ALL FEATURES

1. Read the literature, consulted with people on LinkedIn.
2. Step by step, created the dataset—from testing the API on one session to building the whole dataset for 2022–2025.
3. Trained the models and found that the base random forest is the best one.
4. Tried feature engineering: instead of one-hot encoding compounds, set them as numbers and removed features(not C\_1 == 0 || C\_1 == 1, but C == 1 || C == 2).
5. Normalized the data and removed outliers.
6. Removed some features (year, stint\_start)

Зображення, що містить текст, знімок екрана, Графік, ряд

Вміст на основі ШІ може бути неправильним.

1. Feature importances showed that the model focuses on the wrong features.

Зображення, що містить текст, знімок екрана, Шрифт, число

Вміст на основі ШІ може бути неправильним.

3 FEATURES

1. Decided to make a model with three features: circuit length, compound name(SOFT), and tyre life as the label. In the process, I understood how wrong normalization can be—I can’t just normalize circles between all tyres; for one track, it is soft, for another, it is hard.
2. Зображення, що містить текст, знімок екрана, число, ряд

   Вміст на основі ШІ може бути неправильним.New normalization – visualized data, removed ouliersЗображення, що містить текст, знімок екрана, ряд, число

   Вміст на основі ШІ може бути неправильним..

Too mutch data collision points