

COMP257 Project proposal for **group l**

Minhyung Kim

Kuk Kim, 40410633

Formula One Stats and searching for the Winning Formula

7/10/2018

## PROJECT TITLE

Formula One Stats and searching for the Winning Formula

## DESCRIPTION AND GOALS OF PROJECT

Formula One World Championship is the highest level and premier forms of auto car racing in the world.

The *formula* in this context refers to *the set of rules* that all participants’ racing cars must conform to. F1 is known to be all about speed and efficiency.

With the growing capabilities of the big data, data driven collaboration and the real time data feeds, the *real* race is happening behind the scenes.

F1 racing team uses data gathering techniques to design and build machine with optimum performance based on the collected data. With the combination of other racing factors such as weather, driver, and specific racing track details, the team pre-empt the race winning formula.

Our goal in this project is from the perspective of racing fans, we will show which racing factors contribute to race winning results by analysing the data.

### SUMMARY OF DATASETS

We will be using the datasets from various online sources stated below.

*http://www.statsf1.com/en/statistiques/pilote.aspx*

*https://www.racefans.net/2018-f1-season/2018-f1-statistics/*

We will be obtaining Grand Prix race and drivers related data such as fastest laps, pole positions etc. These data is in the form of texts and tables in html which we will be converting them to csv format for the analysis.

### RESEARCH TECHNIQUES

We expect to use data clustering to look for the common group of properties within dataset.

We will be using linear regression on the variables within the dataset to develop the predictive model.

We will then be using logistic regression to establish a baseline classification performance.

Further to that, we will be applying neural network to see if the performance can be improved. The techniques we will be using is K-nearest algorithm and cross-validation to conduct the parameter tuning.

### MILESTONES AND SUCCESS CRITERIA

Our first milestone is to do data gathering and applying clustering algorithms to find the structures within the dataset in preparation for data analysis.

The success criteria for this milestone is to obtain a clean dataset for the analysis. This can be evaluated by observing the descriptive statistics on the dataset and look for any abnormalities.

Our next milestone is successfully develop a simple working model to define a baseline performance.

The success criteria is to obtain a working model that produces baseline results. This can be evaluated by applying multiple test datasets and achieving the consistent result.

The last milestone is to apply a neural network for possible performance enhancements.

The success criteria is to get a better working model that produces results with enhanced performance. Evaluation process can be achieved by examining the outcome results against the baseline performance.