**CS 4641 Project Proposal: Predicting NBA Player Performance Post-Injury**

**Intro/Background:**

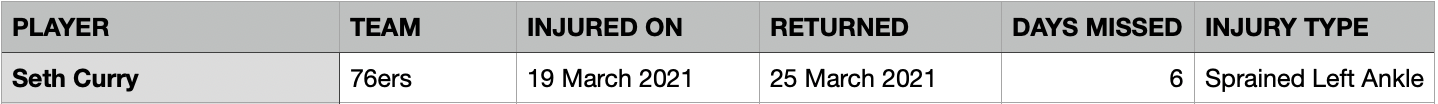
In professional basketball, determining when an injured player should return to the team is often a difficult decision that involves conflicting interests. Team pressure may lead to a premature return to action by the player, putting them in a compromised position physically. On the other hand, the player may want more time for recovery than necessary in an attempt to nullify the risk of re-injury and secure their financial future. Across 3 seasons, 56.8% of players have at least 1 game-preventing injury (Lewis, 2018). This indicates the frequency of the dilemma between returning or resting.

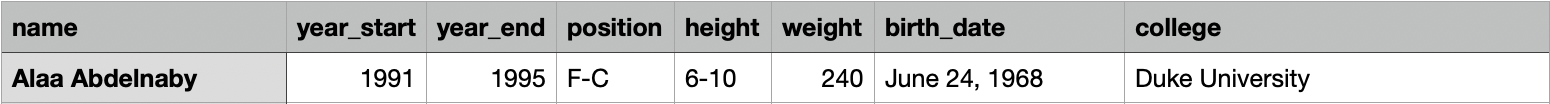
**Problem Definition:**

To lessen the divide between the conflicting interests of the team and the player, this project will formulate two models: one that predicts player performance post-injury and one that predicts whether or not a player will re-aggravate a previous injury. These predictive models would allow both the team and the player to make more educated decisions regarding the amount of treatment needed for certain injuries.

**Data Collection:**

Two data sets were used for the models below: one with NBA injury data from 2010 to 2020 and one with attributes (height, weight, birth date, etc.) for all NBA players from 1950 to 2018. Both data sets contain information retrieved from basketball-reference.com and prosportstransactions.com, which seem to be the go-to sources for NBA statistics. Additionally, sample entries from both these data sets are shown below.





Creating a final conglomerate data set took a few steps. First, the player attribute data set was significantly reduced to include only desired player attributes from 2010 onwards. Then, this player attribute data was paired with each injury data entry to get entries such as the example entries shown below.



Table

Description automatically generated

The “injury type” column gives the numerical encoding of the injury the player sustained. In the entries above, “7” denotes a left knee injury. Additional data was ascertained from the original data sets, such as that seen in the “reinjured” and “times previously injured” columns. The “reinjured” column gives a Boolean output that represents whether or not the player sustained the same injury in the following year. Additionally, the “times previously injured” column gives the number of the times previously that the player sustained that injury. For example, Kemba Walker had 7 previous left knee injuries before sustaining another knee injury on March 27th, 2021.

**Methods:**

Presently, we have data on all of the injuries that have occurred in the NBA from 2010 to 2020. Additionally, we have access to player statistics over the last decade. Cleaning and organizing this data will likely be the most involved and critical part of the model-making process because for each injury, we must combine the injury data (time missed, injury type, etc.) with the relevant player statistics before and after the injury. We will categorize the data by injury in order to keep variables that are not accounted for in our models as consistent as possible.

We will utilize regression modeling to determine which statistics are most relevant to the two models. After determining the most relevant input parameters, we will use a portion of our data to train a neural network that, given a particular injury and the relevant input parameters, outputs the predicted performance of the player post-injury (for example, the model may predict a 5% scoring decrease post-injury) and whether or not the player will re-injure himself in the same manner within some time frame. The remaining data that was not used for training will be used to test the accuracy of the models.

**Portential Results/General Discussion:**

Our models could lead us to identify the worst type of injuries—having the greatest negative effect on a player’s performance—as well as the likelihood that a player will reaggravate an injury considering multiple factors such as recovery time and game load after returning from injury. A study by Sports Health looked at injuries in the NBA over a 17 year period and concluded that “patellofemoral inflammation [was] the most significant problem in terms of days lost in competition” (Higgins et al., 2021). According to another study, ankle sprains, the most common injury, have elevated rates of incident ankle sprains (Herzog et al., 2019). Our models could come across similar findings and find correlations between the type of injury and effect on the player’s performance.

**References:**

Herzog, M. M., Mack, C. D. F., Dreyer, N. A., Wikstrom, E. A., Padua, D. A., Kocher, M. S., DiFiori, J. P., & Marshall, S. W. (2019). Ankle Sprains in the National Basketball Association, 2013-2014 Through 2016-2017. The American Journal of Sports Medicine, 47(11), 2651–2658. https://doi.org/10.1177/0363546519864678

Higgins, M. J., DeFroda, S., Yang, D. S., Brown, S. M., & Mulcahey, M. K. (2021). Professional Athlete Return to Play and Performance After Shoulder Arthroscopy Varies by Sport. Arthroscopy, sports medicine, and rehabilitation, 3(2), e391–e397. https://doi.org/10.1016/j.asmr.2020.10.001

Lewis, M. (2018). It’s a Hard-Knock Life: Game Load, Fatigue, and Injury Risk in the National Basketball Association. Journal of Athletic Training, 53(5), 503–509. https://doi.org/10.4085/1062-6050-243-17