

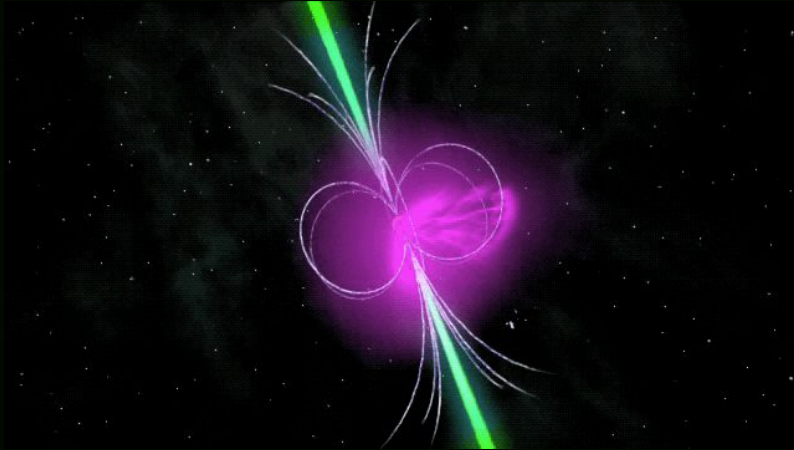


# Predição de Estrelas Pulsares: como Data Science nos ajuda a entender o Universo

07 de Dezembro de 2018

# Embasamento Teórico

# Estrela Pulsar



- Feixe estreito de radiação emitido dos polos magnéticos;
- Direções variam consoante sua rotação;
- Produz um sinal detectável por radiotelescópios aqui na terra;
- Cada pulsar possui um único fingerprint (pulse profile).

# Estrela Pulsar

Possui um **considerável interesse científico:**

- Características das galáxias
- Sondas de espaço-tempo
- Meio interestelar
- Estados da matéria

**Objetivo do projeto:**

**Desenvolver um algoritmo  
eficiente para predição de  
estrelas pulsares**

# Dados

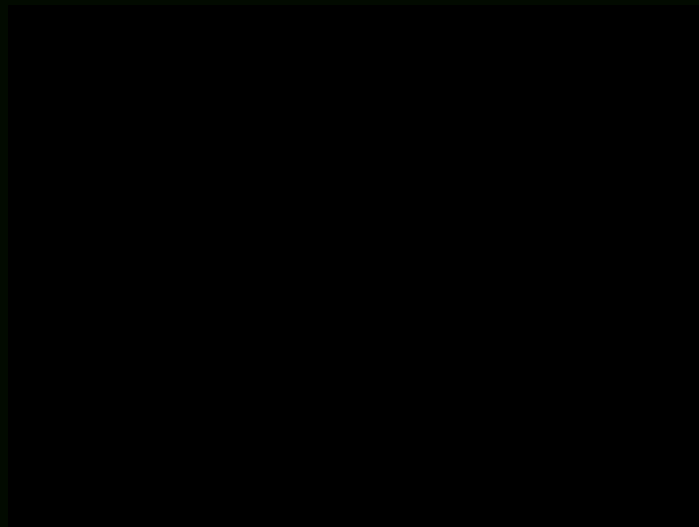
## Estatisticamente obtidos do High Time Resolution Universe Survey

Cálculos estatísticos em cima de um conjunto de sinais captados.

- Média
- Desvio Padrão
- Curtose
- Assimetria

## Atributos do sinal

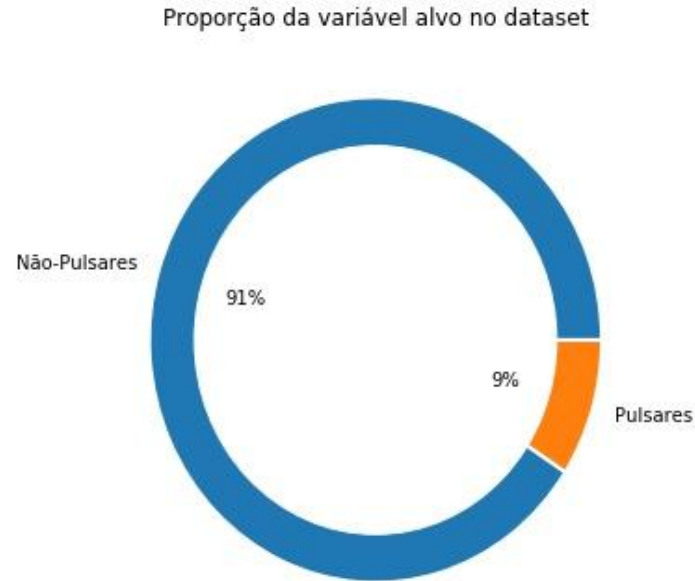
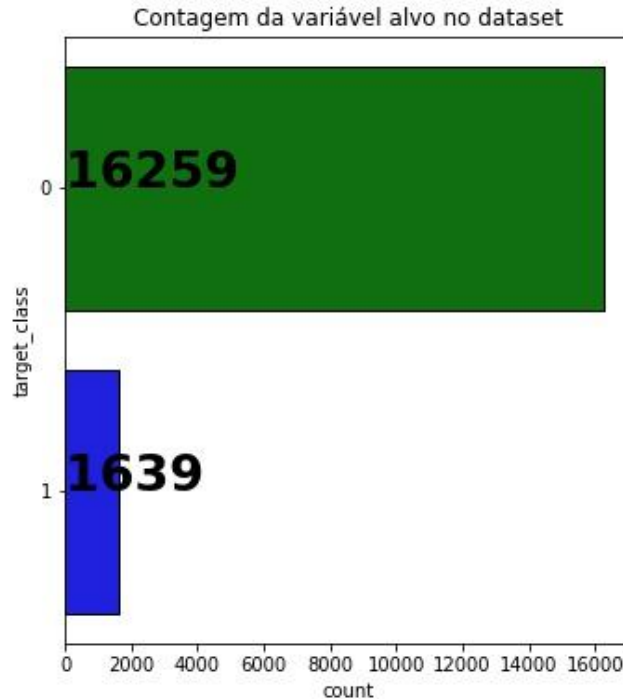
- Perfil integrado do pulso
- Curva DM-SNR



# Exemplo

	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	target_class
0	140.562500	55.683782	-0.234571	-0.699648	3.199833	19.110426	7.975532	74.242225	0
1	102.507812	58.882430	0.465318	-0.515088	1.677258	14.860146	10.576487	127.393580	0
2	103.015625	39.341649	0.323328	1.051164	3.121237	21.744669	7.735822	63.171909	0
3	136.750000	57.178449	-0.068415	-0.636238	3.642977	20.959280	6.896499	53.593661	0
4	88.726562	40.672225	0.600866	1.123492	1.178930	11.468720	14.269573	252.567306	0

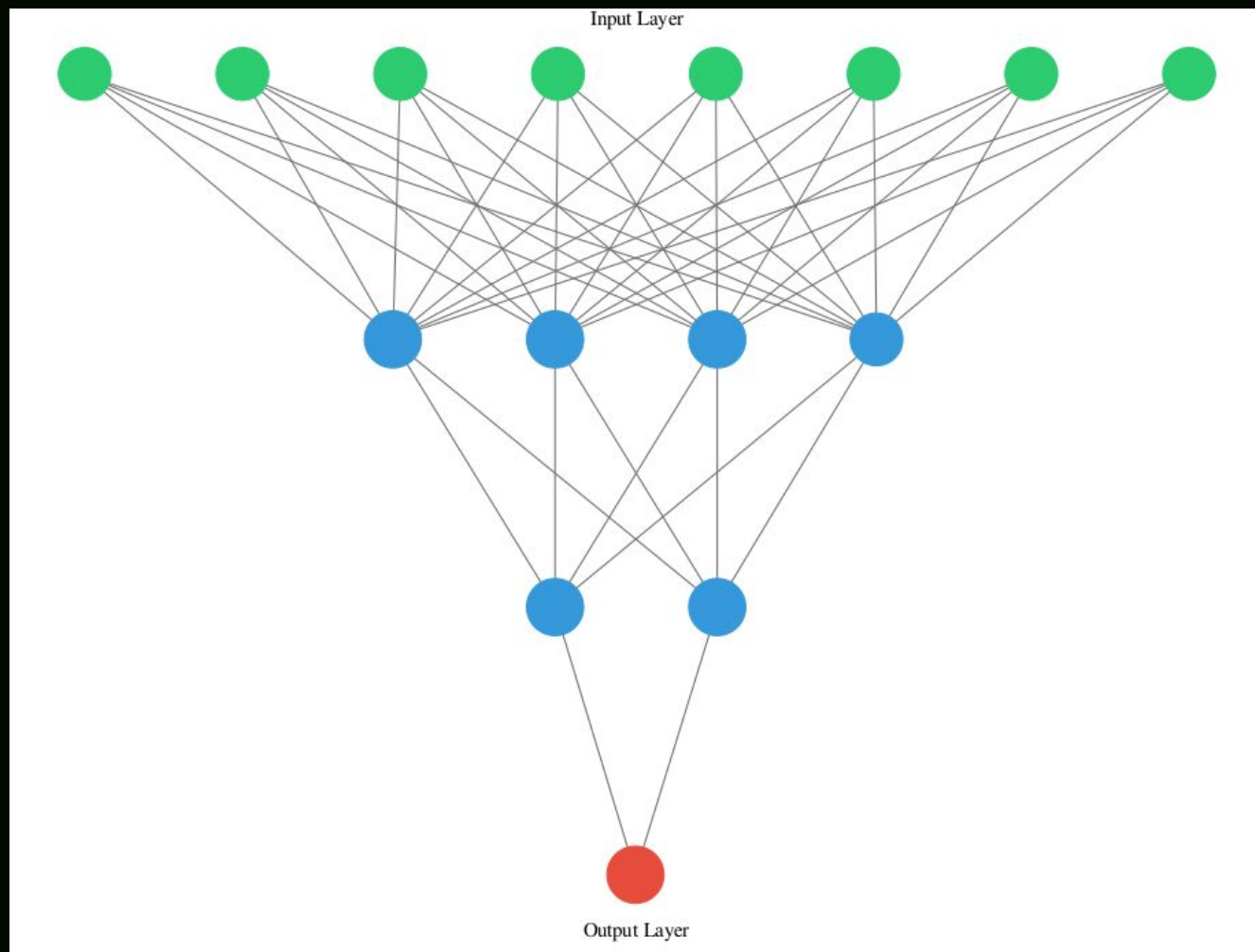
# Problema: Dados desbalanceados





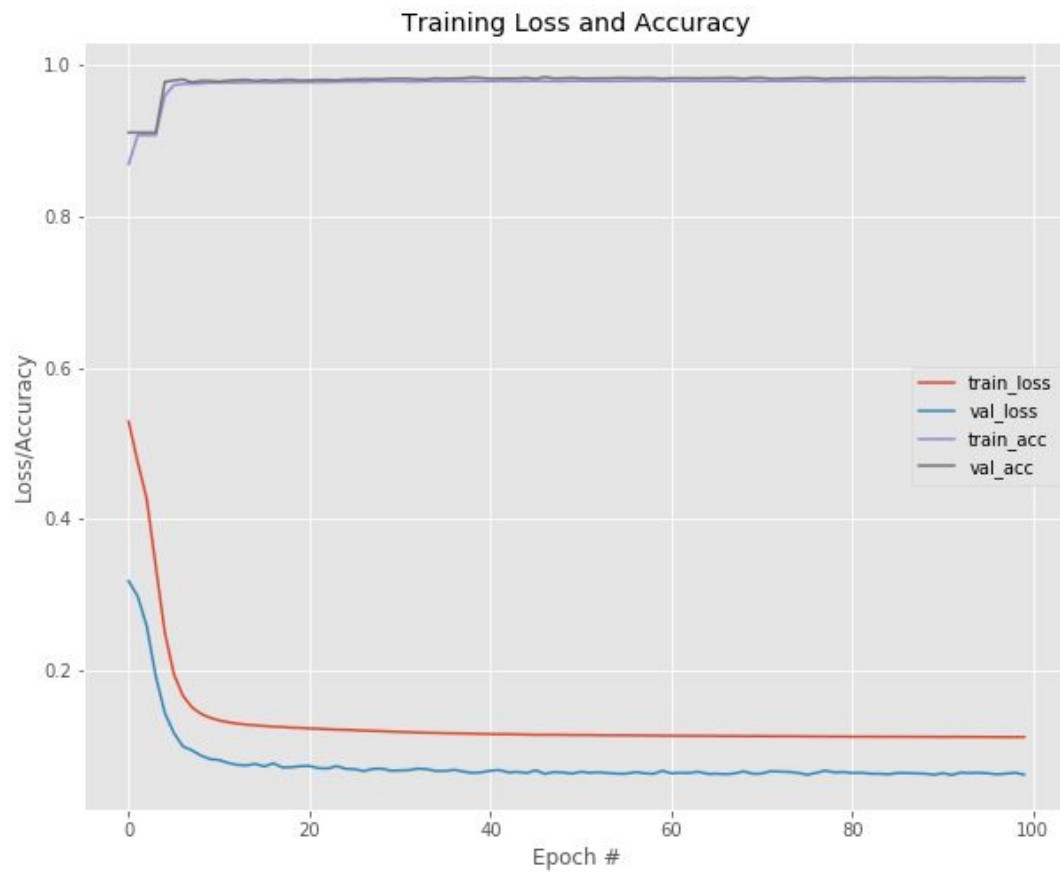
**Modelo**

# Rede Neural

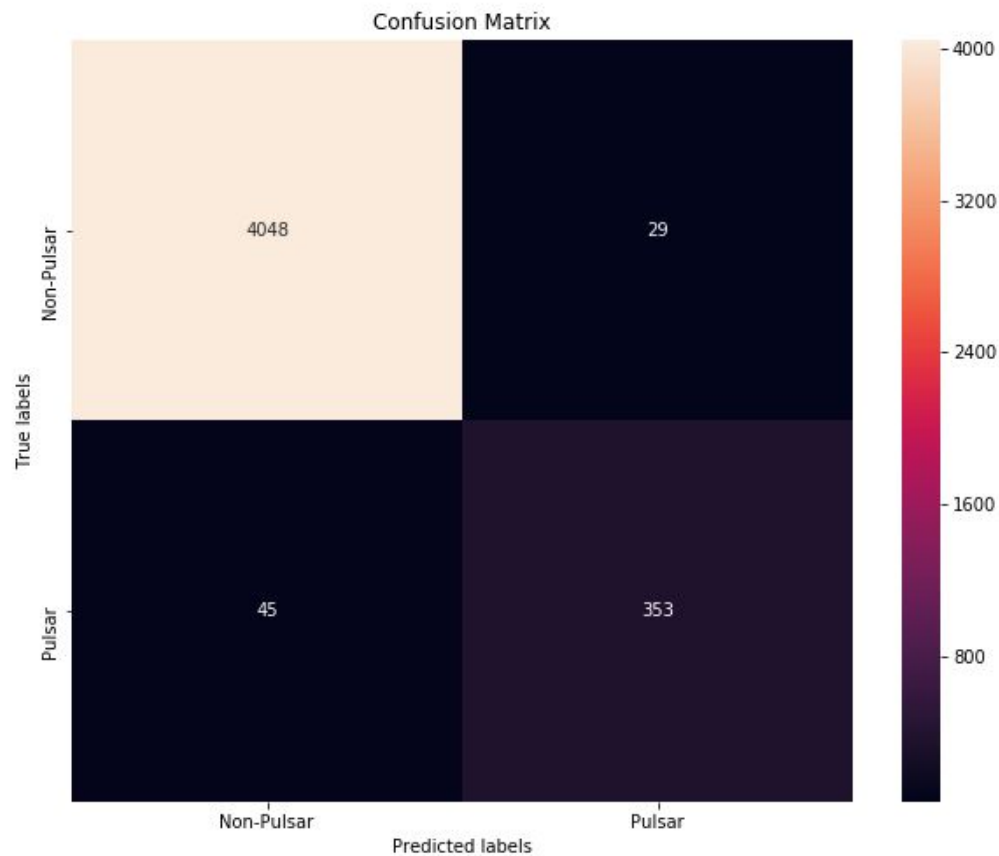


# Resultados

Acurácia de 98.34%



## Matriz de Confusão



	precision	recall	f1-score	support
Non-pulsar Star	0.99	0.99	0.99	4077
Pulsar Star	0.92	0.89	0.91	398
micro avg	0.98	0.98	0.98	4475
macro avg	0.96	0.94	0.95	4475
weighted avg	0.98	0.98	0.98	4475

## Análise de Recall

## Comparando com outros algoritmos

Modelo	Ocorrência de Falsos Negativos	Recall	Acurácia (%)
Rede Neural	45	0.89	98.34
Gradient Boosting	78	0.85	97.50
Random Forest	82	0.84	97.82
Gaussian Naive Bayes	86	0.83	93.79
Decision Tree	92	0.82	96.46
Ada Boost	93	0.82	97.63

**Dúvidas?**