Predicting Formula 1 Race Performance from Free Practice Data Using Machine Learning

Simen S. Gåsland

University of Stavanger

17. November 2025

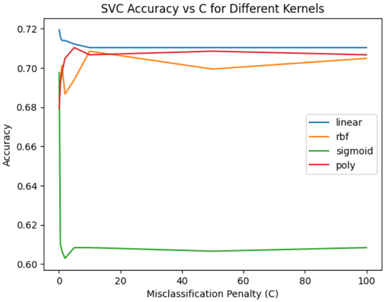
Data Collection, Cleaning and Feature Engineering

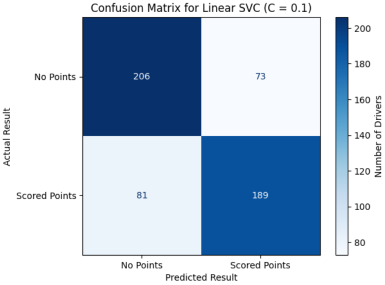
- FastF1 Python package via the **jolpica-f1** API
- A custom **DataAquisition** class

Key features:

- **Driver performance:**
  FastestFPLap, MeanFPLaps, StdFPLaps, DeltaBestFPLap
- **Weather:**
  TrackTempAvgFP, AirTempAvgFP, RainAvgFP
- **Race outcomes:**
  FastestLapRace, FasterThanTeammateRace, PointFinishRace

Predicting Point-Scoring Drivers Using an SVC

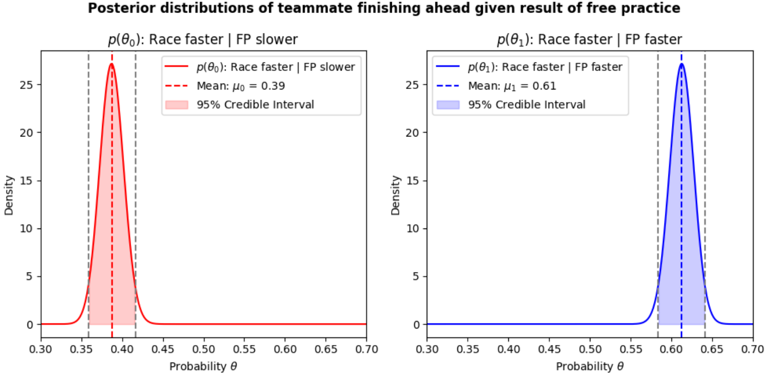




Class	Precision	Recall	F1-score
No Points	0.72	0.74	0.73
Scored Points	0.72	0.70	0.71

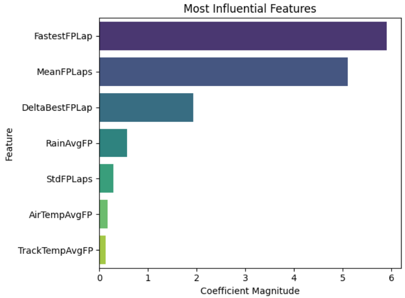
Modeling Teammate Performance Using a Beta-Binomial Model

Prior: $\theta \sim \mathrm{Beta}(1,1)$ $\Rightarrow$ **Posterior:** $\theta \mid$ Slower in FP $\sim \mathrm{Beta}(426, 674)$
Posterior: $\theta \mid$ Faster in FP $\sim \mathrm{Beta}(674, 426)$



Predicting the Fastest Race Lap Using Linear Regression

Deg	Model	α	MAE	R^2
1	LinearReg.	–	1.435±0.292	0.882±0.065
	Lasso	0.027	1.431±0.278	0.885±0.063
	Ridge	7.055	1.428±0.290	0.883±0.064
	ElasticNet	0.005	1.429±0.291	0.883±0.064
	EN (l_1=0.7)	0.007	1.429±0.290	0.884±0.064
2	LinearReg.	–	2.094±0.939	0.686±0.296
	Lasso	0.248	1.734±0.669	0.734±0.227
	Ridge	95.455	1.874±0.916	0.744±0.285
	ElasticNet	0.067	1.839±0.887	0.744±0.286
	EN (l_1=0.7)	0.142	1.814±0.739	0.738±0.238
3	LinearReg.	–	3.339±2.226	-0.621±2.117
	Lasso	0.171	2.827±2.307	-2.005±3.617
	Ridge	12.328	3.282±2.828	-1.113±3.632
	ElasticNet	0.007	3.170±2.860	-1.056±3.664
	EN (l_1=0.7)	0.007	3.161±2.896	-1.065±3.705



Final choice: ElasticNet model with an $L1$-ratio of 0.7

Testing on new unseen 2025 data