Engineering Track: Final Project Proposal

B351

Basic Information

Project Title

Simulating Formula 1 Bets with Predictive Models

Team Members

1. Jacob Johnson

2. Benjamin A. Jacobs

3. Adit Patel

Short Project Statement
Our project aims to build a machine learning model that predicts Formula 1 race outcomes. By analyzing historical race data, weather conditions, and driver performance, we want to develop a system that can simulate betting strategies and make accurate predictions, including top finishes. The goal is to create a model that outperforms bookmaker odds and maximizes return on investment. We'll focus on overcoming challenges like real-time data updates and the ever-changing nature of racing.

1 Problem Space

1. Describe the problem space. What are the objectives, challenges, and constraints? What are some of the variations found in the problem space?
The objective of the model is to predict sports outcomes with a higher accuracy than bookmaker's odds, and maximize return on investment over time. Some challenges will be dealing with constantly changing sportsbook odds, balancing recent performance with long term statistical performance, and accounting for real time data. Some constraints might be accessing live data from bookmakers. Variations in the problem space consist of the modeling approach to the data, how data of different types will be interpreted, and how to weigh the data.

2. What are some historical attempts to tackle the problem space? Include links and references where appropriate.
Several past projects have attempted to predict Formula 1 outcomes using machine learning. For example, Will George developed a model that combined race and weather data to forecast driver performance.
Link 1: https://medium.com/@shuirmanm/beating-the-odds-simulated-formula-1-betting-with-machine-learning-50368c424e09
Link 2: https://www.elon.edu/u/academics/communications/journal/archive/spring-2024/spring-2024-heath-foster/
Link 3: https://mpra.ub.uni-muenchen.de/106821/

2 Algorithms

1. What solution are you proposing? How will this compare to historical approaches?
Decision Tree - This algorithm is appealing in the aspect of formula 1 where there are many competitors and only three winners. In the context of sports betting you can bet different people to get top 3, top 5, etc., which we would be able to classify with a decision tree.

2. What algorithms will you implement? Include links and references where appropriate. We decided to pick a sport like Formula-1 where there are many racers and one winner a decision tree may be the best algorithm to choose. This would also allow us to classify racers in categories as top 3 finish, top 5 finish, etc.

3 Third-Party Libraries and Technologies

If you intend to use third-party tools or technologies, please explain the following for each technology:

1. What technology will you be using?
   One of our main resources will be an API for getting sports data found at https://api-sports.io/
   Another library that we will use that will be very helpful is pandas, which will be used to organize our sports data in our python file.

2. What will it be used for / how will it assist you in your project?
   This api will be used to get the preliminary sports data before we clean it, and potentially do some feature engineering to add features that may be helpful.
   Pandas is going to very valuable because after we get our data from the api and store it in a csv file, we will use pandas to convert the csv data into a data frame to use in our script and use our api on.

3. How will you demonstrate your knowledge of the topic area despite off-loading work to the third-party technology?
   This api will only be used to get the data, all prediction and AI technologies will be implemented ourselves, and the data will just be used to give an answer.
   Pandas is only going to be used to organize our data, as previously said the algorithms will be implemented ourselves.

List this for **all** non-standard libraries you will use. For example, the first item for many Python developers might be numpy, and the first item for many Javascript developers might be jquery. You may always opt to use more third-party tools later by presenting the proposal modification request form to your mentors at one of your check-ins.

4 Project Goals

In this section, please list the specific action items that you intend to complete by the end of the project. Include a range of reach (A-range), target (B-range), and safe (C-range) goals. Each set of goals should build on the previous set. This section will serve as a rubric used to assign a majority of your overall project grade, so be as specific as possible. You may use a bulleted format. This section should be no longer than 1 page single-spaced.

4.1 C-range Goals

Submit a working algorithm that takes our data that is acquired and performs a prediction for the outcome at an accuracy level of below 50 percent. No live testing (only train and test on older data, for example, 2022 F1 race data).

4.2 B-range Goals

Similar plan as the C-range except the accuracy is below 75 percent, and potentially getting to test with live data. Incorporate some baseline sports betting strategies when finding profitable lines. This will potentially be done with an api to find live betting odds.

4.3 A-range Goals

Integrate real-time or frequently updated data using API calls to simulate a live betting environment. Achieve a model accuracy on test data when predicting top 3 or race winners above 75 percent.

Timeline

Week 1: Obtain all data
Week 2: Clean data and begin work on algorithm
Week 3: Train and Test algorithm
Week 4: Implement Betting Odds into algorithm

Acknowledgement

Instructor Mentor 1 _____   Signature _____

Instructor Mentor 2 _____   Signature _____

Team Member 1: Benjamin A. Jacobs   Signature: Benjamin A. Jacobs

Team Member 2: Adit Patel   Signature Adit Patel

Team Member 3: Jacob Johnson   Signature: Jacob Johnson