Formula 1 Race Predictor

Iram Tazim Hoque, Zhenhuan Su, Jiangqiong Liu

Motivation

- **We found this interesting as someone who enjoys sports**
- **Growing Audience(4% increase in 2021-1.55Billion People)**
- **No public tool like this**

Related Work

1. **A Machine Learning Framework for Sport Result Prediction**
   - By Rory et. al. (DOI: 10.1016/j.aci.2017.09.005)
   - Talks about Artificial NN as a tool for doing generalized sports result prediction

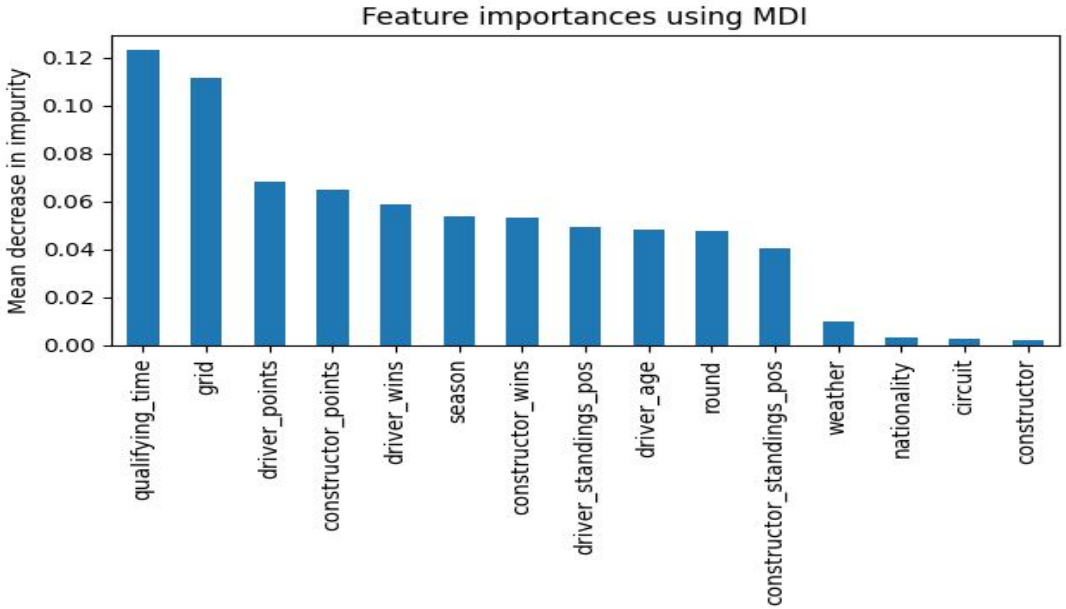
2. **Horse racing prediction using artificial neural networks**
   - By Elnaz et. al. (DOI: https://dl.acm.org/doi/10.5555/1863431.1863457)
   - ANN for Horse racing prediction

3. **A Learn-to-Rank Approach for Predicting Road Cycling Race Outcomes**
   - By Leonid et. al. (DOI: 10.3389/fspor.2021.714107)
   - ML Based method Learn-to-Rank for Road Cycle Racing

4. **Towards better analysis of machine learning models: A visual analytics perspective**
   - By Shixia et, al. (DOI: https://doi.org/10.1016/j.visinf.2017.01.006)
   - Talks about various interactive tools used in XAI

Data and Features

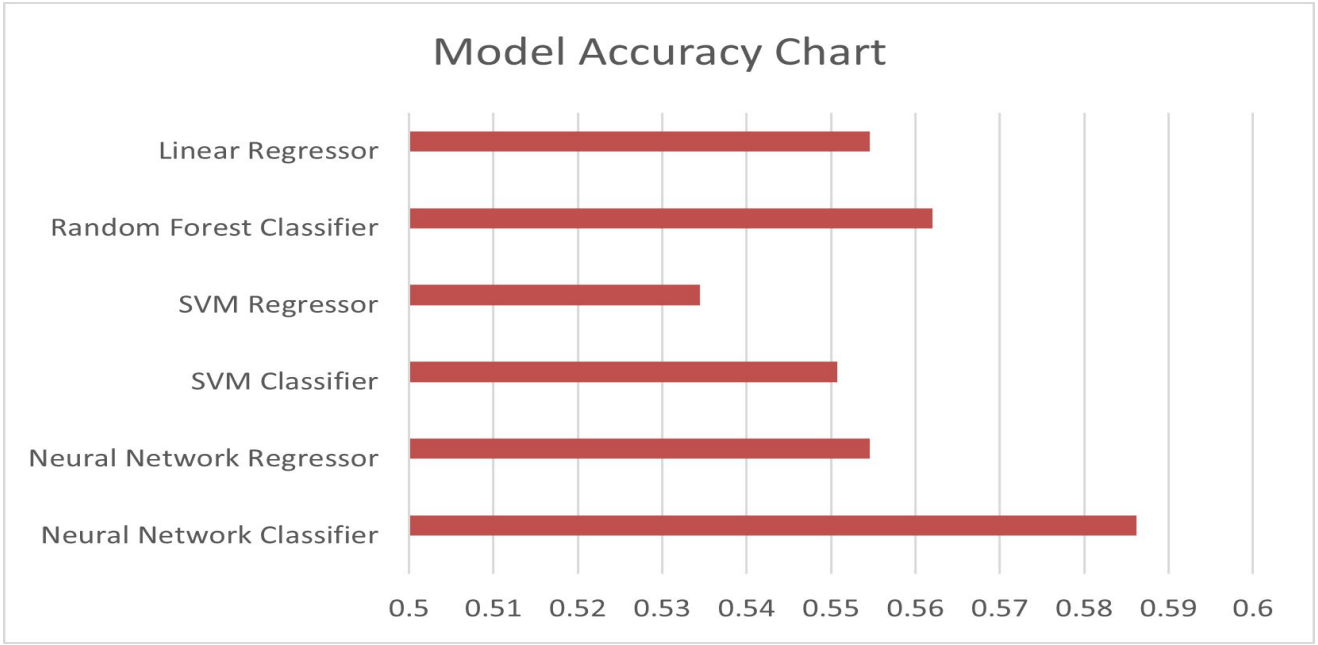
- **Ergast Developer API(**http://ergast.com/mrd/**)**



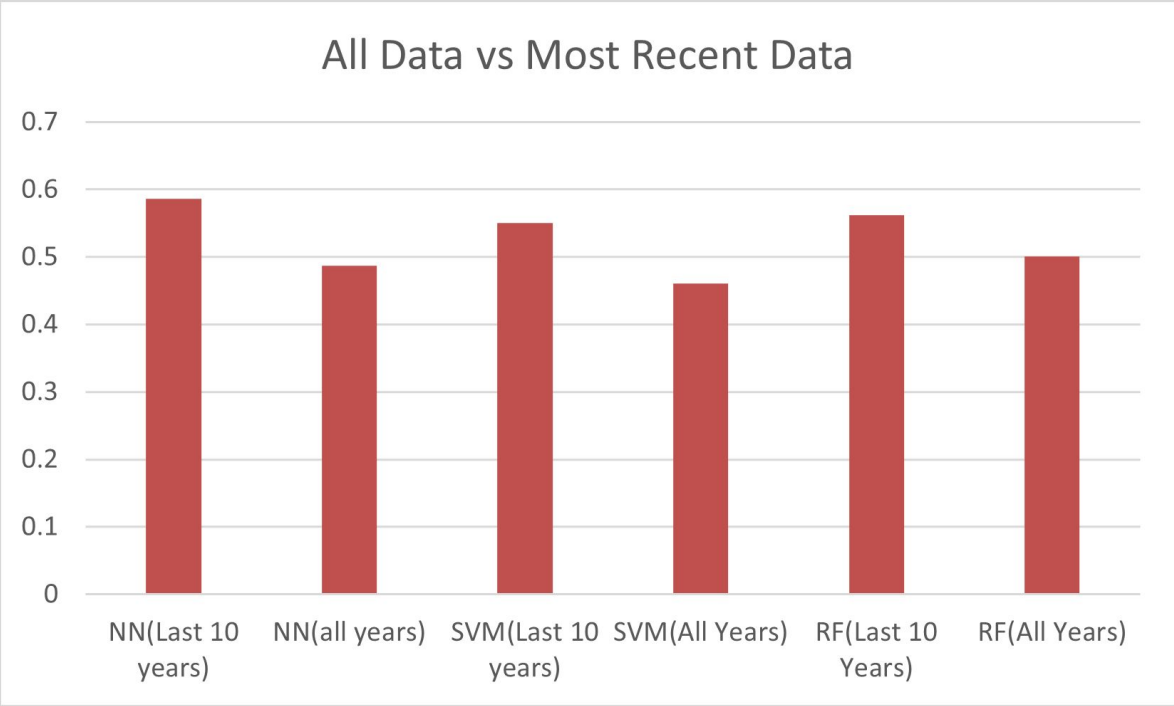
Model

1. **Classification**
   - Neural Network
   - SVM
   - Random Forest
2. **Regression**
   - Linear
   - SVM
   - Neural Network

Model

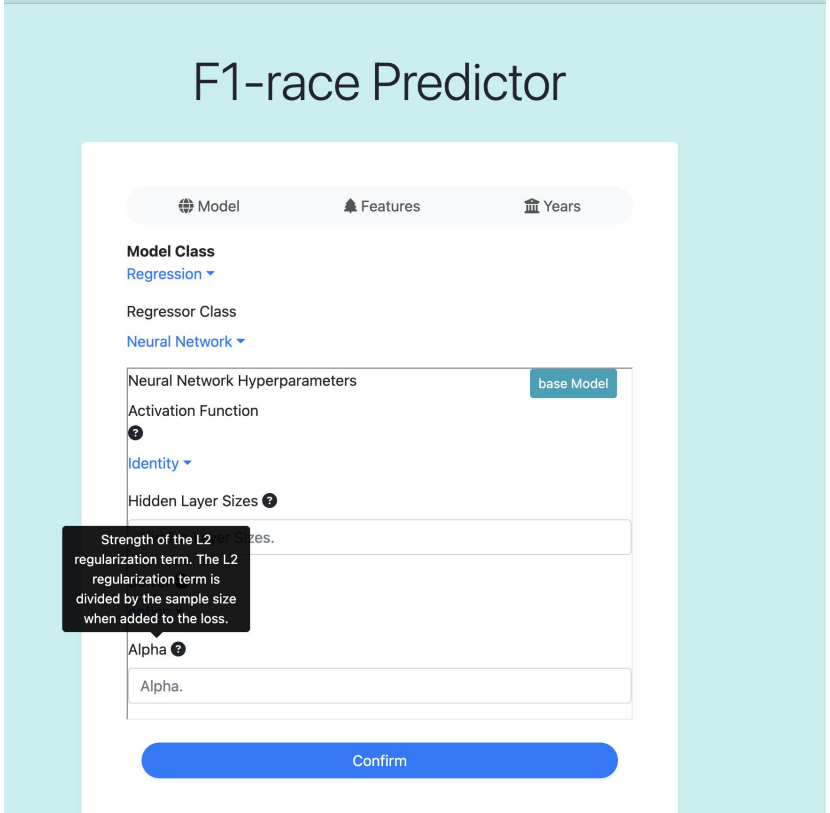

Model Accuracy Chart
Linear Regressor
Random Forest Classifier
SVM Regressor
SVM Classifier
Neural Network Regressor
Neural Network Classifier
0.5
0.51
0.52
0.53
0.54
0.55
0.56
0.57
0.58
0.59
0.6

Model


All Data vs Most Recent Data
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
NN(Last 10 years)
NN(all years)
SVM(Last 10 years)
SVM(All Years)
RF(Last 10 Years)
RF(All Years)

Demo

User Study



Hypothesis: it takes users less time to make a winner prediction using question mark icons showing additional information on hover

Participants:

- studying Computer Science at Purdue University
- pre-selected according to the criteria that they needed to have some background in Machine Learning

Experiment: between-subjects with 12 participants, task completion time

Interface design 1: showing the information about each required input directly on the page

Interface design 2: showing information only when users hover over question mark icons

Task: to predict a driver with the highest winning probability among several drivers with random order of input sections

Questionnaires:

- Pre-testing survey
- Post-testing questionnaire

Discussions and Future Work

Task completion time:

- Unpaired t-test with significance level 0.05
- T-score is 4.3348
- Two tailed P value is 0.0015

User feedback:

- Think aloud
  - Design1: the page looked a little messy with the amount of text and felt they were forced to read them
    the explanations were useful during the first round of prediction, but he would not like to read them afterward
    Design2: the information provided was helpful when they got confused
- Questionnaire
  - not much difference between these two designs regarding understanding the inputs needed in each section
    people with less experience thought the explanations were helpful for filling out the inputs for models
    most of them stated they trusted the winner prediction given by this tool
    most of them were not sure if other people interested in Formula 1 race betting would like to use this tool

Future work:

- usability testing with more participants who have done sports betting on other online platforms
- experiment with the novice and test if our tool helps the novice understand how winner predictions are made