Sabrina del Rosal

BrainStation: Sprint 0

Predicting Racehorses' Risk of Injury Based on Historical Race and Biometric Data

Problem Area:

Horse racing is a flourishing industry filled with fanatics who go to enjoy the race, bet, or be an inclusive part of it by owning racehorses. However, these horses undergo a lot of intense activity and can worsen their performance through time based on age, injury, and other factors. It is difficult to know exactly when and how a horse may begin to lower its winning chances throughout the races. A well performing predictive model could help provide valuable insights for trainers to prevent them from over pushing their horses causing performance decline and illness.

User:

- Horse trainers
- Jockeys
- Horse owners
- Veterinarians

Big Idea:

A machine learning model could use historical race data, biometrics records of individual racehorses, track intensity and/or conditions, and jockey statistics to help find patterns that may be associated with stress or risk of injury. By highlighting early indicators of fatigue, or poor performance, the model may be a crucial tool for racehorse trainers. This will be fairly similar to the injury-prevention models I have seen used for athletes playing basketball or football, for example. It can eventually allow trainers to have an idea as to when a horse might need to wean off racing for a bit in order to prevent massive injury that can lead to retirement.

Firstly, creating a model that can help predict winning positions will be in place to see if there are any trends in number of races a horse can feasible do well in. Then, biometric data will begin to play a part to see if height, weight, age, as well as pedigree plays significant roles in winning. We can look at conditions of track and weather as well to further investigate. If anything, I plan to begin with the most current year and create a model using that smaller dataset if needed to start small.

Impact:

- Helps trainers optimize race strategies and training schedules

- Can potentially reduce the risk of injury in racehorses by detecting performance anomalies that indicate fatigue or strain which saves the owners a lot of money in physical therapy and recuperation of their horses

Data Sources:

- **Kaggle: https://www.kaggle.com/datasets/hwaitt/horse-racing/data**
- **Kaggle: https://www.kaggle.com/datasets/jpmiller/equine-safety/data? select=pedigree.csv**
- **To find pedigrees: https://sporthorse-data.com/pedigree**

Some Extra Info that could be nice UK specific: https://www.britishhorseracing.com/regulation/reports-and-statistics/racing-statistics/

Alternative Approaches:

- Develop a model that can track for potential winning horses in the US for bettors to use.
- By using historic data of past wins based on track conditions, we can possibly predict the top horses for upcoming races.
  - https://horseracingdatasets.com/all/