

Machine Learning para el tratamiento de datos y la detección de exoplanetas mediante el método de tránsito



Universidad
Politécnica
de Madrid

**ETSI SISTEMAS
INFORMÁTICOS**

Trabajo de Fin de Grado

Grado en Ingeniería de Sistemas de Información

Curso 2019/2020



POLITÉCNICA

Autor

José Javier Gómez de Diego

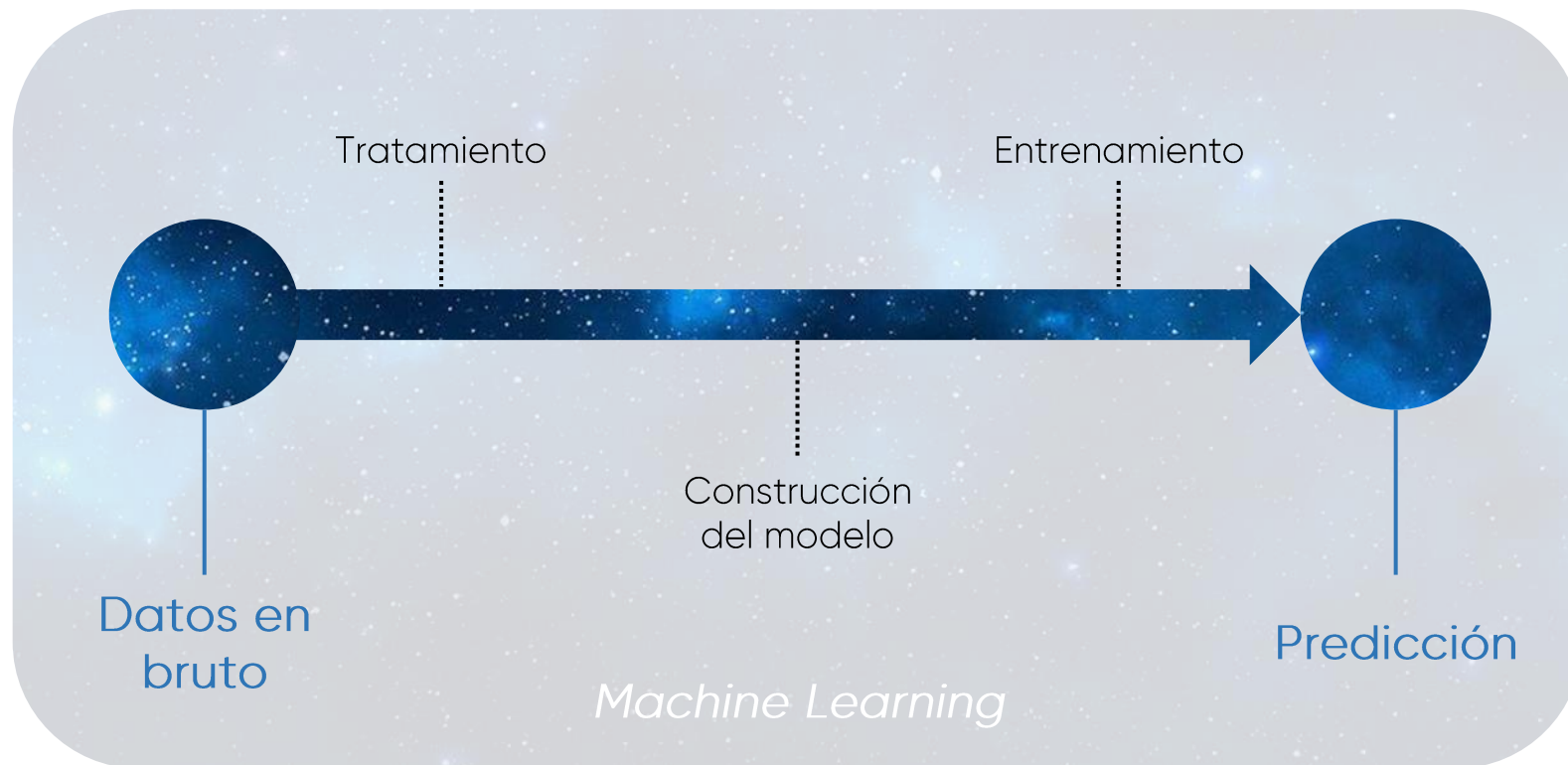
Tutor

Fernando Ortega Requena

Agenda

- Introducción
- Machine Learning
- Exoplanetas
- Datos
- Tratamiento
- Modelo
- Conclusiones

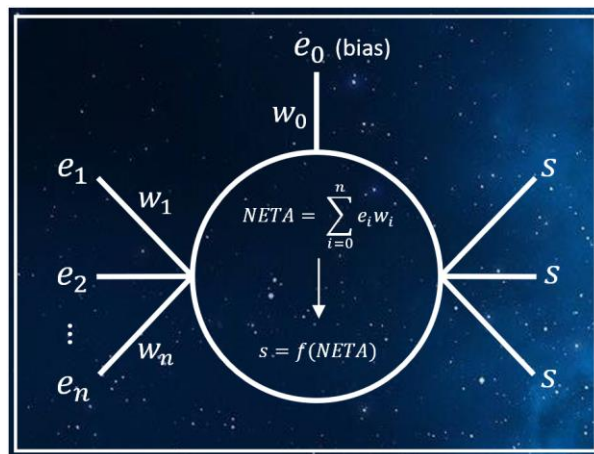
Introducción



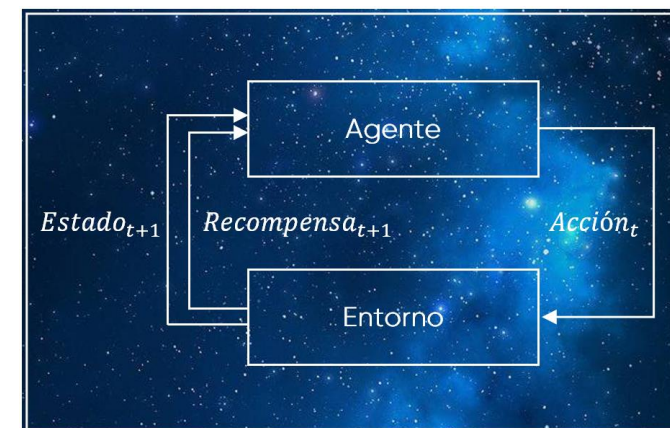
Machine Learning

Aprendizaje

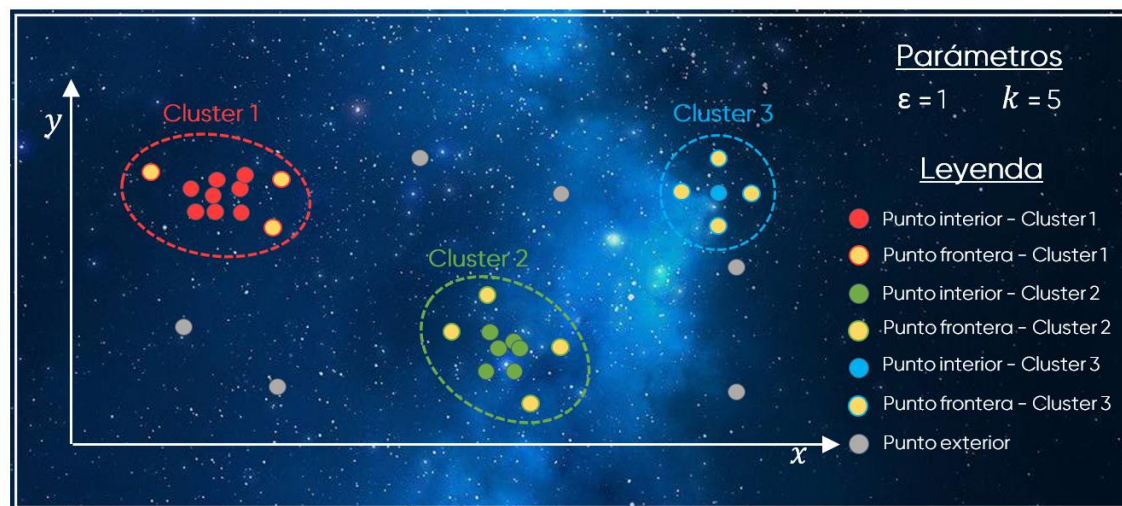
Supervisado



Refuerzo



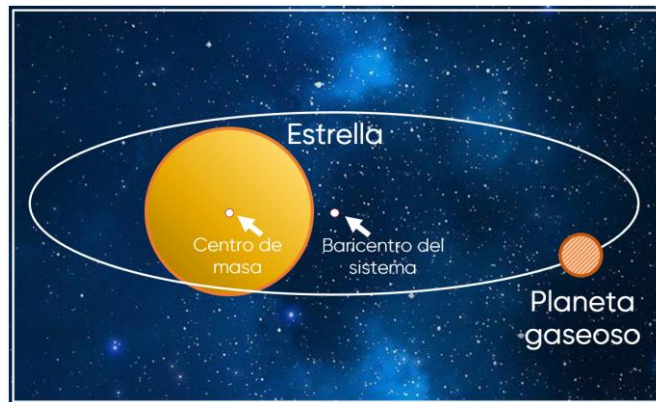
No supervisado



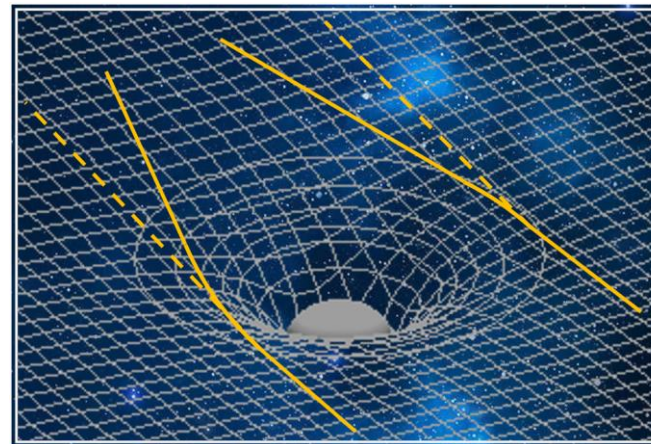
Exoplanetas

Métodos de detección

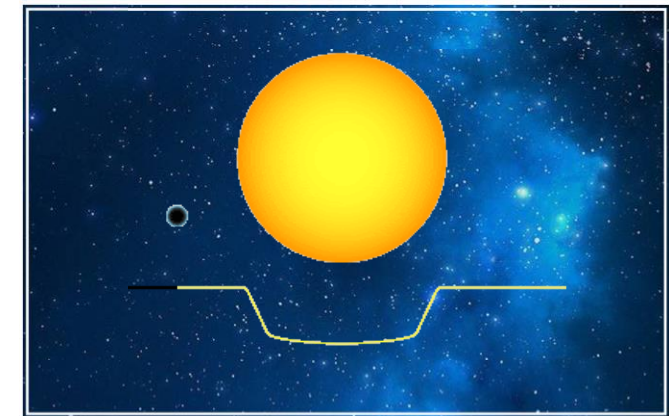
Velocidades
radiales



Micro lente



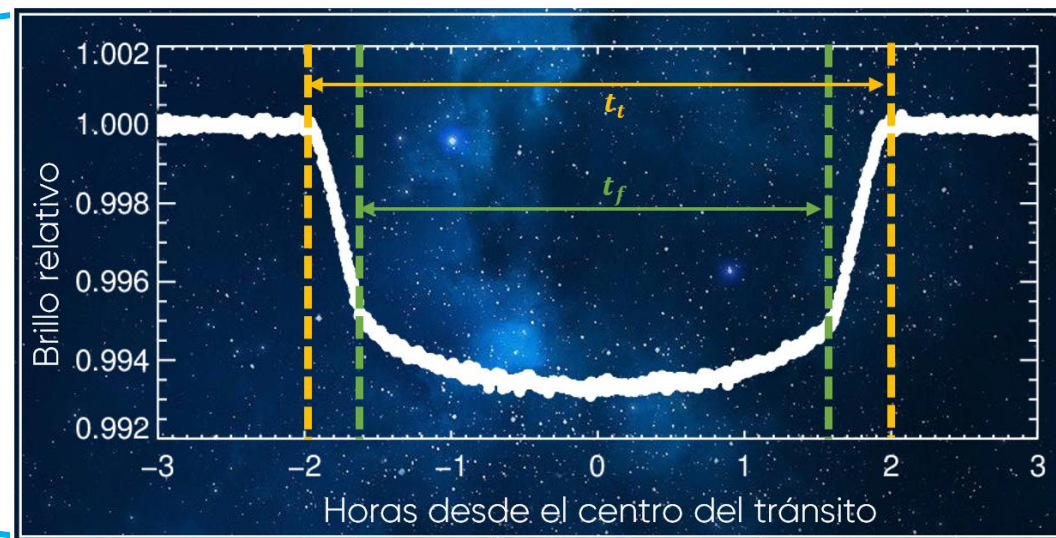
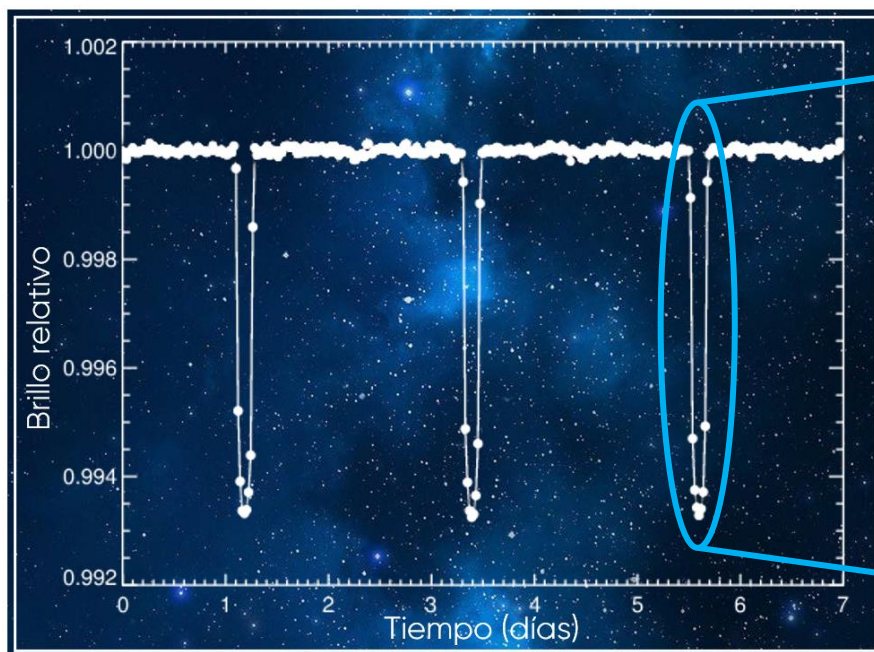
Tránsito



Exoplanetas

Método de tránsito

Curvas de luz



Datos

En bruto

Ejemplo

Entrenamiento

	LABEL	FLUX.1	FLUX.2	FLUX.3	...	FLUX.3196	FLUX.3197
0	2	93.85	83.81	20.10	...	5.08	-39.54
1	2	-38.88	-33.83	-58.54	...	16.00	19.93
2	2	532.64	535.92	513.73	...	-70.02	-96.67
...
5085	1	3.82	2.09	-3.29	...	-6.41	-2.55
5086	1	323.28	306.36	293.16	...	-14.09	27.82

5087 filas x 3198 columnas

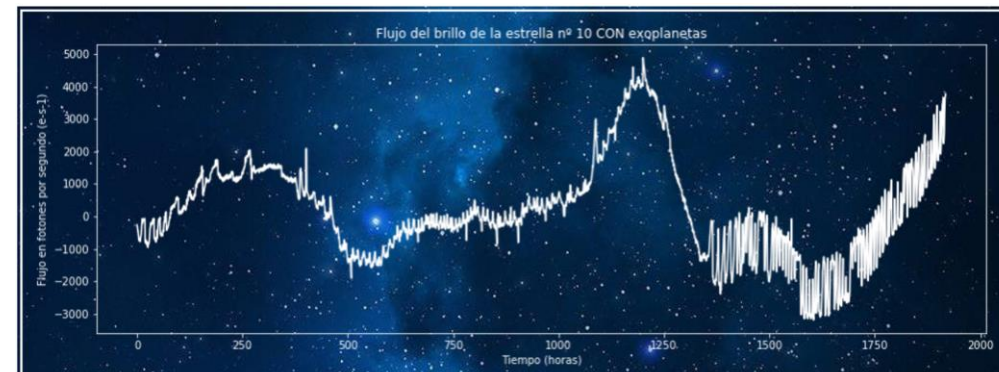
37 con exoplanetas / 5050 sin exoplanetas

Test

	LABEL	FLUX.1	FLUX.2	FLUX.3	...	FLUX.3196	FLUX.3197
0	2	119.88	100.21	86.46	...	269.43	57.72
1	2	5736.59	5699.98	5717.16	...	-2294.86	-2034.72
2	2	844.48	817.49	770.07	...	-36.79	30.63
...
568	1	91.36	85.60	48.81	...	-6.48	17.60
569	1	3071.19	2782.53	2325.47	...	-69.63	121.56

570 filas x 3198 columnas

5 con exoplanetas / 565 sin exoplanetas



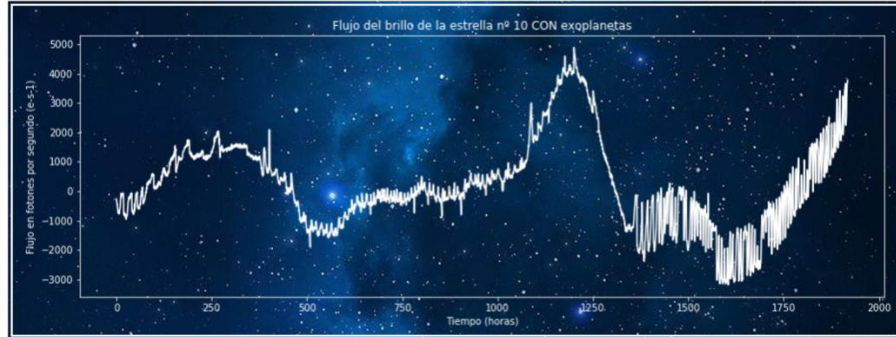
Problemas

- Datos no tratados
- Número de dimensiones
- Desbalanceamiento

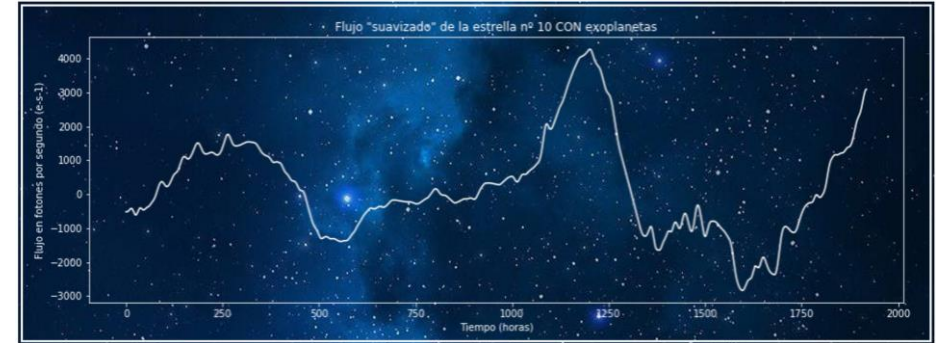
Tratamiento

Adecuación

Original

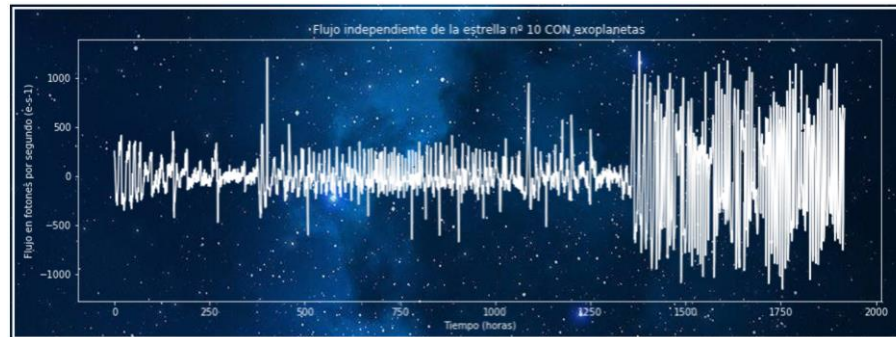


Flujo suavizado



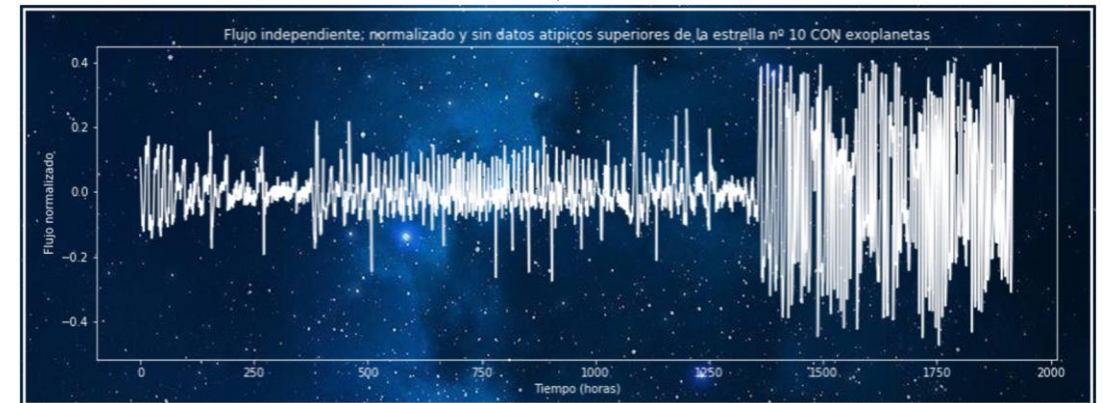
-

=



Eliminar datos
atípicos superiores

Normalización



Tratamiento

PCA

3197
dimensiones

0.9 varianza
→

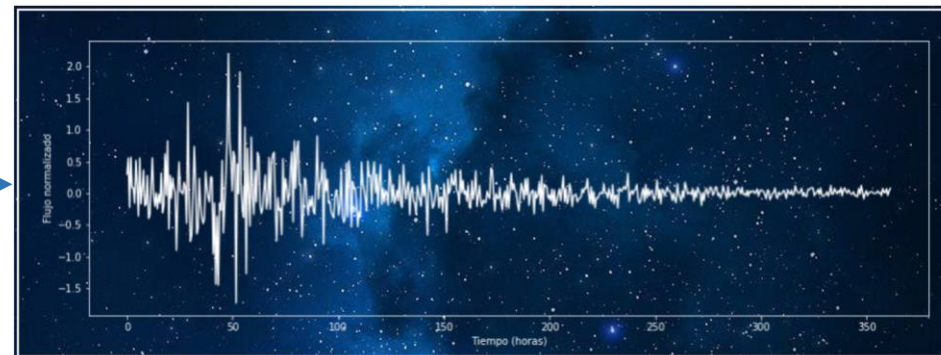
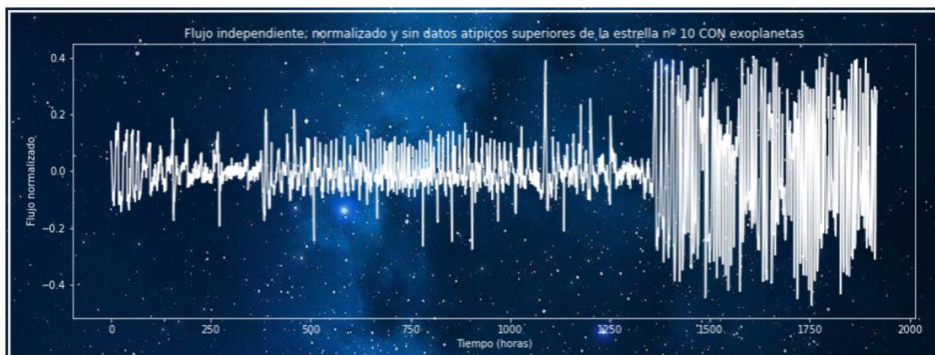
603
dimensiones

	LABEL	FLUX.1	FLUX.2	FLUX.3	...	FLUX.3196	FLUX.3197
0	2	0.1064	0.1	0.0581	...	0.0026	-0.027
1	2	0.0414	0.0507	0.0035	...	0.0398	0.0471
2	2	0.0363	0.0392	0.0188	...	-0.032	-0.057
...
5085	1	0.0072	0.004	-0.006	...	-0.002	0.005
5086	1	0.088	0.0759	0.0672	...	-0.014	0.0167

5087 filas x 3198 columnas

	LABEL	FLUX.1	FLUX.2	FLUX.3	...	FLUX.602	FLUX.603
0	2	0.0646	0.0967	-0.025	...	-0.007	0.0219
1	2	0.2809	0.0516	0.1324	...	-0.004	0.1323
2	2	-0.095	0.2125	-0.554	...	-0.021	-0.089
...
5085	1	-0.004	-0.161	0.0413	...	-0.005	-0.001
5086	1	0.3201	-0.382	-0.145	...	0.0029	0.0281

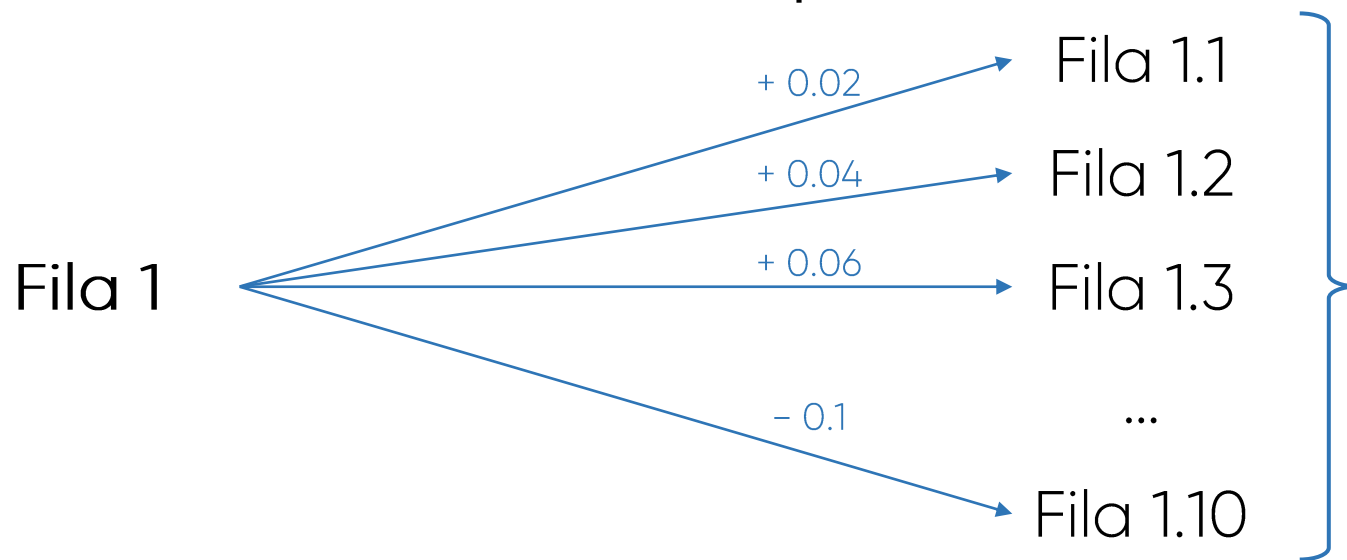
5087 filas x 604 columnas



Tratamiento

Data augmentation

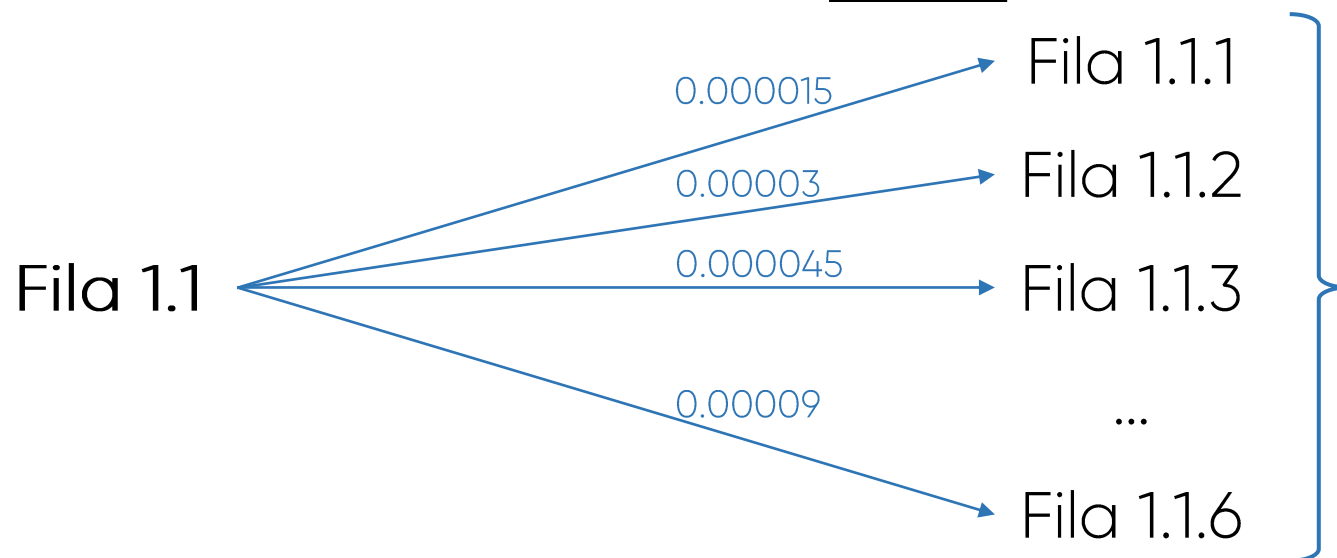
Desplazamiento



$$37 + 37 \times 10 =$$

407 filas

Ruido

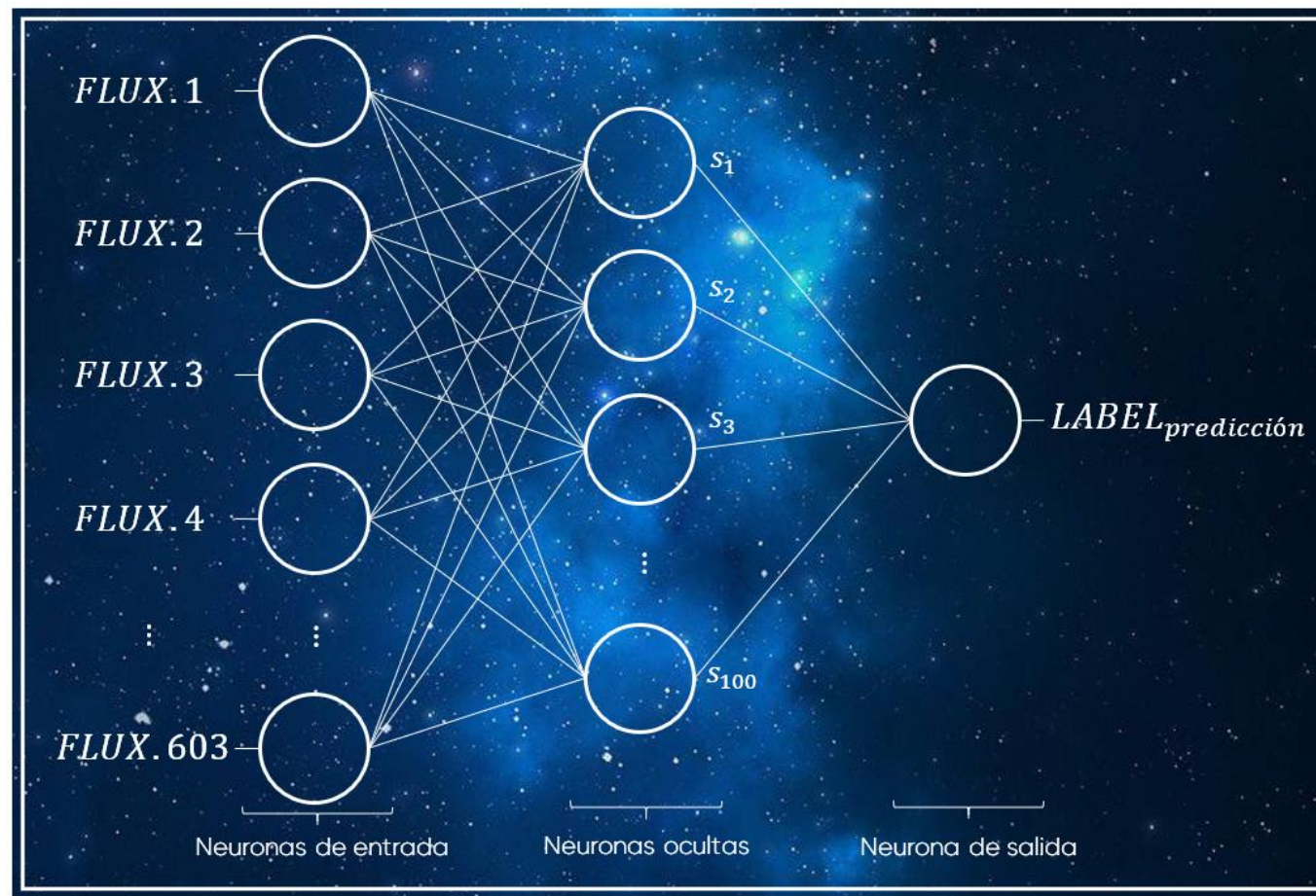


$$407 + 407 \times 6 =$$

2849 filas

Modelo

Red neuronal artificial



Resultados

Sin tratamiento

Predicción \ Real	SIN exoplanetas	CON exoplanetas
SIN exoplanetas	346	219
CON exoplanetas	2	3

1.35%

60%

2.64%

Precisión

Exhaustividad

F1

Con tratamiento

Predicción \ Real	SIN exoplanetas	CON exoplanetas
SIN exoplanetas	562	3
CON exoplanetas	3	2

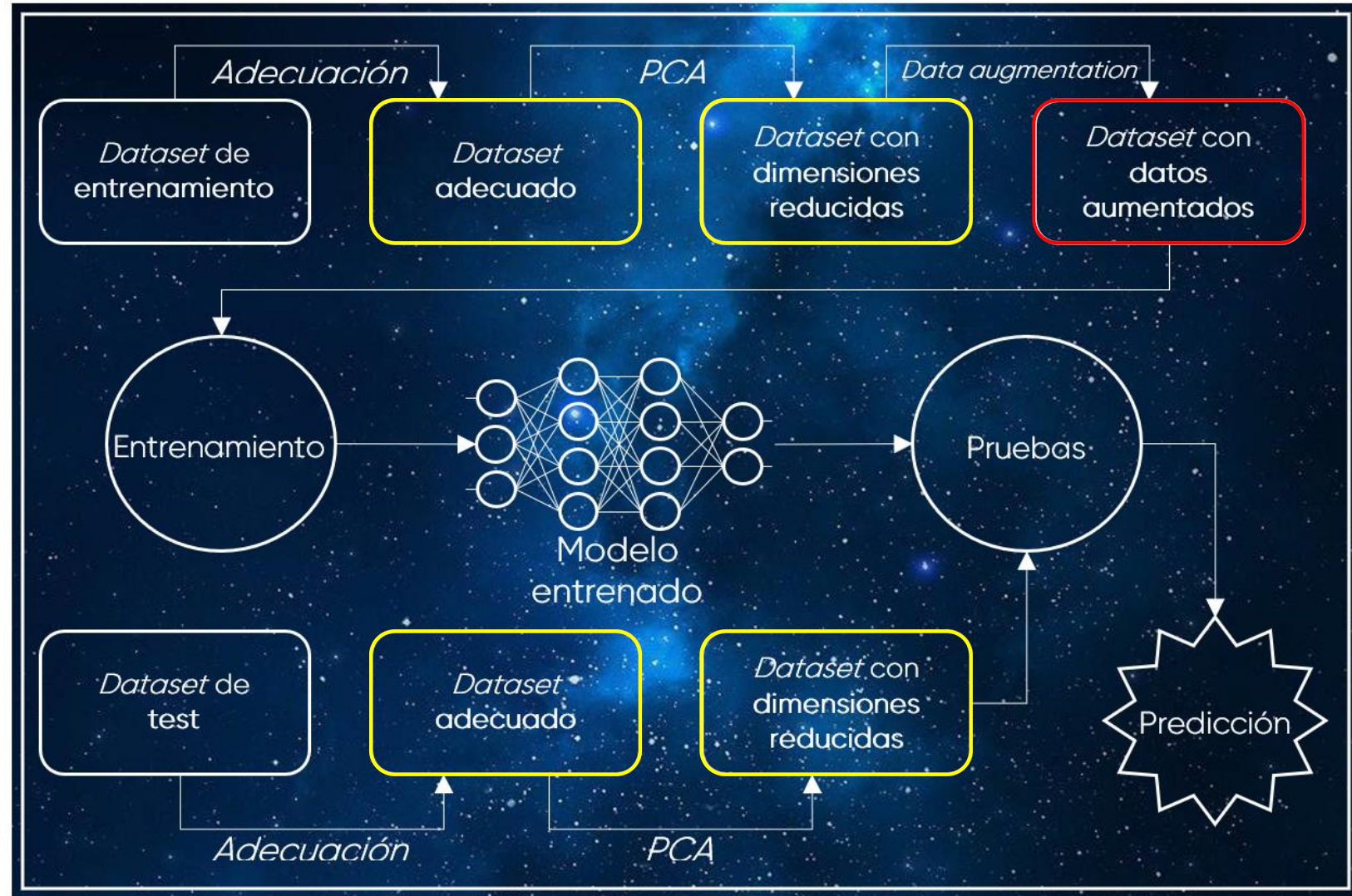
40%

40%

40%

Conclusiones

Tratamiento



Per aspera ad astra

Machine Learning para el tratamiento de datos y la detección de exoplanetas mediante el método de tránsito



Universidad
Politécnica
de Madrid

**ETSI SISTEMAS
INFORMÁTICOS**

Trabajo de Fin de Grado

Grado en Ingeniería de Sistemas de Información

Curso 2019/2020



POLITÉCNICA

Autor

José Javier Gómez de Diego

Tutor

Fernando Ortega Requena