

# Genpact Formula E Racing Lap Prediction

## Executive Summary

-Team SmithRacers101

Formula E Racing is the world's first full electric, international one-seater, street racing championship. Apart from bringing awareness to fighting climate change, it is also creating a technological and sustainable development test bed that helps address mobility and other environmental issues. Furthermore, it is the first global sport with net zero carbon footprint. The prediction question that we have tried to answer is if, using historical data, we can predict the number of laps the driver will need to complete the race. The reason why it's important to solve this business case is because once a driver would know how many laps are left in a race, then that would help him better frame his energy optimization strategy. For example, if our model predicts that the number of laps remaining are 30 with low battery, then the racer could choose to drive conservatively and save energy. On the other hand, if our model would predict the number of remaining laps to be 30 with high battery power, then the racer could choose to drive aggressively and expend energy. Either way, if the racer can better manage their battery expenditure, then that would lead to a better strategy at winning the race.

Within this prediction, we came across a few key insights related to more optimized racing which are as follows:

- Second day of double header races seem to have faster average speeds
- We observe a slight increase in mean kph as wind speed/track temp increases
- We observe a slight reduction in median speed (KPH) before the drivers speed up again in the final lap

We built a total of four prediction models:

- K-Nearest Neighbours (Baseline Model)
- Random Forest (Baseline Model)
- ElasticNet (Baseline Model)
- Weighted Average of all three baseline models (Ensemble Model)

From our results, we observe that the weighted average ensemble model performs better than the rest of the models while predicting the total possible laps for a particular match. Essentially, for the weighted model, we assigned weights to each of the baseline model based on a linear regression equation determining the significance of each baseline model on the predictor variable, which is total number of laps predicted. Thus, since our ensemble model predicts total number of laps, our real-time feedback to the racer would be the total remaining laps which we would get from subtracting the laps already completed from the total laps predicted by our model. In this way, at each point of time during the race, the racer would know an estimate of how many laps are remaining in the race and would be able to come up with a more effective battery expenditure strategy for winning the race.

To conclude, we would like to:

- Further improve on our ensemble model by including weather features
- Clearly, weather has an impact on the average speed of the racer and would be an important factor to consider for prediction of total laps