Yehonatan Mileguir, Krishna Ravishankar, Aman Patel, Aidan Ehrenhalt

Description:

We want to analyze and classify a running data set in order to predict when someone will get injured in running. This prediction will be based on the attributes of the data set, such as number of running sessions per week, distance run before injury, etc.

Why it Matters:

Our idea is important because it will allow us to get a better understanding of what causes running injuries. This is useful to know when designing a training routine as factors such as running duration and frequency can be regulated if they affect chances for an injury. These results could even be extended to other sports, as factors such as intensity and rest are common across all sports and have an impact on the health of training athletes in different fields.

Motivation:

The motivation behind the topic is that one of our team members is interested in athletics and its intersection with data science. Many athletes world wide dont have the access or financial capabilities to have a coach that can tell them whether they are running too much. If there is a system that takes into account running data and can predict injuries it can help a lot of beginner runners avoid the injuries.

Plan of Action:

The data is already cleaned and organized. We plan on starting with basic statistical analysis and looking for correlations between certain variables and then trying to use these patterns to classify them into machine learning algorithms.

Measure Success:

We will classify success based on how many useful patterns/trends we can discover.

Data Collection:

The data has already been collected through a third party research organization.