



MD4: Notebook Kaggle

F1 Race Predictor Models



Juan Carlos Soriano Valle
1493037



Introducció

La F1 és la categoria reina de l'automobilisme. La seva història es remunta a l'any 1950, quan es va organitzar la primera carrera oficial de F1 al món.

En la temporada 2020, 20 pilots diferents repartits en 10 escuderies intenten aconseguir el campionat de pilots i d'escuderies (SPOILER: Hamilton i Mercedes l'han guanyat).

L'objectiu d'aquest projecte és predir de la manera més precisa possible el guanyador de cada carrera d'una temporada en específic.





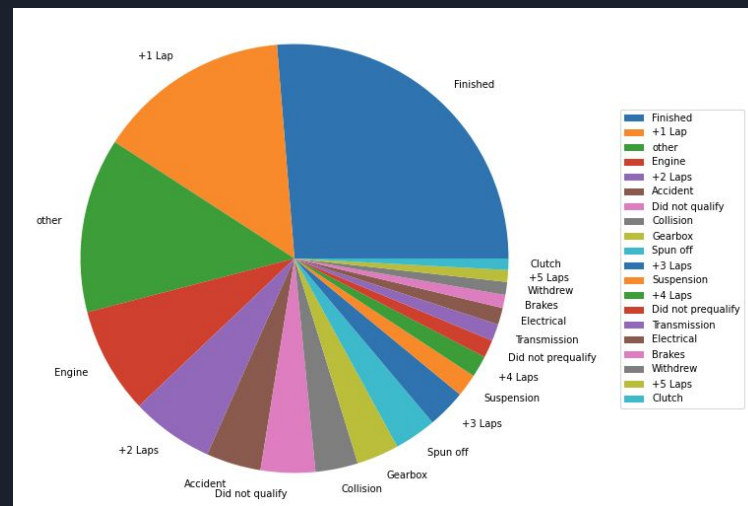
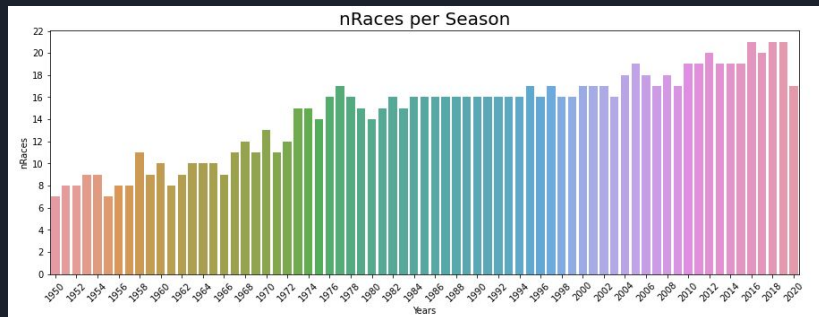
Data Exploration

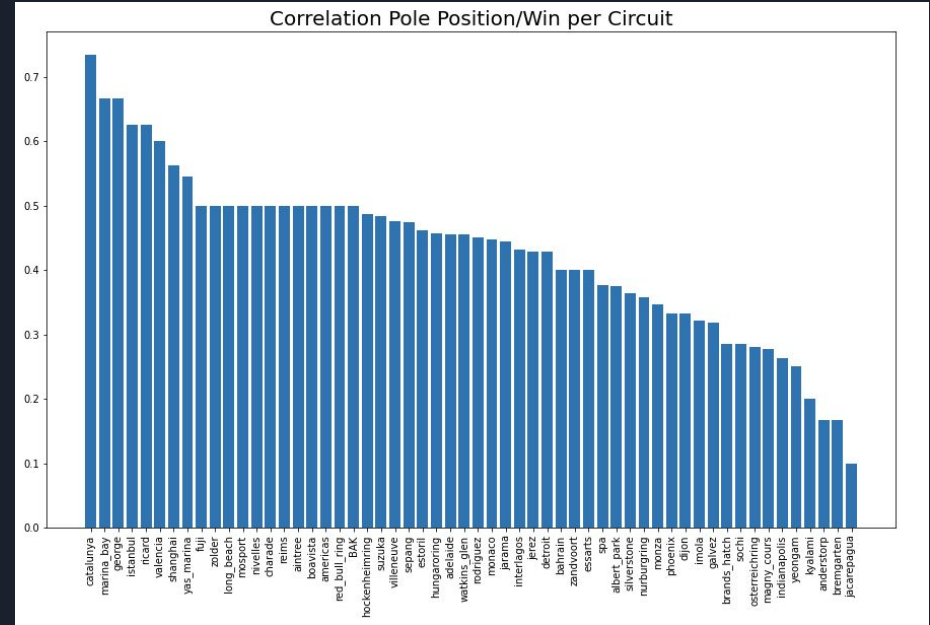
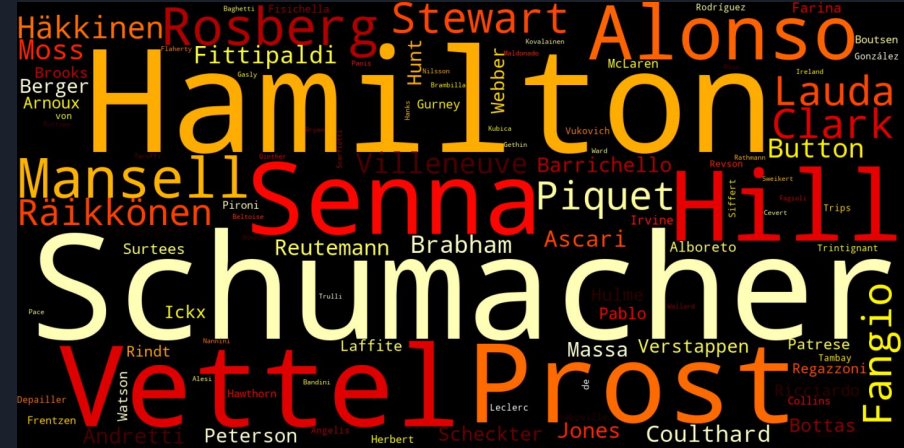
Dades de tota la història de la F1 (1950-2020)

13 arxius .csv:

- Circuits.csv (76,9)
- ConstructorResults (11700,5)
- ConstructorStandings (12456,7)
- Constructors (211,5)
- DriverStandings (32857,7)
- Drivers (848,9)
- LapTimes (487314,6)
- PitStops (7911,7)
- Qualifying (8634,9)
- Races (1035,8)
- Results (24900,18)
- Seasons (71,2)
- Status (137,2)

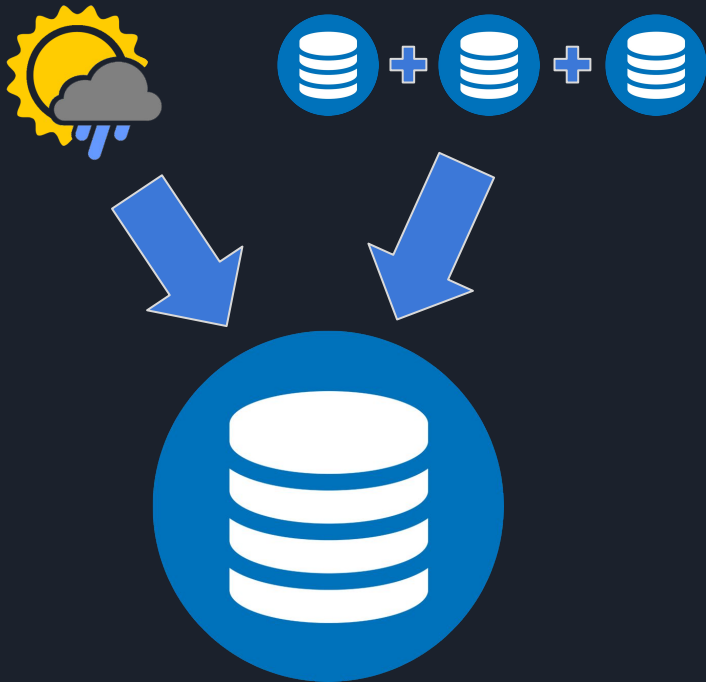
Data Analysis







Data Preparation



- + Edat del pilot
- + Temps Sessió de Classificació
- + Dummies de Constructors, Nacionalitat i Circuit
- Atributs ID's
- Temps total de Cursa



Model Creation

Regressor

- Puntuació obtinguda ascendentment ordenada
- De cada carrera, el que té menys puntuació és el guanyador predit
- Predicció vs Realitat

Classifier

- Probabilitat de ser 0 (no guanyador) o 1 (guanyador) ordenada descendentment en valor de la classe 1
- De cada carrera, el que té més probabilitat de ser de la classe 1 és el guanyador predit
- Predicció probabilitats vs Realitat

-
- Puntuació resultat de totes les carreres de la temporada
 - Resultats de la carrera mapejats en 1-0, on 1 és la classe guanyadora i 0 és la classe no guanyadora



Model Creation

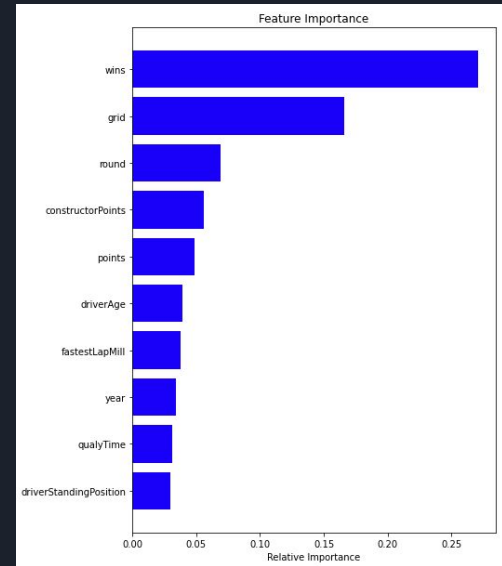
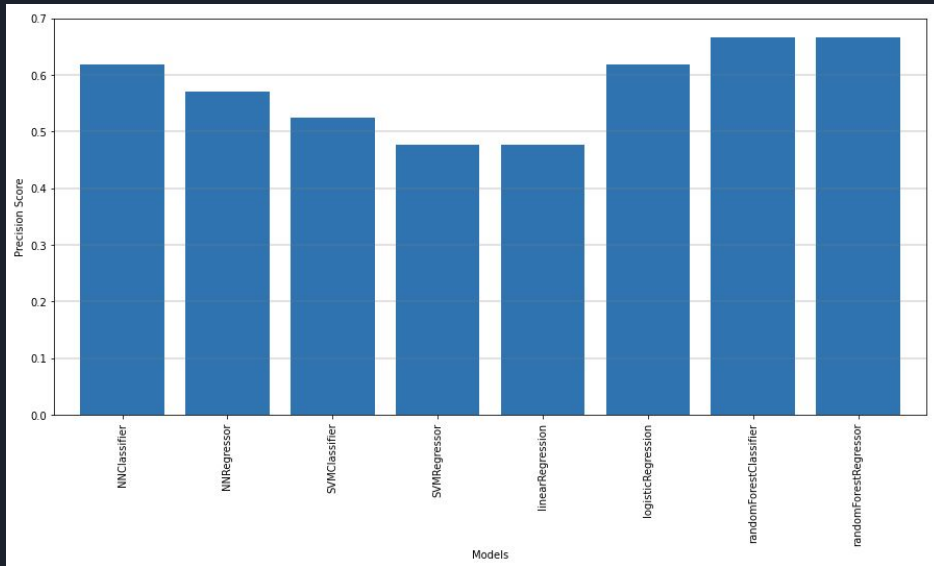
Regressor

	results	driver	positionOrder	real	predicted
0	6.279383	bottas	1	1	1
1	7.154947	hamilton	2	0	0
2	8.124644	vettel	4	0	0
3	8.218787	max_verstappen	3	0	0
4	8.373073	leclerc	5	0	0
5	10.238141	kevin_magnussen	6	0	0
6	11.177567	hulkenberg	7	0	0
7	11.649679	raikkonen	8	0	0
8	12.080526	gasly	11	0	0
9	12.103294	grosjean	18	0	0
10	12.194886	norris	12	0	0
11	13.053382	ricciardo	19	0	0
12	13.238850	perez	13	0	0
13	13.552351	stroll	9	0	0
14	13.742560	kvyat	10	0	0
15	14.239160	albon	14	0	0
16	14.475749	russell	16	0	0
17	14.595040	sainz	20	0	0
18	14.956287	giovinazzi	15	0	0
19	15.019627	kubica	17	0	0

Classifier

	prob0	prob1	driver	real	predicted
0	0.392033	0.607967	hamilton	1	1
1	0.735696	0.264304	bottas	0	0
2	0.881418	0.118582	max_verstappen	0	0
3	0.914869	0.085131	vettel	0	0
4	0.973692	0.026308	kevin_magnussen	0	0
5	0.977356	0.022644	gasly	0	0
6	0.985076	0.014924	sainz	0	0
7	0.991676	0.008324	ricciardo	0	0
8	0.991984	0.008016	kvyat	0	0
9	0.995908	0.004092	norris	0	0
10	0.996585	0.003415	leclerc	0	0
11	0.997357	0.002643	albon	0	0
12	0.998299	0.001701	hulkenberg	0	0
13	0.998407	0.001593	raikkonen	0	0
14	0.998718	0.001282	perez	0	0
15	0.999272	0.000728	grosjean	0	0
16	0.999676	0.000324	stroll	0	0
17	0.999845	0.000155	giovinazzi	0	0
18	0.999874	0.000126	russell	0	0
19	0.999935	0.000065	kubica	0	0

Resultats





Conclusions

- Bons resultats dels models.
- Features més importants: Victòries del pilot, posició a la graella de sortida i el constructor.
- F1 és molt imprevisible.
- Amb més dades específiques (sessions d'entrenament i dades d'aerodinàmica) es podria millorar el rendiment del predictor.

