

# **Formula 1: Drivers and Constructors Championship Predictor**

Contributors:

Lorenzo Resurreccion

Ashish Priyadarshi

## **Overview**

This project involves using a known set of data and results to create a prediction model that will predict the outcome of each race. We then repeat for various races and compile this information to predict the results of the Driver and Constructors championships for an unknown set of data (current year's races).

## **Background**

There exist a few projects that take in race data, qualifying data, and various other factors such as weather, track, tires used, and driver data that have been used to create a prediction model for the F1 championship. Some projects are on github with relevant data they used as well as their methodology (<https://github.com/Sloopy3333/Formula-1-Prediction>). Other examples include articles explaining the results they found after training and using their model to predict information. ([link](#))

Looking more in depth at some existing projects, many vary in their goals and methodology. While some use individual lap data to predict the standings of each driver after multiple laps as well as the eventual winners/standings of each race, others focus more on predicting the overall standings by looking at more general race data and use it to predict the champion of the whole series.

## **Statement of Work**

### **Datasets:**

Multiple data sets exist such as compiled and preprocessed data that have already been used in github (<https://github.com/JaideepGuntupalli/f1-predictor>) and various online data such as [F1's official data](#) and [Ergast API data](#). These datasets will contain information about every race, drivers, their corresponding team, and their finishing result. We will need to preprocess the data to only include drivers who are currently racing, filtering out the retired drivers and teams. We also may need to determine if some columns need to be edited, for example, if a driver doesn't finish then they are classified as 'DNF' which we will change to position 21 (assuming 20 drivers in the grid).

## Method:

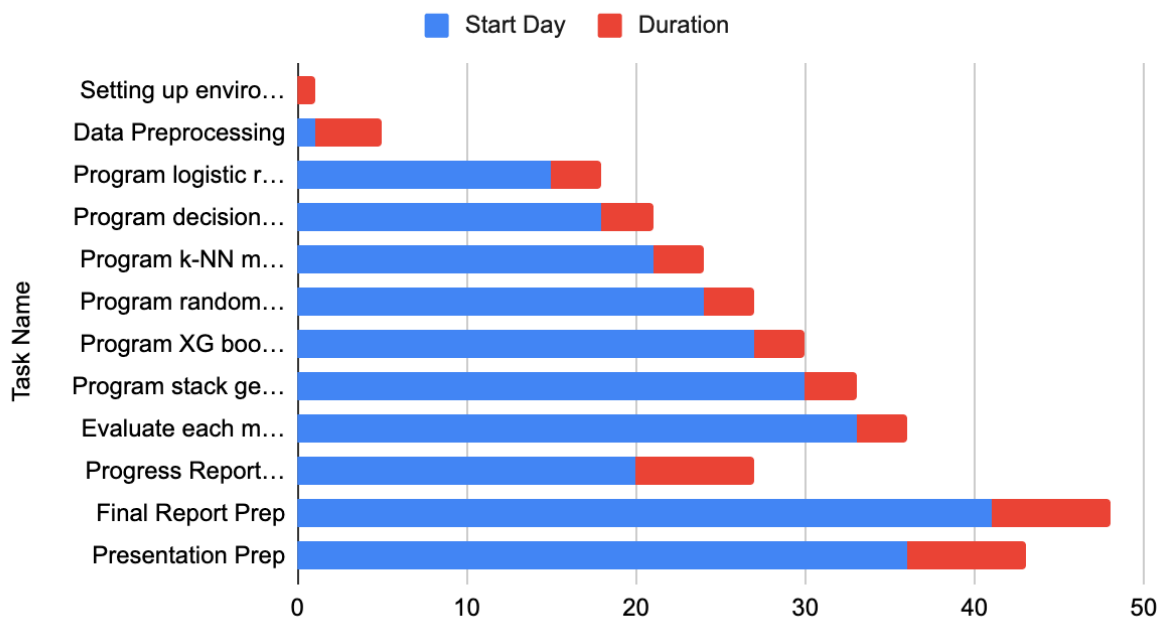
Prior projects have attempted predicting the finishing position of a driver given their qualifying position, year, track, and driver history. We will be attempting to build on this this by testing different models. We will first collect data from the sources listed above. After that, we will preprocess the data, and then start building our models. For this project we will use a logistic regression model, a decision tree, k-NN, random forest, XG Boosting, and also a stack with all of the previous models. Our plan is then to collect metrics for each model, and select the best performing one. We are passionate about Formula 1, and with so much data available we wish to apply our passion for the sport into a model that can predict the results for every year and every race and see how well it matches with the actual results.

## Outcome and Performance:

We will attempt to train our model on more recent years' data sets and test it on known results from those years using cross validation. We then plan to apply our trained model to this year's current races as our test set to monitor how well our model does and verify it doesn't overfit the training data.

## Project Plan

### Start Day and Duration



Link to Github repository being used for this project: [link](#)

## **References**

1. <https://github.com/Sloopy3333/Formula-1-Prediction>
2. [link](#)
3. <https://github.com/JaideepGuntupalli/f1-predictor>
4. [F1's official data](#)
5. [Ergast API data](#)

## **Contributions**

### **Lorenzo:**

I Created the Cover page, and contributed to Overview and Background Sections. Helped make decisions on schedule and method sections.

### **Ashish:**

I contributed to the links and statements of work sections. I also created the Gantt chart for planning our project while helping to make decisions about methodology and scheduling.