F1 Safety car predictor

# 1. Abstract

This is an introductory project that features Formula 1 and Machine Learning. Formula 1 represents the pinnacle of auto-moto sport and as today it is the most popular racing championship in the world.

Goal of this project is to use machine learning algorithm and a given set of data in order to predict whether a safety car will be deployed during a specific race in calendar. Safety car occurs in various situations in Formula 1, and it means that one car (*not on the grid*) will go in front of the first car in the race and slow down the entire column, until the reason for deploying it is resolved. During the safety car deployment, overtaking is not allowed, but pit stops are more profitable as everyone is slower, so this can play a crucial role in fighting for valuable points.

The most usual reason for a safety car is an incident, whether it is one car crashing in the wall or multiple cars colliding together. Rain can also bring it out for some time, however, it the rain is too heavy, the race will be stopped.

In order for this project to make more sense, it is decided that we will not be satisfied with the simple Boolean decision like, **yes**, there will be safety care, or **no** it will not, but rather, we want a percentage of that happening. So for example, in a circuit like Monaco, which is a street race, chance for a safety car is higher, like 70%, while on a traditional track like Austria, that percentage can go down to 35%. Of course, these numbers are just an assumptions at this point.

# 2. What will you need?

Like mentioned, there are various parameters that can increase the chance of a safety car deployment, and most of them are *static* and interchangeable over the years, meaning that there are very little dependencies between these variables.

For this prediction we will use following information:

* **Circuit name** – We have to know which track the race is held on due to various track parameters.
* For starters we will use a **year gap** from 2020 to 2024 making this a 5 year dataset. If it all works out, we will expand the gap, making a model more accurate and precise.
* **Weather information** – Weather is fairly important for this project since rain is the one of the biggest reasons for incidents and most likely, safety car deployment.
* **Number of DNFs** – DNF stands for “*Did not finish”* and simply means that the car did not make it to the end of the race. Reason could be an incident, but also a mechanical issue. We will use this information as well, since a lot of mechanical issues are connected to a geographical location of a track.
* **Circuit type** – Track configuration also plays a role. Street circuits are more narrow and difficult to drive, creating a more risky situation to overtake. Traditional circuits have more space and therefore are easier for less experienced players not to crash their cars.
* **Physical difficulty of a circuit, for both driver and car** – This is an experimental parameter, and it indicates whether a circuit is particularly difficult for drivers due to extreme weather conditions like high temperature and humidity. Also, cars don’t behave the same way on every track due to same parameters mentioned previously.

For this project, we will use Python and Jupyter Notebook to make it more readable. We will discuss the choice of a machine learning algorithm in the next chapter.

# 3. Choosing a ML algorithm

* Random Forest
* Neural network
* Logistic regression

… waiting for a complete research