Project Writeup

Formula One - the zenith of motorsport, avant-garde technology and marvels in physics and automotive engineering, grabs the gaze of millions of spectators around the world. Featuring a formation of 20 drivers, belonging to one of 10 corporate constructors contending for the World Championships, for which points are accumulated throughout the season. Our group was interested in taking part in the sports track. We wanted to focus on a sport that was interesting, had plenty of recorded data to be used to analyze and train data, as there were many variables to be accounted for. As car enthusiasts, Formula One stood out to our group the most. We wanted to use data to predict a driver’s chance of winning a race on any given track. We knew that since we needed to find the correlation between the variables, such as pitstops, lap times, driver standings, qualifying position, and circuits, we needed to implement machine learning methods. We used Kaggle to find information about specific data sets. After careful optimization, we were able to sort through the values and pick out the relevant information needed for this implementation. While brainstorming for ways to train a machine learning model, we were introduced to Amazon’s SageMaker during their workshop. This gave us an opportunity to engage with SageMaker on a deeper level, and to even bring it into implementation for our project. By using the CSV data files obtained from Kaggle, we were able to train the model, and were able to fine tune it to run accurate predictions. By monitoring these factors alongside historical race outcomes, we utilized meticulously structured data to train a machine learning model using Amazon SageMaker. The machine learning model went through stages of testing and training in a regression model, from which we can anticipate meaningful forecasts of the chosen driver’s performance on any specified circuit. Our AI driven prediction system involves several key components - analyzing historical data and real time inputs to determine the probability of various race outcomes, condensing large datasets to extract the relevant information, to reduce computational complexity, and validating predictions by comparing them with supplementary sources to ensure accuracy and reliability. We kept training the model, and in the meantime, started to create a website to showcase this program. The user picks a driver, the circuit, and their starting position, and it gives real time data about their win probability.