**Success of an NBA team**

**Branden Knorr**

**Introduction**

When discussing sports and certain teams that have had great success, the sports community tends to measure success by focusing on number of championships, playoff wins, or just overall record. By highlighting these core factors of what makes a team great, fans are able to debate and determine whether or not the success will carry over to the next season, and if the franchise has established a dynasty in their respective league. The goal of this project is to use various methods such as ridge regression, lasso regression, and linear regression to create prediction models that will use previous NBA team averages to try and estimate the number of wins for their next season.

**Related Work**

# To get an idea of measuring an NBA team’s success based off a previous season, information and models were taken into account from the article, The Complete History of The NBA, by the analysis group *FIVETHIRTYEIGHT*. *FIVETHIRTYEIGHT*’s work highlights the success of an NBA team throughout the history of the franchise and ELO ratings given after every game. ELO ratings for a team are given at the start of season with a league sum of zero. ELO ratings are determined by wins and losses, and then points are taken from the losing team and given to the victors. From the work that FIVETHIRTYEIGHT has completed we are able to see when a team has had their peak ELO for certain seasons, a league average ELO, and championship years for certain franchises. FIVETHIRTYEIGHT has used their ELO ratings in combination with a model that rates NBA players to make predictions on the odds for all teams to make the playoffs. FIVETHIRTYEIGHT tends to release a preseason prediction of teams with the highest odds to make the playoffs and ultimately win the NBA Championship.

# The work for this project has broken away from using FIVETHIRTYEIGHT’s methods for predictions and focused on average team stats to try and compete with the preseason predictions made by FIVETHIRTYEIGHT and focus on predicting number of wins for a team and the 8 playoff teams for each conference.

# DATA

# Data for this project was obtained through FIVETHIRTYEIGHT’s GitHub repository to recreate a similar ELO plots, and other data was taken from Basketball-Reference for the project’s predictions. The ELO data frame included 126314 rows and 23 columns which focused on who played in the game, when the game was played, and the outcome. Instead of recreating an ELO plot for the complete history of every team in the NBA the project only recreated 3 plots of the Lakers, Knicks, and Celtics.

# Chart Description automatically generated

# Figure 1: New York Knicks ELO rating over time

# In the figure above we see the graph that was produced by *FIVETHIRTYEIGHT* for the New York Knicks franchise. We see the graph starts in the 1940’s and ends in present day, and some of the figure’s features include an average ELO rating for the history of the franchise as well as years that the Knicks were world champions. We can see from the figure that the lowest ELO rating for the Knicks occurs in the 1940’s and late 2010’s, and the franchise’s average is right around a rating of 1497.62.

# Chart, line chart Description automatically generated

# FIGURE 2: Los Angeles Lakers ELO rating over time

Above is the recreated figure for the Lakers ELO rating throughout their franchise’s history. This plot was recreated in R and includes some of the same features as the figure from *FIVETHIRTYEIGHT*. The Lakers figure also has an average rating line of 1585.32, however, the plot does not include the championship seasons for the Lakers, because the project does not focus on predicting number of championships. Predictions for the 2015-2016 season were also taken from the *FIVETHIRTYEIGHT* in order to compare them to the predictions of the project’s models.

Data from Basketball-Reference was also taken for training and testing our models. Team averages from the 2014-2015 season were taken and designated as the training set, and team averages from the 2014-2015 season were taken and designated as the testing set. The training and testing tables both included 30 rows one for each team and 25 columns for the average statistics throughout the season.

**TABLE 1: Team Average Stats 2016**

Table

Description automatically generated with low confidence

Above are the average statistics for the 2015-2016 NBA season we can see that some of the variables measured throughout the season are field goals per game, field goals attempted, three point percentage, and more. The rows for the data were initially loaded in no particular order, however, it was processed alphabetically so it could make it easier to rename the rows after removing the team names. Team names were removed in order for the prediction models to perform better and then reassigned later on.

**Methods**

Linear regression was used initially to make predictions on the data set and all features of the data were used on the model. The model was implanted using the lm function in R and mean square error was calculated to evaluate the model. The linear regression model was summarized just so we could see what predictors were most significant to a team’s success, and the summarization highlighted field goals, steals, and turnovers. Ridge regression was then used next on the data to make a prediction. The ridge model used a grid and cross validation to determine the best lambda for the model, and by doing so some variables were reduced towards zero and this helps the prediction model not overfit the data. Once the best lambda was determined it was then input into the predict function on the 2015-2016. Following the prediction, the mean squared error was calculated again for the ridge model and prediction values stored away in a variable. Finally lasso regression was implemented on the data, and like our ridge model a grid of lambdas were created and then by using cross-validation the best lambda was determined. The lasso method allowed for the model to shrink some of the variables down to zero if need be and this too allowed for the model to not become overfit on the data. The predictions from the lasso model were saved into a variable as well, and the mean squared error was calculated as well as the r-squared value.

**Results**

All models were evaluated using mean squared error and r-squared values, and the 2014-2015 data was used to train the models before finally using them to predict on the 2015-2016 season. With regards to the ridge and lasso models implemented the MSE and r-squared values are very close and both models could probably be used interchangeably considering that there weren’t a large number of predictors and samples for the dataset that the project required. If the number of samples were larger then lasso it would be preferred as that usually performs better when the sample size is much larger than the number of predictors, and the opposite would be true for ridge regression. Both models performed very well, and in comparison, to the linear model they greatly outperformed. The r-squared for ridge, lasso, linear regression are as follows 0.6573546, 0.6573546, and 0.9304, and for MSE the values are 59.73, 60.8, and 589.4. We can see that ridge and lasso are very close with how their predictions turned out, and linear regression doesn’t fit the data well at all. All three model correctly predicted 13 of the 16 playoff teams that ended up making it to the playoffs, however, the main difference in these predictions could be made by grouping lasso and ridge together against linear regression. The linear model seemed to rearrange the order and seeding of teams, much more than lasso or ridge who fit the fata more accurately.

**Conclusion**

Overall, the models for the data seemed to perform well on predicting the number of wins for a team, however, the goal of this project was to try and beat or at least match the *FIVETHIRTYEIGHT* predictions of which teams would make it into the playoffs. The *FIVETHIRTYEIGHT* predicted all 16 playoff teams correctly and our models did not as all three only predicted 13 of 16 teams. This can still be seen as a success for the models though because there is a lot that can occur year to year between and NBA season with regards to injury, trades, and free agency. It should also be noted that the *FIVETHIRTYEIGHT* model took into account a hybrid model that relied on player ratings for the season which aided the predictions in the preseason. If this project were to be completed again player rating and efficiency would definitely be taken into account to better improve the model and give better odds for a team in the upcoming season.

**References**

**Fischer-Baum, Reuben. “The Complete History Of The NBA.” *FiveThirtyEight*, 7 Dec. 2015, projects.fivethirtyeight.com/complete-history-of-the-nba/#knicks.**

**https://www.basketball-reference.com/leagues/NBA\_2021.html**