Final Project Proposal: Predictive model comparison for the rating of FIFA19 players

[redacted]

1 Motivation

A sport that should require minimal introduction, football is the world's most popular sport. From the most southern tip of South America to far east Asia, Football spans the globe. A quick Google search will show that the estimated amount of football fans world wide is somewhere around 3.5 billion strong, almost half the human population.

While we've established the sport itself is prevalent in countries everywhere, what truly makes the sport, are the clubs, and players make the clubs. What are the exact credentials and skill set to quantitatively determine what separates players such as Lionel Messi from a Eden Hazard? More so, what goes into determining wage for these players? While this exact data may be hard to collect due to discrepancies, the FIFA video game has assigned players a wide variety of attributes. We will be analyzing this data set as it is most closely related to the real life skills and physical features of world class players everywhere. This data set includes skill set ratings such as strength of weak foot, tackling abilities, ball striking skills, and many more. Upon initial glance, the amount of variables can be overwhelming.

2 Method

- **Data preprocessing:** PCA
- **Method:** Classification, Regression
- **Algorithm:** Linear regression(regression), SVM(regression), KNN(classification)
- **Verification:** MSE, Purity score

Through the method of PCA we hope to reduce the dimensionality of the set to determine which factors actually contribute most to a players overall rating (how good they are). This rating directly influences wage, or so we currently believe. Other methods such as linear regression or polynomial can be performed with ease. The data set is a scalar set, and therefore regression models should be able to predict accurately. We can use these methods to help predict what wages or ratings a player is worth given a set amount of variables such as speed, height, dribbling ability, etc. The set can also be further analyzed through methods of classification, by determining what skills and physical attributes coincide with a specific position for example, midfielders tend to be shorter and lighter as they have to be more agile in passing and dribbling. A few methods than can be applied here are: linear classifiers, SVM, or the k-nearest neighbor classifier (KNN).