Optimizing Risk Assessment for Duke University Student Athlete Injury Prevention

Maria Henriquez and Jake Sumner

Project Manager: Brinnae Bent, Project Leads: Dr. Tim Sell and Dr. Mallory Faherty

We are using athlete strength, flexibility, and balance data collected by the Michael W. Krzyzewski Human Performance lab to develop a machine learning model that can predict the likelihood of prospective injury in student athletes. We know an student athlete’s injury status before and after their test results. We also want to develop a dashboard for clinicians to aid analysis, and a dashboard for the student athlete report.

Overview

Goals

1. Improve KLab data pipelining with intuitive algorithms, strategies, and visualizations
2. Create an injury risk model using machine learning techniques
3. Develop individualized injury risk reports through an interactive web-application

Data Pipeline

[Diagram showing data pipeline stages: Directory with Athlete Info → Python Code that loops through all files and scrapes necessary data; processes data → Aggregate Data immediately visualized in web application dashboard]
Predicting Injury Risk through Machine Learning

- "Balance" variables contributed significantly to injury risk
- Random Forest and Gradient Tree Boosting models produced ~ 65% accuracy
- Accuracy can be maximized as our dataset increases
Web Applications

Application made for clinical use:
- Assess athlete risk and contributing variables
- Compare athlete stats to team
- Explore team distribution of risk

Application made for student athlete use:
- Explore where athlete stats stand compared to team
- Color-coded based off test value

Future Steps

- Cleaning up and updating our web apps as needed
- Getting our apps hosted on a Duke server with Duke authentication - security
- Receive outside opinions on applications from potential users (i.e. clinicians, coaches, athletes, etc.)
- Implement the apps into the KLab's current framework for reporting
- Continue to optimize and improve prediction model with additional data