Data TheraPy

Our goal was to help coaches and players make decisions regarding fatigue. We investigated whether we could predict the player's future fatigue.

We created a model for predicting a player fatigue based on the survey values athletes report along with recent survey results and game data.

The features we extracted in preparation for our model were:

- Total distance ran in game-action each day
- Total acceleration load in game-action each day
- Previous three days of fatigue ratings
- Previous three days of soreness ratings

We also min-max scaled all of the survey entries by player to account for personal variation in reported scores.

Our choice of model, XGBoost Regression, was influenced by the desire to understand the role of the features in prediction as well as the need to handle missing values.

XGBoost is an ensemble technique that employs gradient boosted decision trees. In short, it adds trees to together that will subsequently correct the error of the previous trees.

Our model produced an out-of-sample RMSE of 0.186 on a 0-1 scaled personal fatigue rating. (0 being most fatigue, 1 being least fatigue reported)

Examination of feature importance showed the usefulness of our generated features. Since importance doesn't indicate a numeric relationship between a feature and tomorrow's fatigue, we used simple linear regression and visualizations to look for counterintuitive relationships between the top features and tomorrow's fatigue and found none.

The Flask generated website makes the model usable in real-time. Players enter their survey results in the morning and the values are sent immediately to the model for a prediction for tomorrow's fatigue.

Players and coaches can then can adjust today's behavior with the prediction in mind. If tomorrow is a key match, a prediction of worse fatigue could encourage a player to get more sleep than normal. Likewise, a coach can reduce the intensity of today's workout.

The model we generated this weekend is by no means perfect, but Flask allows the model to be easily modified. The output can also be modified to suit player and coach preferences.