PREDICTIVE MODEL FOR FORMULA 1 USING DATA ANALYSIS

Parthiv Yagnesh Natrajan, Pravalika Arunkumar, Rajsekar Balaji, Rohan Suresh
21011101085, 21011101089, 21011101095, 21011101101
Dr. Sundharakumar K B

SHIV NADAR UNIVERSITY CHENNAI

OVERVIEW

A ***Formula 1*** season generally consists of 10 constructors, each comprising 2 drivers who race for them. A race weekend consists of 3 practice sessions, followed by qualifying, which decides the starting position of the drivers for the race.

Our project makes use of a dataset consisting of qualifying results and race results from the year 2010 to 2020 to predict the winner (constructor) of the 17 races in 2020. Logistic Regression and Random Forest Classifier algorithms were used, with RFC giving us the better accuracy of 82%, compared to 76% of LR.

OBJECTIVE & PROJECT STAKEHOLDERS

- F1 Analysts
- Online-betting Industry

The objective is to predict the winners of Formula 1 Grand Prix (plural) by building a model using Data Analysis and Machine Learning algorithms.

STEPS INVOLED

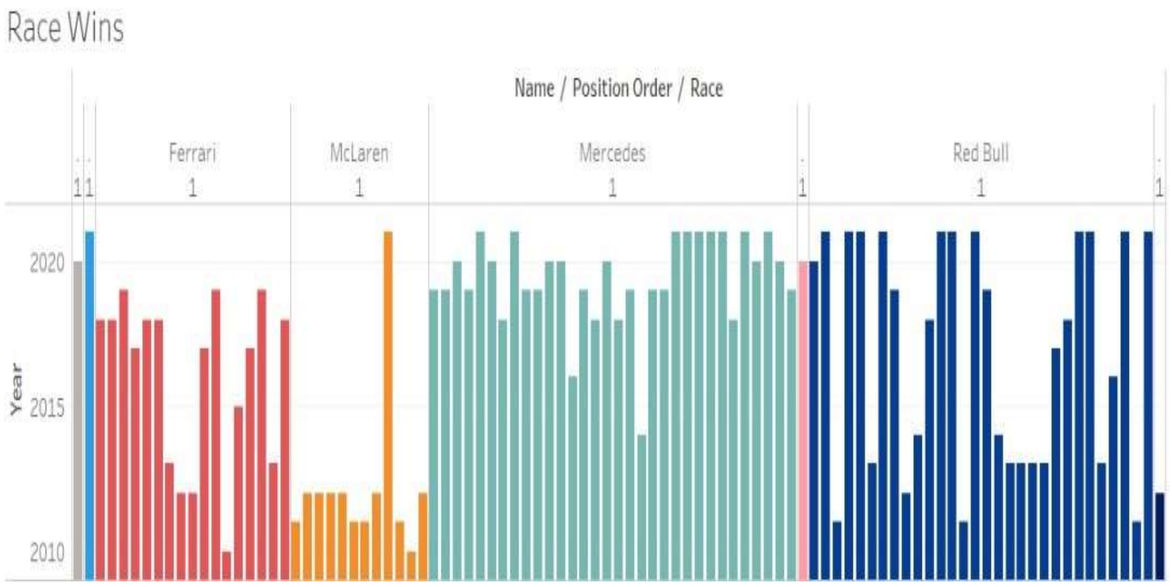
- Collection of data
- Cleansing of data
- Analysing the data
- Applying ML algorithms and testing their accuracy
- Predicting the results

SOFTWARE REQUIREMENT

- Python
- Pandas
- Scikit Learn
- Tableau
- MatPlotLib
- NumPy

DATA COLLECTION & ANALYSIS

The data was collected from Ergast F1 and analysed and visualised from Tableau.



Sheet1 Constructor Id	Sheet1 Name	Sheet1 Race Id	Sheet1 Result Id	Sheet1 Grid	Sheet1 Position Order	Sheet1 Race	Sheet1 Year
6	Ferrari	337	20323	3	1	Bahrain Grand Prix	2010
6	Ferrari	337	20324	2	2	Bahrain Grand Prix	2010
6	Ferrari	341	20420	4	2	Spanish Grand P...	2010
6	Ferrari	341	20424	9	6	Spanish Grand P...	2010
6	Ferrari	343	20473	8	7	Turkish Grand Prix	2010
6	Ferrari	343	20474	12	8	Turkish Grand Prix	2010
6	Ferrari	346	20552	3	14	British Grand Prix	2010
6	Ferrari	346	20553	7	15	British Grand Prix	2010
6	Ferrari	348	20588	3	2	Hungarian Grand...	2010
6	Ferrari	348	20590	4	4	Hungarian Grand...	2010
10	Force India	337	20331	12	9	Bahrain Grand Prix	2010
10	Force India	337	20334	10	12	Bahrain Grand Prix	2010
10	Force India	341	20425	11	7	Spanish Grand P...	2010
10	Force India	341	20433	16	15	Spanish Grand P...	2010
10	Force India	343	20475	11	9	Turkish Grand Prix	2010
10	Force India	343	20479	18	13	Turkish Grand Prix	2010
10	Force India	346	20546	11	8	British Grand Prix	2010
10	Force India	346	20549	20	11	British Grand Prix	2010
164	HRT	337	20341	23	19	Bahrain Grand Prix	2010

PREDICTIVE MODEL FOR FORMULA 1 USING DATA ANALYSIS

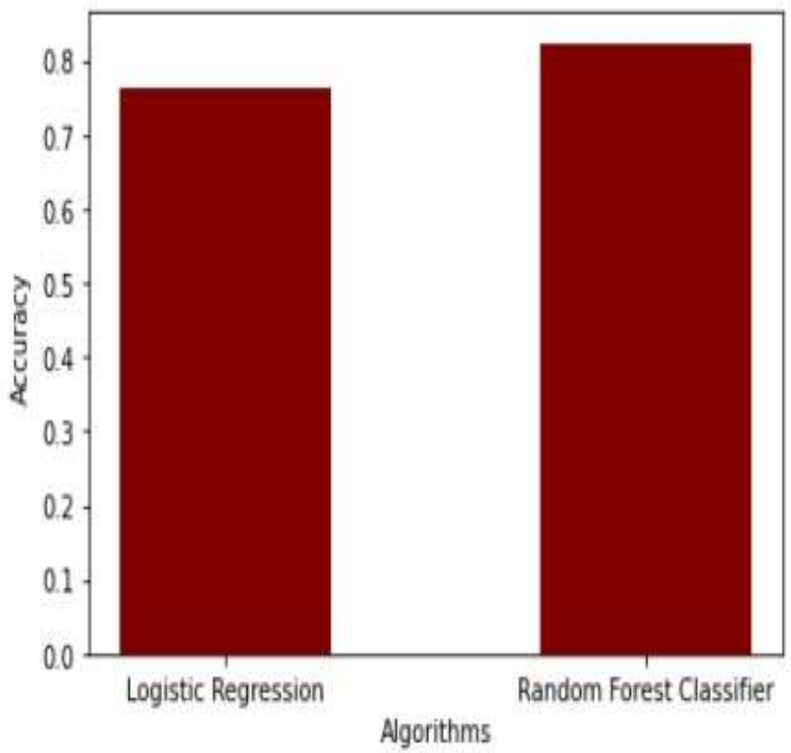
Parthiv Yagnesh Natrajan, Pravalika Arunkumar, Rajsekar Balaji, Rohan Suresh
21011101085, 21011101089, 21011101095, 21011101101
Dr. Sundharakumar K B

RESULT

Bar graph showing the comparison of the accuracy of both the algorithms, indicates that RFC performs with better accuracy than Logistic Regression.

Bar graph of Accuracy

```
labels=['Logistic Regression','Random Forest Classifier']
plt.bar(x=labels,height=y_axis,width=0.5,label=labels,color='maroon')
plt.xlabel('Algorithms')
plt.ylabel('Accuracy')
plt.show()
```



Using Random Forest Classifier, winners of 14 of the 17 races were predicted correctly by our model.

Race	Year	Grid	Position Order	Actual Winner	Predicted Winner
Austrian Grand Prix	2020	1	1	Mercedes	Mercedes
Styrian Grand Prix	2020	1	1	Mercedes	Mercedes
Hungarian Grand Prix	2020	1	1	Mercedes	Mercedes
British Grand Prix	2020	1	1	Mercedes	Mercedes
70th Anniversary Grand Prix	2020	4	1	Red Bull	Red Bull
Spanish Grand Prix	2020	1	1	Mercedes	Mercedes
Belgian Grand Prix	2020	1	1	Mercedes	Mercedes
Italian Grand Prix	2020	10	1	AlphaTauri	Mercedes
Tuscan Grand Prix	2020	1	1	Mercedes	Mercedes
Russian Grand Prix	2020	3	1	Mercedes	Mercedes
Eifel Grand Prix	2020	2	1	Mercedes	Mercedes
Portuguese Grand Prix	2020	1	1	Mercedes	Mercedes
Emilia Romagna Grand Prix	2020	2	1	Mercedes	Mercedes
Turkish Grand Prix	2020	6	1	Mercedes	Mercedes
Bahrain Grand Prix	2020	1	1	Mercedes	Ferrari
Sakhir Grand Prix	2020	5	1	Racing Point	Ferrari
Abu Dhabi Grand Prix	2020	1	1	Red Bull	Mercedes