NBA Prediction Model Report

Regis University

Anderson College of Business and Computing

MSDS 692

Eric Peter

NBA Prediction Model

Over the past year, sports betting has become one of the biggest markets in sports. It seems every week a new sports book opens, with promises of the best promotions and the easiest lines. Even ESPN's SportsCenter, live broadcasts of NBA games, and apps like Yahoo Sports are keen on pandering to the betting man, displaying spread and total points projections right next to the game itself. The Denver Nugget's broadcast partner Altitude Sports give updates at halftime highlighting the likeliness of player prop bets to hit by the conclusion of the game. It seems the old assumption that sports gambling is niche and unscrupulous has been long forgotten.

It is clear that there is a lot of money to be made for a bookmaker, as well as for the hopeful sports fan. But how do bookmakers create lines that are both accurate and yield a return no matter the outcome? The answer lies in incredibly complicated algorithms using advanced statistics that take nearly every factor imaginable into account— injuries, recent form, historical matchups, game location, and even public perception to name a few. Because there is so much money on the line, bookmakers need to ensure that their lines are as accurate as possible.

Background

This model attempts to recreate the process and algorithms by which bookmakers are routinely raking in money by the millions, while comparing it directly to Vegas' best. In no way do I expect this simple model to outperform that of the professional's, however there is much to be learned in even the most diluted model.

By using statistics updated live from nba.stats.com provided by an NBA API, data can be easily collected and analyzed. Betting sites provide their lines free online for anyone to see. VegasInsider.com displays all the major casinos lines for every major sporting event, and compiles them into a single "Consensus" line. For this project, we will focus solely on the Spread (winning margin) as well as the Over/Under (total points to be scored in the game).

*When reading the predicted Spread, TEAM A: -6.5 means that Team A is predicted to win by 6.5 points. The bettor can gamble that the winning margin will be more or less than this prediction.

*When reading Over/Under, 220.5 means that the book predicts 220.5 points to be scored. The bettor can gamble that there will be more points or less points than this prediction.

Analysis

With data provided by the NBA API, the bulk of the heavy lifting is then done by only a few lines to predict the Spread and Over/Under.

Statistics are recorded from the NBA API and a summary is displayed and saved in a text document. In this case, Denver both scores and allows more points on average, but only marginally. Also displayed are the standard deviation of points scored and allowed, average made three-point field goals (for and against), average made free throws (for and against), and average offensive field goal percentage.

```
RESULTS

Denver Nuggets at Chicago Bulls
OFFENSIVE STATS
        Pts  SD Pts  FG Pct   3Pt     FT
CHI  114.69   11.36    0.49  13.09  15.66
DEN  115.52    9.49    0.48  13.52  16.00


DEFENSIVE STATS
     Pts Allowed  SD Pts Allowed  3Pt Allowed  FT Allowed
CHI        114.84           10.76        11.72        19.0
DEN        115.52            9.49        13.52        16.0
```

Next, interactive graphs are created comparing points scored and points allowed via Plotly Express. Here we see a similar distribution for points scored between the Denver Nuggets and the Chicago Bulls.
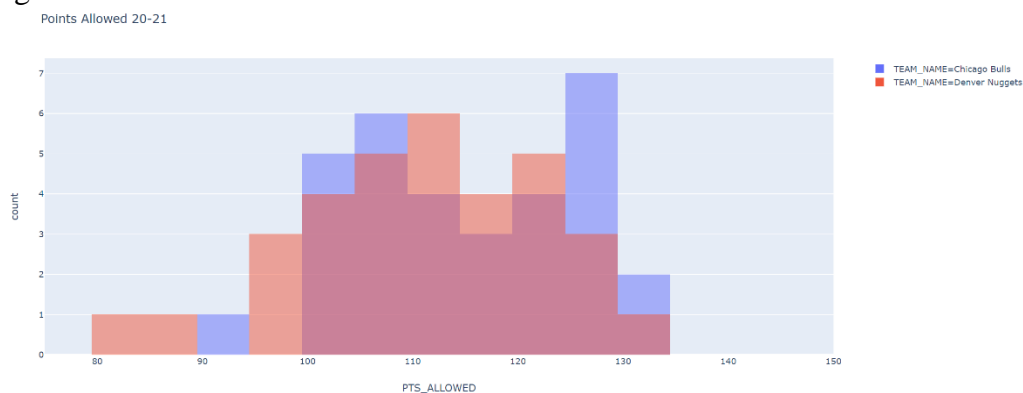
Games are predicted by running 100,000 simulations. The outcome it determined by randomly generating a score for each team using the mean and standard deviation points scored by that team. It is them averaged a randomly generated points allowed from the mean and standard deviation points allowed by the opposing team. This means that if a Team A doesn't score a lot of points per game, but Team B on average allows more points scored than Team A scores, we can expect Team A to score above their average points per game, and vice versa. When this is done for both teams, we can compare the randomly generated scores and predict a winner. By doing this 100,000 times we hope to eliminate as much of the statistical deviation as possible and determine the team that wins the simulation most often, and by how much they are expected to win by on average. Additionally, we can count the cumulative points and divide it by 100,000 to find the expected total points scored between the two teams.

```
for i in range(0, simNum):
    '''...'''

    homeSimPts = (rnd.gauss(homeStats["Mean Pts"],homeStats["Std Dev Pts"])+
                  rnd.gauss(awayStats["Mean Pts Allowed"],awayStats["Std Dev Pts Allowed"]))/2

    awaySimPts = (rnd.gauss(awayStats["Mean Pts"],awayStats["Std Dev Pts"])+
                  rnd.gauss(homeStats["Mean Pts Allowed"],homeStats["Std Dev Pts Allowed"]))/2

    homePtsCount = homePtsCount + homeSimPts
    awayPtsCount = awayPtsCount + awaySimPts

    if homeSimPts > awaySimPts:
        homeWinCount = homeWinCount + 1
    elif homeSimPts < awaySimPts:
        awayWinCount = awayWinCount + 1
    else: tieCount = tieCount + 1
```

Once all the simulations are complete, the results are compared and used to make a final prediction. The simulation results and predictions are written to a text file and saved.

```
100000 GAME SIMULATION:
Chicago Bulls won 49.71 percent of games. (49707 games)
Denver Nuggets won 50.29 percent of games. (50293 games)

SPREAD:
Denver Nuggets are favored at -0.5

MONEYLINE:
Denver Nuggets to win: -111
Chicago Bulls to win: +109

OVER/UNDER:
The projected total points is 230.0
```

Finally, this is output and compared to the lines provided by VegasInsider.com and stored in a text document.

```
Games for 03-01-2021:

                                Game Team  Vegas Line   My Line      Result
  0   DENVER NUGGETS  @ CHICAGO BULLS   DEN        -4.5      -0.5   Cover: +1.5
  1   DENVER NUGGETS  @ CHICAGO BULLS   CHI       225.5     230.0    Over: 230
```

Here we see a game played on March 1st between the Denver Nuggets and the Chicago Bulls. As we can see, both Vegas and my model predict Denver to win, by 4.5 and 0.5 respectively. Vegas predicts 225.5 points to be scored, while my model predicts 230.0 points. As seen on the right, Denver won by 6 (4.5 + 1.5) and 230 points were scored. In this case, the game had already been played, however the program can predict games being played presently or in the future. In this case, the results are not yet available and will instead simply display "N/A" in the results column.

Conclusion

I've found that my model often predicts smaller winning margins than Vegas does. This is in part because Vegas considers many more factors than just average points scored, and average points allowed. However, I have been pleasantly surprised with how much my model agrees with Vegas' lines, both in predicting an overall winner and in predicting the points scored. I will continue to build upon this model to make it more accurate and more efficient.

 With sports betting growing at an unbelievable rate, more and more people will soon be introduced to the nuisances of sports statistics and just how incredibly accurate these predictions really are. While I am not likely to start a sportsbook any time soon (as I would likely lose a lot of money if lines are off even slightly), it was fascinating to make my own prediction model.

References

Cheng, C. (2020, October 09). Predict NBA games, make money - machine learning project. Retrieved March 07, 2021, from https://towardsdatascience.com/predict-nba-games-make-money-machine-learning-project-b222b33f70a3

Dekanovsky, V. (2020, November 12). Histograms with Plotly EXPRESS: Complete Guide. Retrieved March 07, 2021, from https://towardsdatascience.com/histograms-with-plotly-express-complete-guide-d483656c5ad7#6791

Jee, K. (2019, December 23). How to Get NBA data using The nba_api Python Module (beginner). Retrieved March 07, 2021, from https://www.playingnumbers.com/2019/12/how-to-get-nba-data-using-the-nba_api-python-module-beginner/