SEng 474: Project Final Report

Predicting All-NBA Teams Using Machine Learning

Ryan Russell, Ken Dajani, Omar Abdulaziz, Justin Bao, Xiang Zhao, Akhil Tadimeti
August 7, 2020

1. Introduction

Our group decided to tackle the problem of predicting the All-NBA teams for the 2019-20 NBA season. Our initial reason for choosing this topic was a common interest in basketball among the group members. That being said, there is tremendous value in a model that can predict the All-NBA teams. For example, there are numerous websites and forums dedicated to betting on different outcomes of NBA seasons. One of these outcomes is, of course, betting on the All-NBA team selections. Furthermore, many NBA players have financial incentives built into their contracts if they achieve a selection to an All-NBA team.

Unlike winning a basketball game, where a team simply must score more points than the other team, being selected to be on the NBA first team does not have a clear criterion that must be achieved. There is no singular statistic that dictates the outcome of All-NBA selections. This complexity makes the problem interesting and promotes in-depth discussion about specific players and the voting process among the NBA community, especially at a time when the NBA is expanding in popularity. Specifically, the NBA broke viewership records during the 2019 Finals, with an "aggregate average audience of 20.5 million viewers" across Canada and the US for Toronto's first NBA Championship [1]. Considering each of these factors, this project not only provides many avenues to experiment with, but the results will have the potential to engross the NBA's worldwide fanbase.

At a high level, our approach to solving this problem involved using data from past NBA seasons to train several machine learning models. These trained models were then tuned to optimize accuracy and precision. Finally, data from the 2019-20 NBA season was fed into each model to predict the All-NBA teams for the current season.

2. Background

The All-NBA teams essentially consist of the 15 top performing NBA players for the entire regular season. There are three teams of 5 players, where the first team consists of the league's 5 best players, the second team is players ranked 6 to 10, and the third team is players ranked 11 to 15. It should be noted that players are also restricted by position. Each team can have two guards, two forwards, and one center. In addition to prestige, players may strive to achieve an all-NBA selection because of financial incentives built into their contracts that kick in when they are named to a team. The teams are selected through a voting process involving approximately 100 journalists who cover the NBA. The primary factors considered when selecting All-NBA players are basic counting statistics (points, assists, rebounds, etc.), advanced metrics, and team success. However, due to the human factor in voting, narrative and voter fatigue also play an important role in the selection process. For example, a player making an All-NBA team for five straight years prior to the current season may lead some voters to select said player regardless of their play in the current season.

2.1 Problem and Goals

Put simply, the problem is that we do not know which players will be selected to the All-NBA teams for the 2019-20 NBA season. This problem definition provides us with an explicit goal: to accurately predict the 15 players that will be selected to the All-NBA teams. While the NBA regular season would normally have concluded in April 2020, the season was paused due to the COVID-19 crisis, and resumed last week, on July 30, in Orlando, Florida [2]. Therefore, the 2019-20 All-NBA team selections likely will not be announced until at least September. An additional (and more difficult to achieve) goal that our group has for the project is to not only correctly predict which 15 players will be selected, but also to accurately predict which 5 players will be on the 1st team, 2nd team, and 3rd team. In some cases, a single vote separates a 1st team player and a 2nd team player, which greatly increases the difficulty of this second goal.

There are several possible outcomes for our project when considering these goals. Firstly, our model is inaccurate and fails to predict the 15 All-NBA players. Whether the model predicts one player or fourteen players accurately, the result is the same: the goal is not met. Secondly, it is possible that the model predicts each of the 15 All-NBA selections, but predicts the team which some players are assigned to incorrectly. With a robust model, this would likely be the expected outcome. Due to the human factors outlined in the selection process, even a very well-trained model may have difficulty accurately predicting a difference in voting due to an unexpected media narrative shift. A situation like this could easily lead to a player vaulting several spots to land on a higher team. The final possible outcome is that the model accurately predicts each of the 15 All-NBA selections and which team each player is assigned to. That is, the resulting model achieves both goals successfully.

3. Datasets

All of the data used in the experiments were from the Basketball Reference website, which houses a vast repository of NBA-related data that is used by professional NBA analysts and fans alike [3].

3.1 Preprocessing

For the past NBA data, which we used to train the models, we made several restrictions to reduce redundancy and improve both accuracy and training time. Firstly, we restricted the time range of the data to the 39 seasons from 1979-80, when the 3-point line was introduced, to 2018-19, given the prevalence of 3-pointers in today's game. We had initially planned for the training dataset to include all NBA player-seasons from the 1979-80 season to the 2018-19 season. However, we quickly realized that most of the players would never even be considered for selection to an All-NBA team (0% probability), so the model accuracy could be skewed and the majority of training data would be useless. Therefore, the second restriction was to include only players that had made either the All-Star team or an All-NBA team in a given season. After closer inspection, we discovered it is very rare for a player to make an All-NBA team without being an All-Star caliber player. In fact, fewer than 4% of the samples in our past NBA dataset were players that made an All-NBA team but not an All-Star team. As a result of these restrictions, our past NBA dataset includes 956 player-season samples. The current season NBA data, or unseen data, has also been restricted to those players who made the All-Star team, with the addition of Bradley Beal, who was widely seen as an All-Star snub in the NBA community and is averaging 30.5 points per game, ranking second in the NBA.

3.2 Key Features

A key correlation we found when examining which advanced stats to use when training our models was that the relationship between the Win Shares (WS) and Value Over Replacement Player (VORP) of a player had a strong correlation with making an All-NBA team. In other words, the higher a player's Win Shares and VORP were, the more likely they were an All-NBA caliber player. Descriptions of the aforementioned statistics can be found in the Glossary. Figure 1 shows Win Shares plotted against VORP, where red dots are All-NBA players, and blue dots are All-Stars who were not selected to an All-NBA team.

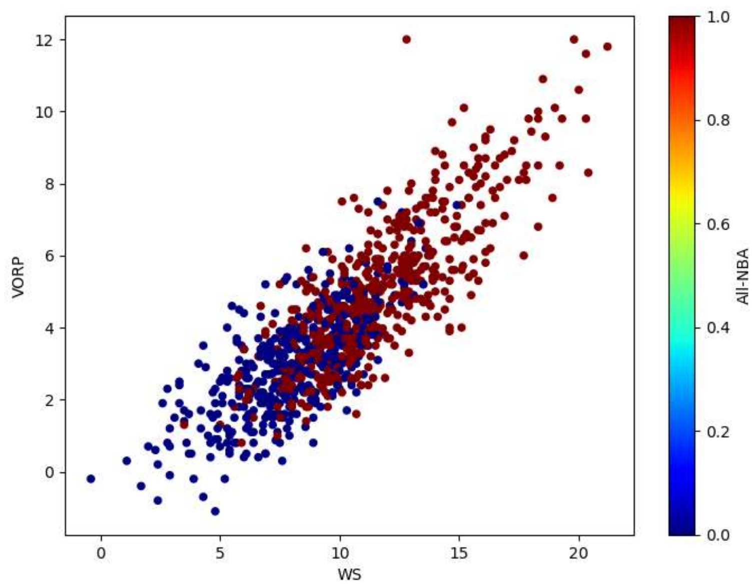

Figure 1: WS vs. VORP

As we can see from Figure X, there is a distinct linear correlation between WS, VORP, and All-NBA selections. These findings aided our feature selection process, and led us to include them in the datasets. The past NBA dataset includes the following 15 features: Games Played (G), Wins, Team Standing (Seed), Minutes Played (MP), Points Scored (PTS), Total Rebounds (TRB), Assists (AST), Steals (STL), Blocks (BLK), Field Goal Percentage (FG%), 3 Point Percentage (3P%) , Free Throw Percentage (FT%), Win Shares (WS), Value Over Replacement Player (VORP), Box Plus/Minus (BPM), All-NBA (Binary Class), All-Star (Binary Class).

4. Approach

The approach our team decided to take in tackling this problem was to divide into three sub teams of two members, develop independent predictive models, and compare the results of each model. The three selected models were

Neural Networks, Decision Trees (with Random Forest tests), and Support Vector Machines. The experiments performed for each model are discussed in the following sections.

5. Neural Network Model

For the neural network implementation, we decided to use Python 3 and Scikit Learn's Multilayer Perceptron Classifier [4]. After tuning the hyperparameters, we determined the best balance of performance and training time was obtained using the following: the ReLu activation function, two hidden layers of size 100, and a maximum iteration setting of 2000. Furthermore, we decided to take the average of 100 predictions after observing that on some runs the probability for certain players jumped by significant margins. A three-layer, 100-neuron neural network was initially tested, but this network's results didn't differ too greatly from the two-layer network. The final model outputs a probability ranging from 0 to 1 for each player. That probability is the likelihood of that player making it to an All-NBA team.

5.1 Hyperparameter Tuning

Before attempting to predict the All-NBA teams for the current season, we wanted to ensure a reasonably high test accuracy. To do this, we varied the test/training split of the past NBA dataset from 10% test to 35% test. Figure 2 shows the training accuracy, test accuracy, and precision of the model for each test/training split.

Figure 2: Accuracy vs. Test Case Split

As can be seen in the graph above, the 10% and 15% test splits provided a decent accuracy rating of 0.81. However, the precision of the models was less than ideal at 0.76 and 0.83 respectively. 20% and 25% splits produced the most accurate results, and the test accuracy hovered consistently around 80%. At this point, we were ready to make our predictions.

5.2 Results

The model was trained 100 times to obtain the average result and limit the variance of the experiment. The resulting probabilities can be seen in Figure 3.

```
['James Harden', 0.9829018770191732]
['Giannis Antetokounmpo', 0.9807273284686703]
['LeBron James', 0.9611870375044319]
['Anthony Davis', 0.9472211499139128]
['Nikola Jokić', 0.9245934515928947]
['Damian Lillard', 0.8995642567816857]
['Luka Dončić', 0.8779351084479494]
['Rudy Gobert', 0.7790426815033367]
['Kawhi Leonard', 0.7627333574905035]
['Domantas Sabonis', 0.6130727809461918]
['Jimmy Butler', 0.5742265983743272]
['Jayson Tatum', 0.5726438096431903]
['Bam Adebayo', 0.5719183533935113]
['Bradley Beal', 0.5426181165283402]
['Chris Paul', 0.508224647099904]
['Russell Westbrook', 0.4838440967080876]
['Donovan Mitchell', 0.4731195094801458]
['Pascal Siakam', 0.44043226021678905]
['Trae Young', 0.4281265436172492]
['Kyle Lowry', 0.3998909819019308]
['Khris Middleton', 0.36318094244094734]
['Devin Booker', 0.3598056911805172]
['Ben Simmons', 0.3393592279565996]
['Brandon Ingram', 0.316567088352286]
['Joel Embiid', 0.18428723616387133]
['Kemba Walker', 0.17234450167267085]
```

Figure 3: Neural Network Probabilities

From the list of probabilities, we can distinguish several tiers of players. Firstly, we have the top 3: James Harden, Giannis Antetokounmpo, and LeBron James, who have probabilities above 0.96 and also happen to be the MVP candidates this season [5]. From Anthony Davis to Kawhi Leonard, we have more All-NBA locks. The probabilities for this tier range from 0.76 to 0.94. Following this, we have a fairly steep drop in probability, down to 0.61 for Domantas Sabonis, who actually does not make any team due to the fact that only three centers will be selected. The rest of the players have probabilities below 0.57, but many still have strong chances to make an All-NBA team, except for Joel Embiid and Kemba Walker at the bottom of the list, with 0.18 and 0.17, respectively. The probabilities were then separated into three separate positions: Guard, Forward, and Center. These separations are displayed in Figure 4.

```
Guards:
['James Harden', 0.9829018770191732]
['Damian Lillard', 0.8995642567816857]
['Luka Dončić', 0.8779351084479494]
['Jimmy Butler', 0.5742265983743272]
['Bradley Beal', 0.5426181165283402]
['Chris Paul', 0.508224647099904]
['Russell Westbrook', 0.4838440967080876]
['Donovan Mitchell', 0.4731195094801458]
['Trae Young', 0.4281265436172492]
['Kyle Lowry', 0.3998909819019308]
['Devin Booker', 0.3598056911805172]
['Ben Simmons', 0.3393592279565996]
['Kemba Walker', 0.17234450167267085]

Forwards:
['Giannis Antetokounmpo', 0.9807273284686703]
['LeBron James', 0.9611870375044319]
['Kawhi Leonard', 0.7627333574905035]
['Jayson Tatum', 0.5726438096431903]
['Pascal Siakam', 0.44043226021678905]
['Khris Middleton', 0.36318094244094734]
['Brandon Ingram', 0.316567088352286]

Centers:
['Anthony Davis', 0.9472211499139128]
['Nikola Jokić', 0.9245934515928947]
['Rudy Gobert', 0.7790426815033367]
['Domantas Sabonis', 0.6130727809461918]
['Bam Adebayo', 0.5719183533935113]
['Joel Embiid', 0.18428723616387133]
```

Figure 4: Neural Network Probabilities by Position

From this list of player probabilities, the top two guards, top two forwards, and the top center available were selected for each team. The resulting All-NBA team composition is shown in Table 1.

All-NBA First Team	**All-NBA Second Team**	**All-NBA Third Team**
G: James Harden (98%) G:Damian Lillard (90%) F: Giannis Antetokounmpo (98%) F: LeBron James (96%) C: Anthony Davis (95%)	G: Luka Dončić (88%) G: Jimmy Butler (57%) F: Kawhi Leonard (76%) F: Jayson Tatum (57%) C: Nikola Jokić (92%)	G: Bradley Beal (54%) G: Chris Paul (51%) F: Pascal Siakam (44%) F: Khris Middleton (36%) C: Rudy Gobert (78%)

Table 1: Neural Network Predictions

As we inspect the second and third teams, the effect that specifying positions has is evident, particularly with the third team as we have Rudy Gobert (78%) and Khris Middleton (36%) predicted to be on the same team. The closest two players are Damian Lillard with 90% and Luka Dončić with 88%, and these two could very likely be swapped during the actual voting. A final interesting note is that the model predicts that Bradley Beal has a 54% chance of making an All-NBA team despite not making the All-Star team this season, which likely indicates that some narrative plays a role in underrating his performance.

6. Decision Tree and Random Forest Models

6.1 Decision Tree Model

Advantages such as robustness to noise, handling of irrelevant and redundant predictive attribute values, interpretability, and efficient training time make decision trees an apt choice for the presented datasets. Considering the limited size for the current season data (2019-20 season), the decision tree model has offered decent results in predicting All-NBA players accurately for the season.

The dataset was split into an 80-20 split, meaning 80% of the historical dataset was utilized to train the model and 20% of it was utilized to test it. After that was done, the decision tree model was presented with the current season's dataset to predict the current season's All-NBA teams. Once the datasets were decided upon, classification was performed and in order to achieve more accurate results, and several robust hyperparameters were utilized to further tune the decision trees. The implementation was realized utilizing Python 3 and was composed using a Jupyter notebook. Several Sci-Kit libraries were used in order to generate Decision Tree classifiers and generate the necessary plots.

As mentioned earlier, there are several parameters that the Sci-Kit library provides to tune the decision tree. Hyperparameters such as max depth, criterion, and splitter were considered valuable to tune the decision tree. Further analysis into these hyperparameters provided optimal values to increase decision tree accuracy.

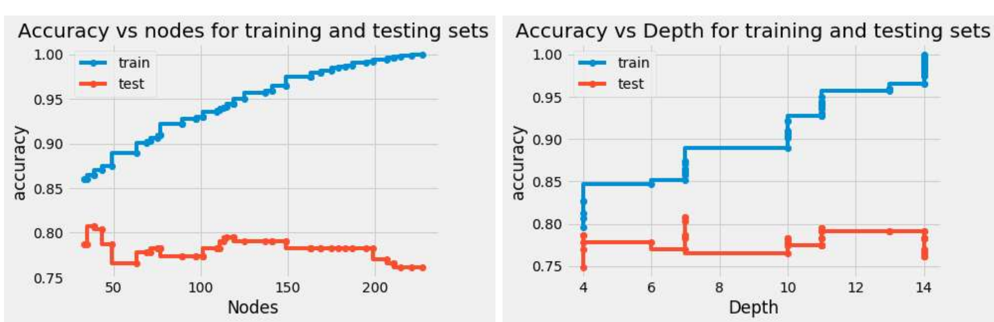

Figure 5: Accuracy vs. Nodes and Accuracy vs. Depth

Figure 5 offers insight into how certain hyperparameters were judged and utilized to predict optimal values for the different hyperparameters utilized. Considering these results, the following hyperparameters were utilized to tune the decision tree:

- Max Depth
- Criterion for measuring the quality of a split
- Parameter to split the node

These parameters were set to: max_depth = 7; Criterion = 'gini'; splitter = 'best'. Figure 6 shows some of the metrics achieved by the model, and Figure 7 displays the individual probabilities produced by the decision tree.

```
Accuracy Score: 0.812
Recall: 0.808
Precision: 0.840
F1: 0.824
Log Loss: 0.437
Area under ROC curve: 0.876
```

Figure 6: Decision Tree Metrics

```
['James Harden', 0.97]
['LeBron James', 0.91]
['Giannis Antetokounmpo', 0.9]
['Anthony Davis', 0.88]
['Kawhi Leonard', 0.84]
['Luka Dončić', 0.77]
['Bradley Beal', 0.7]
['Damian Lillard', 0.65]
['Nikola Jokić', 0.63]
['Rudy Gobert', 0.56]
['Joel Embiid', 0.55]
['Domantas Sabonis', 0.52]
['Russell Westbrook', 0.51]
['Pascal Siakam', 0.45]
['Trae Young', 0.44]
['Jayson Tatum', 0.4]
['Donovan Mitchell', 0.33]
['Jimmy Butler', 0.33]
['Kemba Walker', 0.3]
['Kyle Lowry', 0.25]
['Khris Middleton', 0.24]
['Chris Paul', 0.22]
['Bam Adebayo', 0.16]
['Ben Simmons', 0.12]
['Devin Booker', 0.11]
['Brandon Ingram', 0.04]
```

Figure 7: Decision Tree Probabilities

Finally, Table 2 outlines the Decision Tree predictions for the All-NBA teams.

All-NBA First Team	All-NBA Second Team	All-NBA Third Team
G: James Harden (97%) G: Luka Dončić (77%) F: LeBron James (91%) F: Giannis Antetokounmpo (90%) C: Anthony Davis (88%)	G: Bradley Beal (70%) G: Damian Lillard (65%) F: Kawhi Leonard (84%) F: Pascal Siakam (45%) C: Nikola Jokić (63%)	G: Russell Westbrook (51%) G: Trae Young (44%) F: Jayson Tatum (40%) F: Khris Middleton (24%) C: Rudy Gobert (56%)

Table 2: Decision Tree Predictions

6.2 Random Forest Model

For many of the same reasons as the Decision Tree model, the Random Forest model offers a good choice for predicting the All-NBA teams. The first thing to do was optimize the parameters of the model. For Random Forests, we had two parameters to optimize:

- Number of estimators
- Number of features used to split estimators

Regarding the first parameter, having less than 5 estimators lead to terrible accuracy, so we started at 5 estimators and gradually increased it by 5 for each consecutive test. This ensures that we found the point where accuracy is maximized while keeping the number of estimators relatively low. For the second parameter, the sqrt of all the features versus the full number of features was compared. We will test using a 80-20 split on the past dataset. The results of the test can be seen in Figure 8.

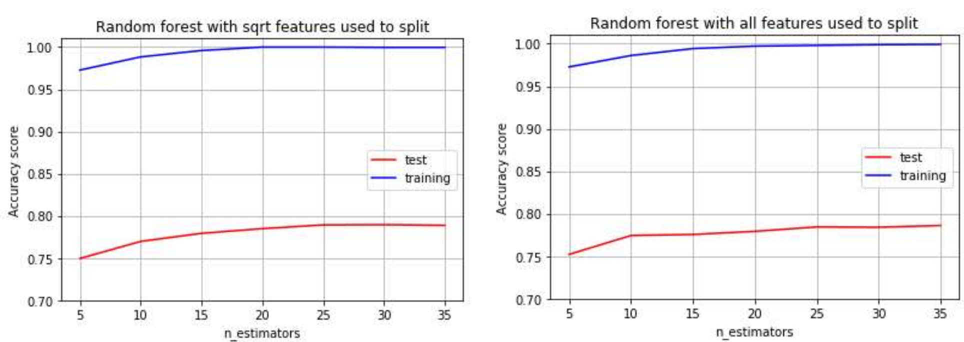

Figure 8: Accuracy Score vs. N Estimators

From these results, the optimal parameters are 25 estimators split using sqrt of the features. Using these parameters, we train the optimal Random Forest with the past dataset and predict the 2019-20 data. We will use the regressor version of the method to return the probability of each player being on an All-NBA team. Higher percentages indicate better chances for higher All-NBA team placement. The according output is shown in Figure 9.

```
0.932 James Harden (G)
0.8340000000000002 Giannis Antetokounmpo (F)
0.3060000000000001 Trae Young (G)
0.508 Damian Lillard (G)
0.7200000000000001 Luka Dončić (G)
0.5720000000000001 Russell Westbrook (G)
0.7180000000000002 Kawhi Leonard (F)
0.9000000000000001 Anthony Davis (C)
0.1640000000000001 Devin Booker (G)
0.8700000000000003 LeBron James (F)
0.12800000000000006 Brandon Ingram (F)
0.4660000000000001 Donovan Mitchell (G)
0.53 Pascal Siakam (F)
0.522 Jayson Tatum (F)
```

```
0.5319999999999999 Joel Embiid (C)
0.3540000000000002 Kemba Walker (G)
0.4040000000000001 Khris Middleton (F)
0.44000000000000006 Jimmy Butler (G)
0.658 Nikola Jokić (C)
0.2760000000000001 Kyle Lowry (G)
0.33600000000000013 Domantas Sabonis (C)
0.3680000000000001 Chris Paul (G)
0.3080000000000001 Ben Simmons (G)
0.21800000000000005 Bam Adebayo (C)
0.6040000000000001 Rudy Gobert (C)
0.2940000000000001 Bradley Beal (G)
```

Figure 9: Random Forest Probabilities

From this output we will select the six guards, six forwards, and three centers with the highest probabilities to form the Random Forest's predictions for the 2019-20 All-NBA teams. The predictions are shown in Table 3.

All-NBA First Team	All-NBA Second Team	All-NBA Third Team
G: James Harden (93%) G: Luka Dončić (72%) F: LeBron James (87%) F: Giannis Antetokounmpo (83%) C: Anthony Davis (90%)	G: Russell Westbrook (57%) G: Damian Lillard (51%) F: Kawhi Leonard (72%) F: Pascal Siakam (53%) C: Nikola Jokić (66%)	G: Donovan Mitchell (46%) G: Jimmy Butler (44%) F: Jayson Tatum (52%) F: Khris Middleton (40%) C: Rudy Gobert (60%)

Table 3: Random Forest Predictions

7. Support Vector Machine Model

For the Support Vector Machine method, the model was implemented using Python 3 and Scikit Learn's Support Vector Classifier [6]. After tuning several hyperparameters, we determined that the most relevant parameters are C, tol and possibility. Several different gamma inputs were used as well, and the observed result was that a scaled gamma value worked the best for our data input. With a maximum iteration setting of 2000 and possibility equal to true, the model takes a longer time to train but provides us with convincing and applicable results so that we can clearly identify the fifteen All-NBA players.

7.1 Hyperparameter Tuning

During our hypertuning process, we first tried to lower the error using a smaller C value, thereby increasing the regularization strength of the SVM. This tuning resulted in a high training accuracy of 89%. In order to have enough players identified as All NBA team players, the tol was lowered to ensure the outcome is a list of at least fifteen players with at least six of them belonging to the Guard position, six to the Forward position, and three to the Center position. The real bottle-neck to this classification is picking the 3 Centers since the number of players who are registered as a center is the fewest amongst the 3 positions.

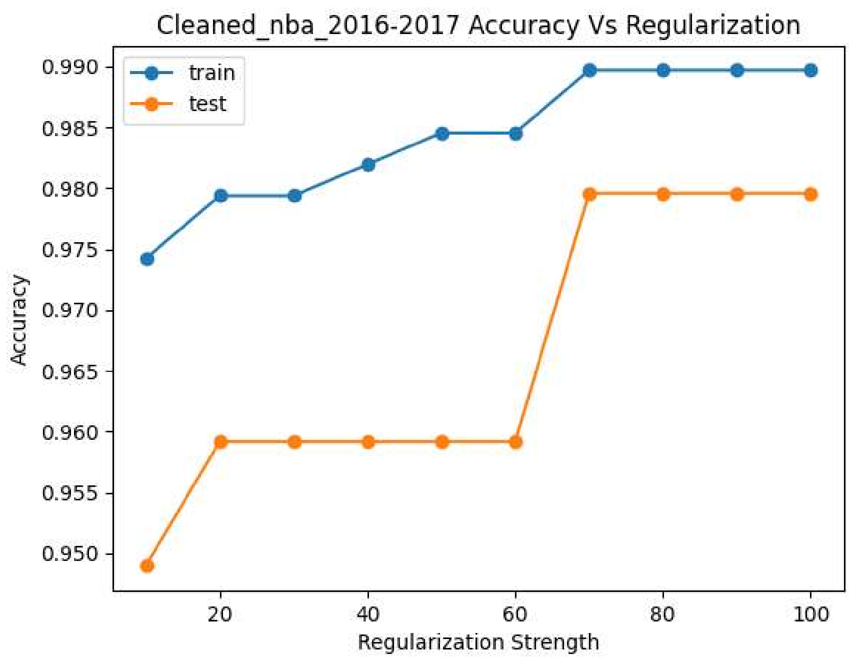

Figure 10: SVM Accuracy vs. Regularization Strength

Figure 10 shows the accuracy plotted against regularization strength. After the result met the conditions shown in Figure 10, the players with the highest probability were placed into the first, second, and third according to their position.

7.2 Results

The SVM model's final predictions are shown in Table 4. As discussed above, the Center of the predicted All-NBA Third Team, Rudy Gobert, has a confidence of 63% which is much lower than other players like

Jimmy Butler who didn’t show up in the final result. However, due to the position restriction, only the best six guards will be voted into the All NBA Teams. Comparing the final result with our raw input data, we observed that the top 3 most impactive attributes are points per game, assists per game and Win Shares (WS).

All-NBA First Team	**All-NBA Second Team**	**All-NBA Third Team**
G: James Harden (95%) G:Luka Dončić (90%) F: Giannis Antetokounmpo (93%) F: LeBron James (91%) C: Anthony Davis (86%)	G: Damian Lillard (89%) G: Trae Young (88%) F: Kawhi Leonard (86%) F: Khris Middleton (85%) C: Nikola Jokić (71%)	G: Bradley Beal (84%) G: Russell Westbrook (82%) F: Pascal Siakam (79%) F: Jayson Tatum (71%) C: Rudy Gobert (63%)

Table 4: SVM Predictions

8. Conclusion

In conclusion, each model performed very similarly in terms of test accuracy, which hovered around 80% (+/- 5%). However, there were some clear differences, one being the actual probability values. The average probability of the SVM predictions was relatively high at 83% between all 3 All-NBA teams, while the Random Forest and Neural Network models had an average probabilities of 64.4% and 74%, respectively. This would indicate that the SVM was more confident in its player predictions. As for the actual selection of the players, four names appeared in all predictions when selecting the first team: Lebron James, James Harden, Giannis Antetokounmpo, and Anthony Davis. However, there was a disagreement between the models regarding Damian Lillard and Luka Dončić as the Random Forest, Decision Tree, and SVM models had Luka Dončić on the first team, while the Neural Network model had Damian Lillard in his place. This can ultimately be chalked up to the similar level of performance exhibited by these elite guards.

Another difference between the models was how each one handled players on losing teams with incredible individual on-court performances. An example of this would be Trae Young, who is averaging 29.6 points and 9.3 assists per game while his Atlanta Hawks are the 26th overall seed (of thirty teams) [7]. The SVM model predicted Young to be selected on the second team with a 88% probability while none of the other models had him selected to any All-NBA team. Many experts believe that players with outstanding individual statistics on losing teams are less deserving of All-NBA selections. In this regard, we suppose that the Neural Network and Random Forest models are more accurate than the SVM. Overall, each model has strengths and weaknesses. From the results gathered during this project, a combination of each model’s predictions would likely be the optimal method to predicting future All-NBA teams with a high degree of accuracy.

9. References

[1] Sports Video Group. “Ratings Roundup: Toronto’s First NBA Championship Shatters Viewership Numbers in Canada,” sportsvideo.org [Online]. Available: https://www.sportsvideo.org/2019/06/19/ratings-roundup-torontos-first-nba-championship-shatters-viewership-numbers-in-canada/#:~:text=The%202019%20NBA%20Finals%20between,Cavaliers%20(18.5%20million%20viewers). [Accessed: June 24, 2020].

[2] NBA. “Everything you need to know about the 2019-20 NBA season restart,” nba.com [Online]. Available: https://www.nba.com/nba-returns-2020-faq#:~:text=The%20NBA%20season%20is%20tentatively,start%20date%20of%20July%2031. [Accessed: June 24, 2020].

[3] Basketball Reference. Basketball-reference.com [Online]. Available: https://www.basketball-reference.com/ [Accessed: June 23rd, 2020].

[4] SciKit Learn. Scikit-learn.org [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html. [Accessed: August 5th, 2020].

[5] Sportsraid. “The Final 2020 NBA MVP Ladder,” medium.com/sportsraid [Online]. Available: https://medium.com/sportsraid/final-2020-nba-mvp-ladder-basketball-awards-giannis-antetokounmo-lebron-james-harden-luka-doncic-brow-cbb3d766f427. [Accessed: August 1st, 2020].

[6] SciKit Learn. Scikit-learn.org [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html.[Accessed: August 2nd, 2020]

[7] Basketball Reference. “Trae Young,” basketball-reference.com [Online]. Available: https://www.basketball-reference.com/players/y/youngtr01.html. [Accessed: August 7th, 2020]

[8] Basketball Reference. “NBA Win Shares,” basketball-reference.com [Online]. Available: https://www.basketball-reference.com/about/ws.html. [Accessed: August 6th, 2020]

[9] Basketball Reference. “Glossary,” basketball-reference.com [Online]. Available: https://www.basketball-reference.com/about/glossary.html#:~:text=VORP%20%2D%20Value%20Over%20Replacement%20Player,to%20an%2082%2Dgame%20season. [Accessed: August 6th, 2020]

[10] Basketball Reference. “About Box Plus/Minus (BPM),” basketball-reference.com [Online]. Available: https://www.basketball-reference.com/about/bpm2.html. [Accessed: August 6th, 2020]

Glossary

Stat	**Description**
Win Shares (WS)	A player statistic which attempts to divide up credit for team success to the individuals on the team [8]. **Side note: This statistic is incredibly complicated, so if one would like to know the exact calculation of this statistic, refer to the cited source.*
Value Over Replacement Player (VORP)	A box score estimate of the points per 100 TEAM possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season [9].
Box Plus/Minus (BPM)	A basketball box score-based metric that estimates a basketball player's contribution to the team when that player is on the court. It is based only on the information in the traditional basketball box score--no play-by-play data or non-traditional box score data (like dunks or deflections) are included. BPM uses a player's box score information, position, and the team's overall performance to estimate the player's contribution in points above league average per 100 possessions played [10]. **Side note: Similar to WS, this statistic is incredibly complicated, so if one would like to know the exact calculation of this statistic, refer to the cited source.*