

INTRODUCCIÓN

Nuestro proyecto utiliza los distintos métodos vistos en clase para analizar datos históricos e intentar predecir mediante clasificación la posición final del piloto.

OBJETIVOS

- 1. Aplicar los distintos métodos de clasificación vistos en clase.
- 2. Comparar los distintos métodos de clasificación para determinar el que mejor se ajuste a nuestro dataset.

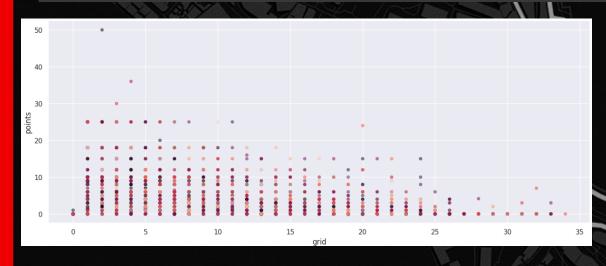


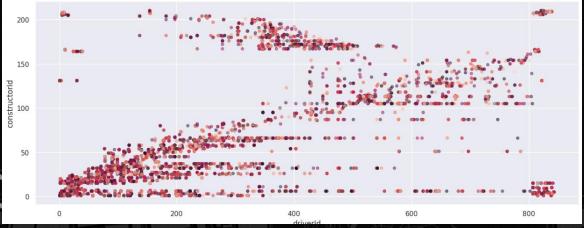
- 3. Obtener un "Accuracy" aceptable de almenos 50%
- 4. Observar el

 "accuracy" de un
 dataset destinado
 al "aprendizaje
 supervisado"
 sometiéndolo a
 métodos de
 "Aprendizaje no
 supervisado"

DATASET

	resultId	raceId	driverId	constructorId	grid	positionOrder	points	laps	milliseconds	fastestLap	fastestLapSpeed	statusId	lapPit
0	1	18	1	1	1	1	10.0	58	5690616.0	39.0	218.3	1	1.0
1	2	18	2	2	5	2	8.0	58	5696094.0	41.0	217.586	1	1.0
2	3	18	3	3	7	3	6.0	58	5698779.0	41.0	216.719	1	11.0
3	4	18	4	4	11	4	5.0	58	5707797.0	58.0	215.464	1	12.0
4	5	18	5	1	3	5	4.0	58	5708630.0	43.0	218.385	1	13.0





Data columns (total 13 columns): Column Non-Null Count Dtype resultId 23777 non-null int64 0 raceId 23777 non-null int64 driverId 23777 non-null int64 constructorId 23777 non-null int64 grid 23777 non-null int64 positionOrder 23777 non-null int64 points 23777 non-null float64 laps 23777 non-null int64 milliseconds float64 6003 non-null fastestLap 5383 non-null float64 fastestLapSpeed 5383 non-null object statusId 23777 non-null int64 lapPit 6251 non-null float64 dtypes: float64(4), int64(8), object(1) memory usage: 2.4+ MB

DESARROLLO

- 0.75

- 0.50

- 0.25

- 0.00

			1.6	96									
resultId	1.00	0.96	0.69	0.48	-0.00	-0.03	0.18	0.08	-0.00	0.25	-0.22	0.04	-0.19
raceld	0.96	1.00	0.72	0.47	-0.00	-0.02	0.15	0.07	0.04	0.16	-0.13	0.06	-0.18
driverId	0.69	0.72	1.00	0.38	0.09	0.04	-0.08	0.08	0.12	0.09	-0.09	0.07	-0.11
constructorId	0.48	0.47	0.38	1.00	0.12	0.10	-0.07	0.01	0.18	-0.03	0.02	0.13	-0.08
grid	-0.00	-0.00	0.09	0.12	1.00	0.13	-0.36	0.08	0.33	-0.00	-0.02	-0.14	0.00
positionOrder	-0.03	-0.02	0.04	0.10	0.13	1.00	-0.56	-0.66	0.61	-0.18	0.11	0.53	-0.04
points	0.18	0.15	-0.08	-0.07	-0.36	-0.56	1.00	0.26	-0.65	0.35	-0.27	-0.27	-0.01
laps	0.08	0.07	0.08	0.01	0.08	-0.66	0.26	1.00	-0.28	0.22	-0.15	-0.35	0.02
milliseconds	-0.00	0.04	0.12	0.18	0.33	0.61	-0.65	-0.28	1.00	-0.33	0.24	0.37	-0.04
fastestLap	0.25	0.16	0.09	-0.03	-0.00	-0.18	0.35	0.22	-0.33	1.00	-0.84	-0.21	0.03
fastestLapSpeed	-0.22	-0.13	-0.09	0.02	-0.02	0.11	-0.27	-0.15	0.24	-0.84	1.00	0.18	-0.05
statusId	0.04	0.06	0.07	0.13	-0.14	0.53	-0.27	-0.35	0.37	-0.21	0.18	1.00	-0.05
lapPit	-0.19	-0.18	-0.11	-0.08	0.00	-0.04	-0.01	0.02	-0.04	0.03	-0.05	-0.05	1.00
	resultid	raceld	driverId	constructorid	grid	positionOrder	points	laps	milliseconds	fastestLap	sstestLapSpeed	statusId	lapPit



- 8 Decision Tree
- ੳ Random Forest
- 8 SVC
- 9 DNN
- **8** Kmeans
- **9** DBSCAN



RESULTADOS

- 8 Se aplicaron con éxito los métodos de clasificación vistos en clase.
- θ Se obtuvo que el método de clasificación que mejor se ajusta a los datos es el DecisionTree.
- ઇ No se logro el "accuracy" esperado (50%), se obtuvo un 38.5% al usar DecisionTree.
- 8 El "accuracy" obtenido con métodos de aprendizaje no supervisado es muy inferior a los obtenidos con aprendizaje supervisado.

Ferrari