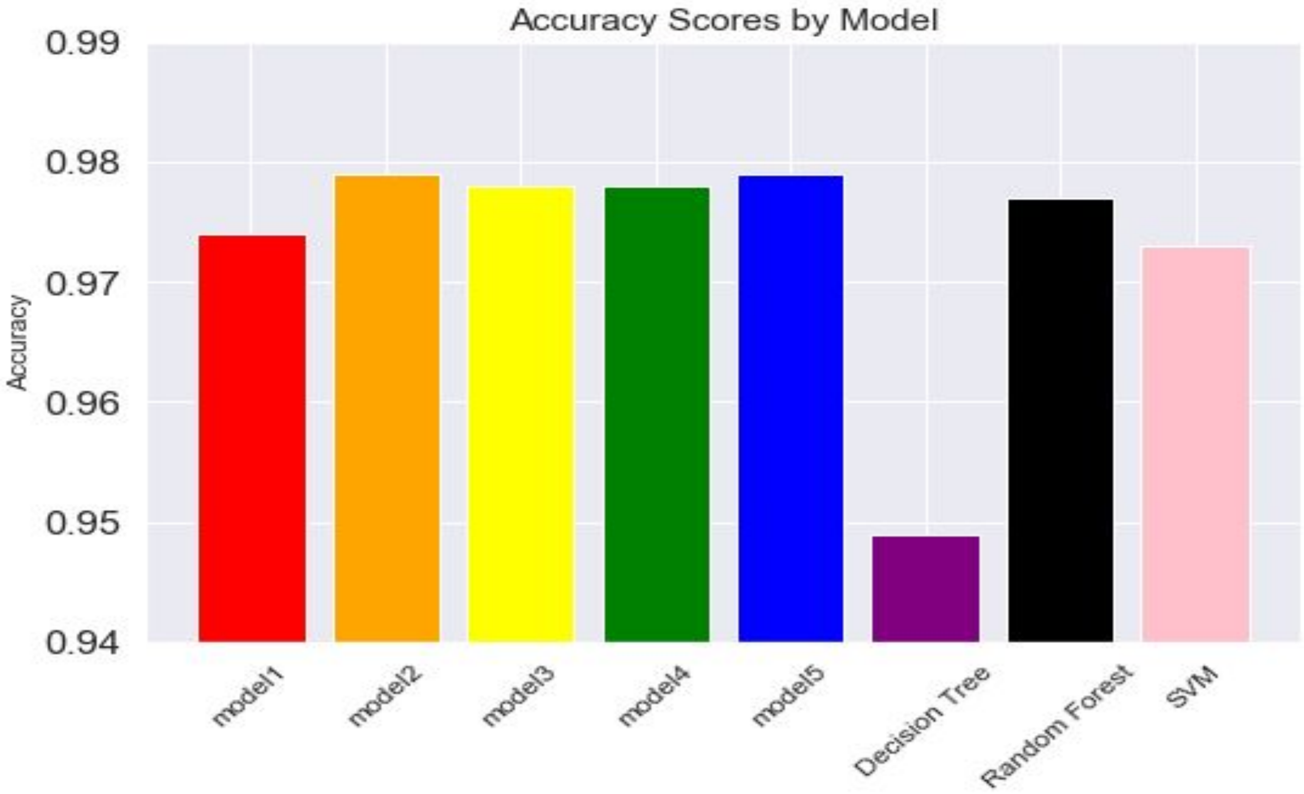
Feature Correlation with Target
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
Mean IP
STD IP
Kurtosis IP
Skewness IP
Mean DM-SNR
STD DM-SNR
Kurtosis DM-SNR
Skewness DM-SNR

Important (and Separable!) Features





Model Performance

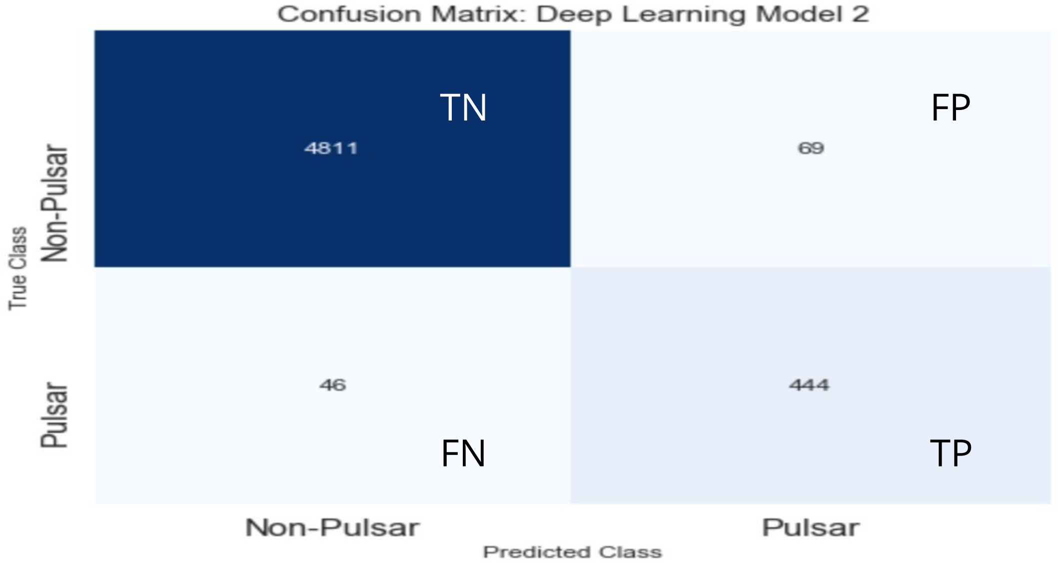


7 of 8 models between 97 and 98% Test Accuracy Score.

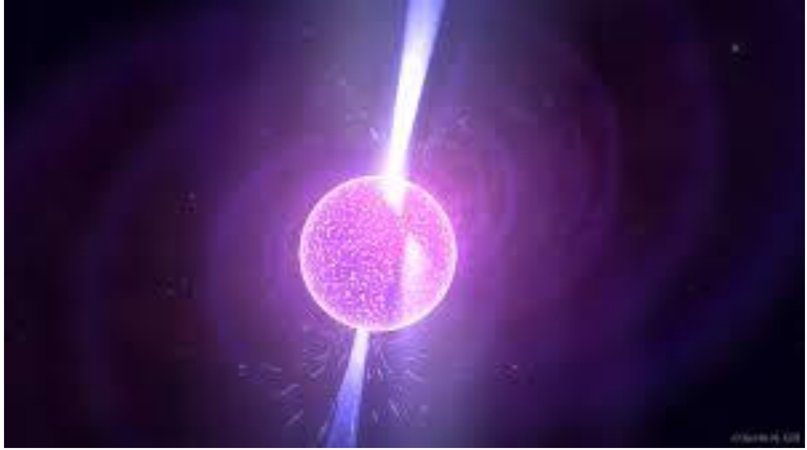
Best Model: Deep Learning Model 2

	precision	recall	f1-score	support
0	0.99	0.99	0.99	4880
1	0.87	0.91	0.89	490
accuracy			0.98	5370
macro avg	0.93	0.95	0.94	5370
weighted avg	0.98	0.98	0.98	5370

```
#Plot confustion matrix for Model 2
plot_conf_matrix(y_true, y_pred2, model_name='Deep Learning Model 2')
```



Conclusions and Future Work



- ★ Good news for the Astrophysicists! Machine Learning is very useful for classifying Pulsar Stars - almost 98% Accuracy even for more basic models.
- ★ Gather more data.
- ★ Continue to investigate overlap region and signal noise to better capture Pulsar Stars signals.
- ★ Experiment with feature selection to develop more efficient models.

Thank you!

Questions?