

Team: MeanCatGoesMew

Topic: Using PPSF to prevent injuries and get a better winning chance

1. Introduction

Our model predicts the individual's Physical Performance Score on Field (note as PPSF) in the game based on the wellness survey that the players have filled out on the morning of the same day. In other words, we will tell you how to efficiently read a player's self-reported wellness survey to see if they should go on the field or to see how long they should stay on the field. By predicting PPSF on the game day, we can possibly prevent injuries and have a better chance of winning the game.

So, why do we use PPSF? Can't we just use the fatigue scores that players provide us to decide if they should be on the field? We know that most of the survey results are based on athletes' feelings, but sometimes what you feel is not what your body is. This means we need to quantify the fatigue more accurately.

To make it more detailed, does feeling fatigue really means that you are fatigued? Does a higher score of readiness on training really indicates a better physical performance score? Obviously, they are not necessarily correct. Sometimes, we feel we are very ready maybe because we know the opponent is weaker than we are, so we are ready to win. Sometimes, we feel we are not ready maybe because we are nervous, but it turns out we perform better than we think we would do because being nervous can be helpful during competition. Thus, our task is to convert the subjective opinions of individual wellness into solid fact and determine which answers in the survey will contribute most in the Physical Performance Score on Field.

2. Method

- a. We get principal component from *gps.csv* to create our response variable by calculating the mean of standardized speed. This method gives us a standard to measure the overall performance of each player based on their past performance. This calculated principal component is our Performance Score.
- b. We extracted useful variables from *games.csv* and *wellness.csv* to get our explanatory variables. Since wellness contains observations for a lot of days, and our goal is predicting performance on game days, we only extract *wellness* data on game days. In this way we get our data set for modeling.
- c. Then, we eliminate more unrelated variables in our data set using exploratory data analysis. The predictors we eliminate have little correlation with our response variable.
- d. Finally, we use a random forest as our model to predict the Performance Score, and we find the variables that are important in predicting the Performance Score.

3. Conclusion:

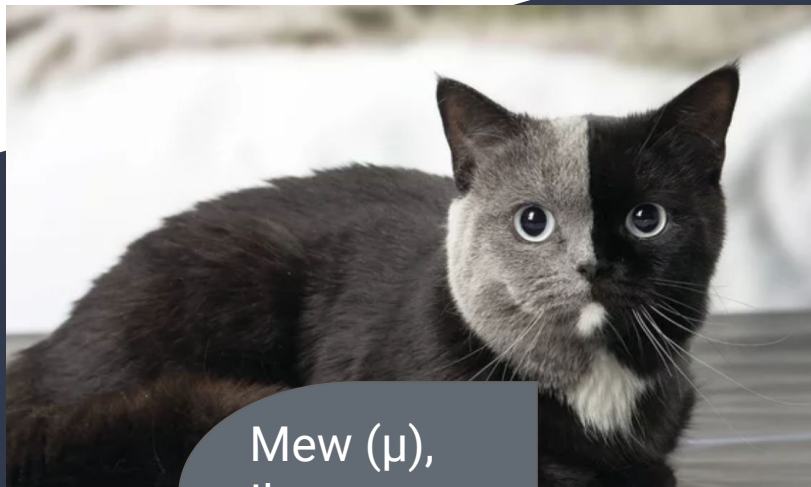
We have found that Sleeping hour, Desire, Soreness and training readiness, are the most important variables to predict our Physical Performance Score. This means that if a person has high sleeping hour, he or she is most likely to do well on the field that day. Same as desire and soreness. On the other hand, the individual answers of Irritability, Sleep Quality, illness, Menstruation and other variables that we didn't mention above are relatively unimportant when predicting individual's PPSF.

4. Future Work:

Since our PPSF is calculated based on the speed from GPS data, we suggest the company to provide more physical performance data when players are on the field so we could get a better PPSF result. At the same time, please continue to keep track of player's wellness evaluations on the game day, so we will have a larger database.

MeanCatGoesMew

Benjamin Liu, Radella Li, Celine Xu,
Eric Yang, Haochun Wang



Mew (μ),
I'm mean

PPSF

- **PPSF - Physical Performance Score on Field**
- **Predicting PPSF by Individual's Wellness Survey**

Why

- Convert objective answers to a subjective fact
- Better Winning chance and less injuries

Method

- Performance Score
- Predictors
- Modeling

- Performance Score
 - Principal component
 - *gps.csv* to get standardized speed for 1 player in 1 game
 - Mean of Standardized *Speed*
- Predictors
 - *games.csv* & *wellness.csv*
 - Used exploratory data analysis to eliminate insignificant predictors
- Modeling
 - Random Forest
 - Predictors of importance
 - SleepHours, Desire, Soreness & TrainingReadiness

Conclusion:



Desire: 1-7; Soreness: 1-7;

Training Readiness: 0%-100%

Future Work:
Better on-field data is
needed for calculating
better PPSF