Optimizing NBA Line-ups Using Genetic Algorithms

Soofi Punjani  
*Department of Electrical Engineering and Computer Science*  
*The University of Tennesee at Knoxville*Knoxville, TN, USA  
[spunjani@vols.utk.edu](mailto:spunjani@vols.utk.edu)

William Connor Parham  
*Department of Electrical Engineering and Computer Science*  
*The University of Tennesee at Knoxville*Knoxville, TN, USA  
 [wparham1@vols.utk.edu](mailto:spunjani@vols.utk.edu)

This project examined whether it is possible to use genetic algorithms to optimize a basketball team’s lineup with the key question being how close we can get to a “perfect” team. To do this, we used the pre-implemented genetic algorithm package Geneal in Python which uses roulette wheel style selection. Using this, we found that based on the 2021-2022 NBA season, it is possible to use genetic algorithms to optimize a basketball lineup if there is an adequate fitness function and player selection algorithm. The fitness function in this case being what dictates ideal player evolution and the player selection algorithm being how we match our ideal evolved player to the most optimal player in our dataset once the evolutionary algorithm has finished evolving said player.

Our results showed that after we found an appropriate fitness function for each positional role in a basketball team it was possible to find the highest rated player in each role. To confirm this, our final roster for the 2021-2022 season is:

* Center: Anthony Davis
* Power Forward: LeBron James
* Small Forward: Devin Booker
* Point Guard: Kyrie Irving
* Small Forward: Jayson Tatum

Keywords—NBA, basketball, field goals, blocks, steals, frontcourt, center, forward, power forward, small forward, backcourt, guard, shooting guard, point guard

# Introduction

## National Basketball Association Information

The National Basketball Association (NBA) is a professional basketball league that was formed in 1949. The NBA is a collection of 30 teams, broken down by two conferences with three divisions a piece. Each team has a roster of up to fifteen players, and at any given point during a game, each team plays three frontcourt players and two backcourt players. The frontcourt consists of a center and two forwards, a small forward and a power forward. The backcourt consists of two guards, a point guard and a shooting guard. Therefore, these are the following positions playing on the floor at any given point during a game:

* Point Guard
* Shooting Guard
* Small Forward
* Power Forward
* Center

Each team plays 82 games in the regular season. At the conclusion of the regular season, the top 6 teams in each conference clinch a playoff spot, while the teams seeded 7-10 compete in the play-in tournament for the remaining 2 playoff spots, for a total of 8 playoff spots in each conference. After the play-in tournament, the playoff teams in each conference play in the postseason. Teams in each conference play each other in a best-of-seven series, where the losing team is eliminated. The winners of each conference meet in the NBA Finals where they also play a best-of-seven series. The NBA has a stats page where they keep track of statistics for teams and players. This data has records of data that ranges back 10+ years. This includes information such as:

* Wins
* Losses
* Points scored
* Rebounds
* Shooting percentages

## Goals and Motivations

The overarching goal of this project is to use an evolutionary genetic algorithm to optimize an NBA lineup.  We chose to look at optimizing NBA lineups because we wanted to answer the question of whether we could apply genetic algorithms to the process of optimizing NBA lineups rather than interpreting each player’s statistics by hand and optimizing a team that way.

We found this topic interesting because we believe that there are real world applications to team building and sports betting within NBA lineup selection.  This project also coincided with the March Madness tournament (an elimination contest for the top college basketball teams), which just concluded a month prior to the writing of this paper, as well as the NBA Playoffs, with conference semifinals (or second round) games being played for conference finals berths at the time of the writing of this paper.

We selected this topic because we found that we each had an overlapping interest in the NBA as well as an interest in finding programming applications surrounding large scale datasets which are extremely easy to find on Kaggle and we agreed that this would help us cut back on data preprocessing.

Finally, we selected Genetic Evolutionary Algorithms for two reasons.  Firstly, we each found the genetic algorithm portion of class to be one of the easier topics to grasp as well as gaining a good understanding of these algorithms conceptually through the programming assignment associated with them.  Second, we talked about how genetic evolutionary algorithms are one of the most generally useful ways to apply Biologically Inspired Computation to a problem.  Because we did not know where to start when creating our team optimizer, we decided that a more general use case algorithm would be the most appropriate when starting out.

# Related Work

When doing research for this project we found a plethora of related works.  These works ranged from team optimizers in different sports to evolutionary algorithms applied to optimizing basketball lineups.  They are all related to our work in that they want to optimize a team lineup or predict a game’s outcome based on player statistics.

One related work focuses on creating an ideal fantasy lineup for the sport cricket, which is popular in England, Australia, South Asia, and parts of Africa. To create the ideal fantasy lineup, players are assessed in a random forest through recursive feature elimination, where their player metrics are considered and their selection is based on a modified genetic algorithm [1]. Such method produced better results than traditional team selection. Although our project has the same goal as this related work, our project differs because it focuses on optimizing each position independent of team dynamics.

Like the previous discussed work, another related work involves predicting a player’s performance in terms of fantasy points using logistic regression, random forests, and an ensemble method [2]. Even though our work is similar since it involves creating a lineup strictly based on individual performances, it is different from this work because it has no salary cap limit. Therefore, the fantasy lineups this work studies are more realistic than the lineup we generate.

An additional related work focuses on using a novel approach to determine the winning team in basketball; in other words, it tries to determine the winner of each basketball contest. Factoring for the complexity of the on-court dynamics for all ten players on the floor (five for each team), players are categorized into different player types based on their size, position, potential, and talent and skill levels, since each type of player impacts the game offensively and defensively in different ways. As a result, this method can predict the winner of a game with a high accuracy level and indicates the presence of a correlation between player stereotypes extracted from individual statistics and the outcome (win or loss) of that player’s team in a game [3]. While our work is also like this one, ours is different for these reasons:

* We found the best players for each position that had the best fitness value based on a combination of per-game field goals, blocks, and steals
* Although the players we aimed for had archetypes matching the clusters discussed in the study, our lineup was made up of individual prolific scorers with some of them able to block shots and make steals, so how they would operate with each other was not necessarily a factor with how our lineup was constructed

Another related work uses neural networks and recurrent neural network models to optimize a line-up based on the current game situation. This work focused on optimizing small-ball lineups designed to:

* increase the pace of the game through the number of offensive possessions and fast-break transition (defensive stop to offense scoring) opportunities,
* pass the ball at a high rate to find open and high-percentage scoring opportunities,
* and space the floor with most or all the players near the three-point arc to draw defenses away from the painted area near the basket to allow for players to drive into the paint and more efficiently score with less defenders or draw more defenders and pass to an open player [4].

Our project, albeit with the same motives, is different from this work because it does not:

* optimize current lineups,
* account for the current game situation,
* or account for how players operate with each other

Compared to the lineups this work focuses on, our lineup would be considered by many as an ideal lineup on paper yet would work in most if not all game situations only because it is an unrealistic lineup composed entirely of All-NBA, All-Star, and potential Hall of Fame players with an overwhelming advantage in terms of talent and skill.

An additional related work looks at Cricket team selection using genetic algorithms. Their algorithm evaluates players based on batting, bowling, keeping, fielding, experience, injuries, and finally that a team is composed of proper players. From here they encode each player into a string and employ crossover, replacement, and mutation. Their algorithm then reevaluates the fitness of the population and chooses the best individual. They employ both tournament selection and a more individualized selection for their unique problem [5]. This work mirrors our work; however, they look to optimize Cricket teams rather than basketball teams. This shows the interaction between genetic algorithms for different problems. We go through much the same process of using crossover, mutation, and replacement, however, the statistics we look at when evaluating players are completely different.

Another related work presents a more general way to optimize team selection using genetic algorithms. The algorithm they are proposing is an “adaptation of [the] island genetic algorithm”. To illustrate this, they use their implementation to solve an optimization problem related to Cricket teams. They propose that using a few conflicting restrictions combined with crossover in which the fitness of frequent solutions can be used to find the perfect solution will lead to an ideal team [6]. This heavily relates to our work. We both employ genetic algorithms to optimize a team lineup. However, because our work is trying to build the best team possible regardless of budget, we do not impose a restriction such as a salary cap that they discuss.

An additional related work aims to solve the problem of team building within sports. To do this they argue that they can use a “novel weights network clustering algorithm" to sort players in the National Basketball Association or NBA for short into categories. They then feed these categories into a model that considers the NBA’s “draft, trade, and free agency aspects of player acquisition.” Using the data that the model provides they then feed these player ratings to a new model that formulates teams based on rules such as a salary limit and how well individuals play together. These teams are then evaluated on a case study of the NBA from the 2019-2020 off-season [7]. This correlates with our work in that we also want to solve the issue of team building within sports. However, we want to look at basketball and use genetic algorithms rather than employing a neural network.

Another related work looks at play-by-play data to infer the aptitude of different five-man lineups employed by basketball teams. Using a plus-minus framework (a way to evaluate a player’s performance) incorporating a medley of statistics instead of only scored points they claim to better evaluate the effectiveness of a given line-up [8]. This relates to our work by examining a group of statistics rather than only evaluating players and team compositions from their points scored. They also must normalize their data, a process that we share because of the way in which we evolve our optimal players.

An additional related work aims to solve the problem of lineup selection within Soccer. They develop a model that integrates factors such as “spatial regions of player actions and defensive interactions with opponent players.” From this data they create a tool that will help coaches to better evaluate each player for their effectiveness from multiple viewpoints [9]. This relates to our work by also aiding in optimization. However, where as they create a machine learning model we opt to use genetic algorithms as a faster more generalized approach to problem solving.

Out of all the related works, the most notable related work we took inspiration from involves optimizing basketball lineups from the Association of Basketball Clubs in the top Spanish premier league using their own implementation of a genetic evolutionary algorithm [10].

However, we differed from these works in a few distinct ways:

* We did not use the same implementation of a genetic algorithm that they did.  Instead, we opted for the Geneal genetic algorithm because of its ease of use, online examples, and documentation.
* We implemented our own player matching algorithm.  This is what determines which individual in our dataset most closely matches our evolved player and assigns them to our team roster.
* We did not include a salary cap when looking at our ideal teams.  Because of this our algorithm chooses to aggressively pursue players who are best in their respective positions without considering the overall cost of a team and whether that team could truly exist.
* We chose to look at a different combination of statistics when evolving our players than what any articles we found looked at.  We chose to look at blocks, steals, and field goals per game because we believe that this best reflects a player’s personal skill rather than skill in a team setting and when evolving the best player on a role-by-role basis we thought it best to look at individual skill.

# Methods

The structure of our experiment is simple:

* Take a csv containing NBA player information and preprocess it to fit our needs
* Find minimums and maximums of each position to initialize our beginning population in our genetic algorithm
* Run the genetic algorithm for 100 generations
* Obtain ideal player
* Match ideal player to the closest existing real player
* Create lineup

The software packages we chose to use are straightforward except for our genetic algorithm package. For all our csv manipulation and preprocessing we chose to use numpy and pandas.  For our visualization we used matplotlib, and for our genetic algorithm we chose to use the Geneal library.  This library implements evolutionary genetic algorithms for both discrete and continuous problems.  In our experiment we chose to use the continuous implementation and only rewrote the fitness function to fit our needs.

The approach we are taking consists of six main steps (as shown by the process flow chart in Figure 2):

1. We found our dataset, which was in this case a CSV containing NBA player statistics and values for each NBA season.
2. We went through and pre-processed the data.  This includes the following tasks:
   * Filtering out nans
   * Sorting by positions and seasons
   * Re-labelling players classified as multiple positions
3. We created our fitness function.  To do this we looked at the statistics that best defined a player’s individual skills.  Because of this we chose to not look at the assists statistic which is commonly observed in basketball.
4. After defining the fitness function, as shown below in Figure 1, we evolved our optimal players for each position based on steals, blocks, and field goals on a per-game basis.

A screenshot of a computer program

Description automatically generated with medium confidence

Fig. 1. Fitness function

1. Next, we created our matching algorithm.  This is a necessary step because as our evolutionary algorithm evolves a best player it does not cap out at a number one player in a position.  Because of this our matching algorithm must rescale the ranges that our fitness function evolved to and then find the nearest player in our pool of real players.
2. Finally, we evaluate our results after finding our five most optimal players with one player in each position.  To do this, we create a bar chart (Figure 3) looking at the averages in each area on the composition we created averaged against 5 random players with one player coming from each position in a pool of all players.

Fig. 2. Process flow chart of project methods

In our experiment, we define a few key features:

1. Each player must be sorted into 1 of 5 pools with each pool corresponding to a position on a NBA team.
2. All players will be evaluated based on an equal fitness function regardless of their role.
3. No salary cap will be introduced to allow our algorithm to create an All-Star or All-NBA team.
4. All players will be evaluated based on blocks, steals, and field goals on a per-game basis.”

To investigate this, we will let our genetic algorithm run for different lengths of time to see how our player’s average fitness grows with generation.

# Results

Our results were better than expected.  Not only did the genetic algorithm succeed in creating optimal NBA lineups it consistently found players within the top five for each role even when operating with a less than ideal fitness function for each role.  Our resulting lineup not only included some of the largest names in basketball such as LeBron James and Anthony Davis; it also went above and beyond and placed them in their best position for that season even if it was not necessarily their main position. (For instance, LeBron James plays and has been a small forward throughout his career but even plays power forward and even center and point guard at times. Anthony Davis plays mostly power forward but can play center at times due to his abilities to offensively finish at and defensively protect the rim. During the 2021-2022 season as well as the time of the writing of this paper, both of those players are teammates on the Los Angeles Lakers, so when both players play together, Davis plays center with James at any other position except shooting guard or Davis plays power forward with James at any other position except shooting guard or center since Davis is the taller player on the team.)

One positive indicator for our results is that all the matched players, in their careers up to the 2021 to 2022 season, have been selected to at least one All-NBA team and at least three All-Star games. Additionally, all these players have started in NBA Finals games that their team won and often had the most field goals made, blocks, or steals compared to the rest of their team or even all players in a game for a number of games. Additionally, compared to our random lineup in which one random player from each position was selected, our lineup had a higher average score for all the statistical categories of per-game field goals made, steals, and blocks, as shown in Figure 3 below.

A picture containing text, screenshot, line, font

Description automatically generated

Fig. 3. Optimized lineup field goals, steals, blocks vs. random lineup

As shown in Table 1 below, each player in our resulting lineup has an average top 5 rating (except for small forward Jayson Tatum who has an average rank of 5.7, albeit in a deep pool of talented small forwards). Regardless, compared to other players in the same position, these players have one of the highest, if not the highest, average ranking (in Anthony Davis’ case, it is impossible for any other center to have a higher average ranking, but for other players, it is mathematically possible, although unlikely). Nevertheless, each optimized player having at least one of the highest average rankings indicates that the fitness function used can be a reliable one.



Table 1. Player rankings by statistical category and average rankings

Our results also depict how important having a good fitness function is.  As you can see in the following graphs, the fitness function fit all of the taller frontcourt players (Anthony Davis, LeBron James, and Jayson Tatum) very well. However, because of how it was weighted, it did not fit the smaller backcourt guards Devin Booker and Kyrie Irving as well. One possible reason for this discrepancy is that the frontcourt players are more likely to have more blocks per game because they operate nearer to the basket on the defensive end, giving them more opportunity to block shots close to the basket, and they are just simply more capable of blocking shots due to their larger sizes and wingspans. Because of these factors, it is paramount to optimize a fitness function for each position.

The following graphs show the mean and maximum fitness for the frontcourt center and forward (small and power) positions. As discussed above, the fitness function fit those players better and can be seen by how the mean and maximum fitness values quickly increase over the first 40 generations before gradually increasing over the last 60 generations.

A picture containing text, screenshot, plot, diagram

Description automatically generated

Fig. 4. Mean and maximum fitness for the center position over 100 generations

A picture containing text, screenshot, diagram, plot

Description automatically generated

Fig. 5. Mean and maximum fitness for the power forward position over 100 generations

A picture containing text, screenshot, diagram, plot

Description automatically generated

Fig. 6. Mean and maximum fitness for the small forward position over 100 generations

In contrast to the frontcourt players, the average and maximum fitness values for backcourt guard players take on a much different trajectory since they slowly increase before quickly increasing over the last 40 generations. Such a trend is indicative of a fitness function that does not fit the backcourt guards quite as well as it does for the frontcourt players.

A picture containing text, screenshot, diagram, plot

Description automatically generatedFig. 7. Mean and maximum fitness for the shooting guard position over 100 generations

A picture containing text, screenshot, diagram, plot

Description automatically generated

Fig. 8. Mean and maximum fitness for the point guard position over 100 generations

We also found that having a good matching function was almost as important as having a good fitness function.  When working on this experiment we found that even after we had our final implementation of a new fitness function our matching function had a tremendous impact on the team.  So much so that further optimizing our matching function improved our results from looking at no-name players to immediately looking at players like Anthony Davis who are consistently within the top five for each category they excel in.

# Discussion

In conclusion we have found that not only is it possible to use evolutionary genetic algorithms to optimize NBA lineups it is not difficult either.  After performing the necessary preprocessing and writing a few supporting functions and code we found our algorithm to be going beyond what we initially expected it to do.  This included placing players in their optimal position even if it is not what they were known to play as well as choosing players we may not have initially expected.

For our future work we plan on implementing a new version of our experiment in which a salary cap is imposed.  This will force the algorithm to get more creative and not allow it to select the most abled player in each role.  Instead it will have to look at the combined values of each player and select off of a totality within the team.  Alongside this we also plan to implement a fitness function appropriate to each role.  We found that even though our fitness function gets top five players in each position it still fails to treat all players equally and because of this some positions are optimized less efficiently because of what they are meant to do versus what our fitness function wants them to do i.e. score points.

##### References

1. Jha, A., Kar, A.K. & Gupta, A. Optimization of team selection in fantasy cricket: a hybrid approach using recursive feature elimination and genetic algorithm. *Ann Oper Res*(2022). <https://doi.org/10.1007/s10479-022-04726-z>
2. A. Winschel, “Daily Fantasy Basketball Lineup Optimizer,” Mercyhurst University, mercyhurst.edu (accessed May 13, 2023).
3. C. Osken and C. Onay, “Predicting the winning team in basketball: A novel approach,” *Heliyon*, vol. 8, no. 12, Dec. 2022. doi:10.1016/j.heliyon.2022.e12189
4. W. Yichen and H. Yamashita, “Lineup Optimization Model of Basketball Players Based on the Prediction of Recursive Neural Networks,” *International Journal of Economics and Management Engineering*, vol. 15, no. 3, 2021. Accessed: May 2023. [Online]. Available: https://publications.waset.org/10011928/lineup-optimization-model-of-basketball-players-based-on-the-prediction-of-recursive-neural-networks
5. S. N. Omkar, “Cricket Team Selection Using Genetic Algorithm,” *Cricket Team Selection Using Selection Algorithm*, 2003. Accessed: May 2023. [Online]. Available: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b364384991f4fc2cace48f2dce02dd658afd1ea1>
6. S. M. A. Burney, N. Mahmood, K. Rizwan, and U. Amjad, “A Generic Approach for Team Selection in Multiplayer Games using Genetic Algorithm,” *International Journal of Computer Applications*, vol. 40, no. 17, Feb. 2012. Accessed: May 2023. [Online]. Available: <https://d1wqtxts1xzle7.cloudfront.net/45344308/A_Generic_Approach_for_Team_Selection_in20160504-89179-1fwkcq1-libre.pdf?1462364663=&response-content-disposition=inline%3B+filename%3DA_Generic_Approach_for_Team_Selection_in.pdf&Expires=1684038816&Signature=KNsUGr8ae1mi5Rd5qKs-WSq04zdSKHeeB-T3jrknx2HL4XMU73kfy3vvOHkopoc0GGCHjmibkuJpt0PQbQfOHBY53pEEwJLYZ3HzWPnfhJm8KRf9z70dURedbZfB2J~vJqNA2YUBojsmuIclYQNvgeRFEFOphvBeVFfwEliFbRETWmieg5NMxuF6P3ZSBCBrqJrUiSZFe64Tn-5Q4Y6~94ZQLa3WbSsp~xTbXAsyVQiOa3qFZL6JruiOBMK-lBXUnaokE-RsmgYClJGM-KV-vRj8tXMwSmQSPPUVodr-TcmoAGrUMQ9kTdOZ4DUw-ciOhWu0L4oSSEcmkGP75K8mFg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA>
7. M. L. Muniz, *Sports analytics and optimization for team formation problems*, 2022. Accessed: May 2023. [Online]. Available: <https://repository.mines.edu/handle/11124/15458>
8. Luca Grassetti, Ruggero Bellio, Luca Di Gaspero, Giovanni Fonseca, Paolo Vidoni, An extended regularized adjusted plus-minus analysis for lineup management in basketball using play-by-play data, *IMA Journal of Management Mathematics*, Volume 32, Issue 4, October 2021, Pages 385–409, <https://doi.org/10.1093/imaman/dpaa022>
9. A. Cao *et al*., "Team-Builder: Toward More Effective Lineup Selection in Soccer," in *IEEE Transactions on Visualization and Computer Graphics*, 2022, doi: 10.1109/TVCG.2022.3207147.
10. Pérez-Toledano MÁ, Rodriguez FJ, García-Rubio J, Ibañez SJ (2019) Players’ selection for basketball teams, through Performance Index Rating, using multiobjective evolutionary algorithms. PLOS ONE 14(9): e0221258. <https://doi.org/10.1371/journal.pone.0221258>