Using a Deep Neural Network to Predict the All-NBA Teams for the 2019-20 NBA Season

Introduction

Inarguably, across all professional team sports leagues, the ultimate prize and goal are to finish the season being coronated champion. Lifting the championship trophy above their heads is the reason many professional athletes push their bodies to the limit and hone their skills. Be it a multimillionaire professional or a child playing ball at the local field, what is better than winning a trophy?

A league championship is the culmination of an often-grueling season for a team. It recognizes a team as the best of the best among all others as they leave defeated opponents behind. However, being a team accolade, what a championship trophy does not do is answer who are the best individual players to the sport. While there is no doubt what team won the final game and is raising the trophy at the end of the year, questions among sports fans often gravitate to individual players. Who is good? Who is bad? Who is better? These all questions not uncommon to hear in any sports fan circles. Thankfully, professional sports leagues have introduced individual awards recognizing player performances for the season. And while some may argue that there are too many such awards (Most Valuable Player, Rookie of the Year, Gold Gloves, among many others), I, for one, find myself caring more about these individual awards than the final championship trophy.

One of these individual awards is the All-NBA Team honors awarded by the National Basketball Association. This is an end-of-season award given to the best players in the league in recognition of their performances. These award winners are voted on by a panel of sportswriters and broadcasters and are composed of three five-man lineups. Since 1988, these lineups are then distinguished as First Team, Second Team and Third Team composed of two guards, two forwards and one center to mirror a traditional five-man basketball team. Before the 1988 NBA season, and going back to the inaugural season in 1946, only a First and Second team were voted on and selected.

Selection to the All-NBA Teams is based on a weighted point voting system with five points for a First Team vote, three points for a Second Team vote, and one point for a Third Team vote. The five players with the highest point totals make the First Team, with the next five making the Second Team and a final five rounding up the Third Team for a total of fifteen players having the distinguished honor. However, for players is more than honor and bragging rights on the line with these awards. Often, very lucrative bonuses and contract kickers are contingent on the player making the All-NBA teams.

We set out to predict what players will be the recipient of All-NBA awards using machine learning. The premise is, we will use a machine learning classifying algorithm to determine the probability distribution for each player to make any of the All-NBA Teams or miss out altogether for the 2019-20 NBA season. We will then weight the probabilities with the point system described above to mirror the real-life point system for All-NBA voting. For this task, we will create a machine learning model based on a deep neural network (DNN), train the model using historical data from past seasons winners and losers, and use it to predict the yet to be determined winners for the 2019-20 season.

Data

Our dataset starts with the 1988-89 NBA season and spans the period until the 2019-20 season. The 1988 NBA season is when the NBA started the current format of three separate All-NBA teams. As mentioned before, prior to that only two teams where selected therefore to model the current format that is as far back as we should go with historical data. The data consists of all countable stats like points, rebounds, shooting percentages, etc. as well as several so-called advanced stats such as win shares, VORP, PER, etc. This information is present for all players in the league during the seasons from 1988 to 2019. In total the database consists of over 12,000 distinct player entries and 50 tracked metrics. For our machine learning model, we will trim the total number of metrics to predict a player's salary to the following 28 features:

Shooting	Passing	Rebounding	Defensive	Advanced	Playing Time
2PA, 2P	AST/G, TOV/G	ORB/G, DRB/G	STL/G, BLK/G	OWS, DWS	G, GS
3PA, 3P	USG%		PF/G	ORTG, DRTG	MP/G
FTA, FT				OBPM, DBPM	
EFG%, TS%				PER, VORP	
PTS/G					

A regular NBA season consists of each team playing a total of 82 games. For our features, we decided to use metrics that are based on a per-game basis This is to account for seasons that were shortened by lockouts, such as the 1998-99 and 2011-12 seasons with 50 and 66 games respectively, and our target season shortened due to the COVID-19 global pandemic. Additionally, we try to avoid features that are colinear with each other that are highly correlated to one another. For example, total rebounds are the sum of offensive and defensive rebounds, therefore we do not include total rebounds in our features. Unnecessary features decrease training speed, decrease model interpretability, and, most importantly, decrease generalization performance on the test set. We hope that these features will paint an accurate picture of a player's performance, and thus have the predictive power to determine All-NBA winners.

We will examine how some of the features correlate with award winners. Theoretically, only the best players each season should be earning All-NBA honors, so we expect the density of award winners to be heavily skewed to top-performers playing heavier minutes.

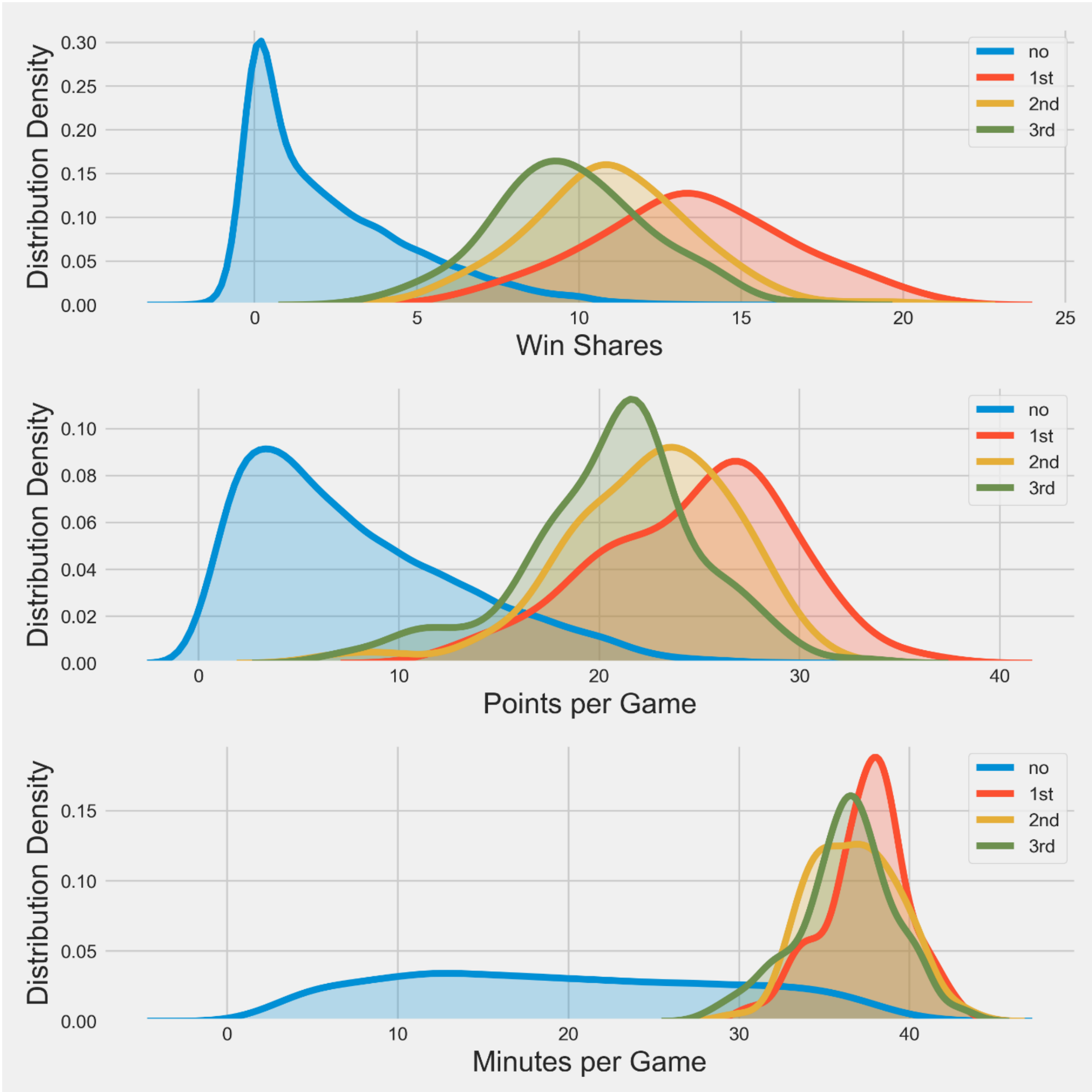

Figure 1. Density distribution of All-NBA award winners and losers from 1988 to 2019 for Win Shares, Points per Game and Minutes per Game.

As expected, we can see that historically players that win All-NBA honors are in the upper echelon of Win Shares, Points scored, and Minutes played. We then expect there to be a relatively large positive correlation between these features and the probability to win the award. Note that there is a distinct separation in the peak of each distribution for 1st, 2nd and 3rd Team indicating that a distinction exists between each. We hope that our machine learning algorithm, in combination with all other features, can separate not only winners versus losers but also the distinction between winners.

Methodology

Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data so that they carry out certain tasks, in this case assign probabilities for individual players to make the All-NBA First, Second or Third Team or miss the honor completely based on their performance on key metrics. With our 12,000-plus samples, consisting of players statistical data and labels indicating past All-NBA winners since the 1988-89 NBA season, we are ready to train a machine learning model to make salary predictions. These All-NBA labels will represent our class labels for each data point. For example, consider Michael Jordan who won First Team honors during the 1995-96 season. Aside from the other statistical features, in this example our class labels would identify Jordan winning First Team with a 1 and all other labels with a 0:

Player	1st	2nd	3rd	NO
Michael Jordan	1	0	0	0

We will randomly split the data using a traditional 70/30 split to train/test our model. That is, we randomly select 70% of our total data to train the machine learning model, and the remaining 30% to test it for accuracy.

The types of machine learning algorithms differ in their approach, the type of data they input and output, and the type of task or problem that they are intended to solve. Since our model is built of a set of data that contains both the inputs and the desired outputs, we can say this is a type of supervised learning algorithm. Additionally, because we are predicting discrete classes (1^{st}, 2^{nd}, 3^{rd} and NO), rather than a continuous variable, this is a classification problem.

In particular, DNN machine learning algorithms are a type of artificial neural network with multiple layers between the input and output layers. Data flows from the input layer to the output layer to find the correct mathematical manipulation to turn the input into the output, whether it be a linear relationship or a non-linear relationship. At first, the DNN creates a map of virtual neurons and assigns random numerical values, or "weights", to connections between them. The weights and inputs are multiplied and return an output between 0 and 1. If the network did not accurately recognize a pattern, an algorithm would adjust the weights and cycle again. Each cycle of the network were weights are updated and the output compared to the validation data is called an epoch.

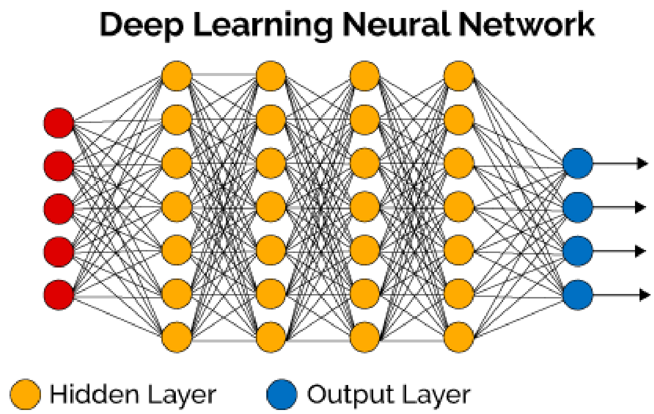

Figure 2. Visual representation of a Deep Neural Network with interconnected nodes (neurons) and 3 hidden layers.

We will use Keras to construct a classifying DNN to predict the probability distribution that each player will be in class equal to the respective All-NBA Team. The model consists of 4 dense layers, with an input layer of 28 nodes, to match the number of features, two hidden layers with 14 and 8 nodes respectively and finally an output layer of 4 nodes to match the number of target classes. The initial layers of our network use Sigmoid activation function while the output layer used a SoftMax activation function. The SoftMax activation function converts inputs into a discrete probability distribution over the target classes. As described below, this is the key for our point scoring system.

We will train our DNN using historical data from the seasons from 1988 to 2018 with our 28 features and class labels. The probability distribution obtained is weighted using the point system for the actual All-NBA selection. For example, if our model predicts Player X to have a 50% probability to be in First-Team, 30% to be in Second Team and 20% to be in Third Team, the total point tabulation based on this distribution would be:

$$Player\ X\ Total\ Points = 0.50 \times 5 + 0.30 \times 3 + 0.20 \times 1 = 3.6\ points$$

With this scoring system, the maximum score possible is 5.0. The predicted First All-NBA Team is selected from the two highest scoring guards, two highest scoring forwards and the highest scoring center. The next five players by position are the Second Team, and the next five are the Third Team. Note, the machine learning model does not account for position so there may be higher scoring players that get left out to accommodate the 2 guards, 2 forwards and 1 center format.

Results

In machine learning overfitting occurs when a model corresponds too closely or exactly to a dataset and may therefore fail to fit additional data or predict future observations reliably. To prevent overfitting of our model, we used early stopping to stop our training process early if we find we have adequate accuracy as measured by a loss function. Instead of training for a fixed number of epochs, you stop as soon as the validation loss rises because, after that, your model will generally only get worse with more training. Additionally, to further prevent overfitting, we added dropout layers to our model. Dropout layers can be an effective way to prevent overfitting by randomly dropping some of the connections between layers during training. This means that their contribution to the activation of downstream nodes is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass. The effect is that the network becomes less sensitive to the specific weights of nodes resulting in a network that is capable of better generalization and is less likely to overfit the training data.

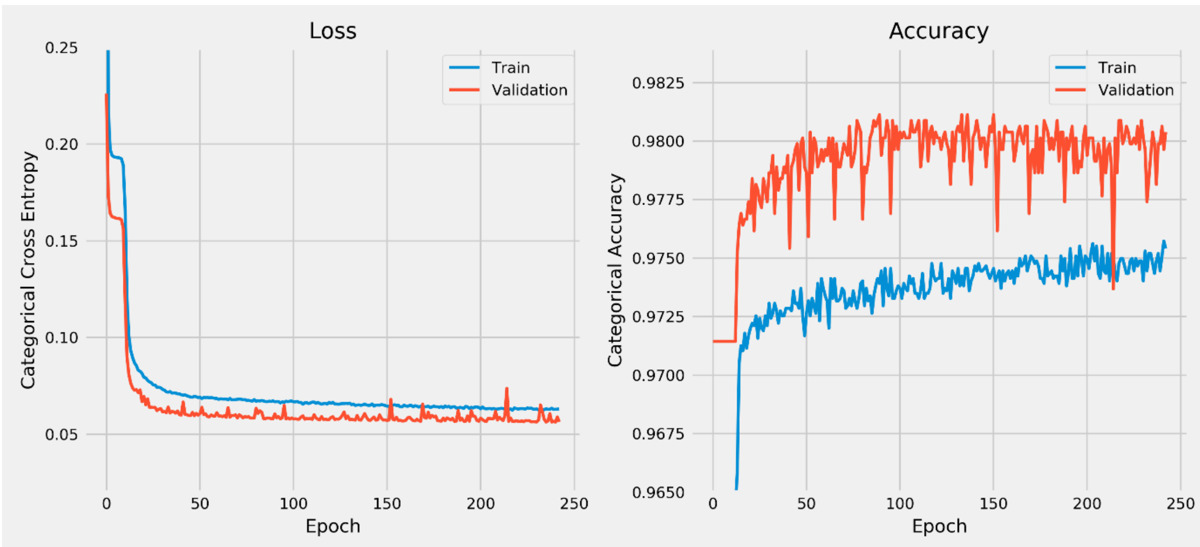

Figure 3. On the left, plot of loss function over number of epochs on the train and validation sets. On the right, plot of model accuracy on train and validation sets over the number of epochs. Both plot show that the model is not overfitting.

As previously mentioned, the machine learning model does not account for player position so there may be higher scoring players that get left out to accommodate the 2 guards, 2 forwards and 1 center format. Then to select the All-NBA teams in the traditional format of 2 guards (G), 2 forwards (F) and one center (C), we used the position the player played the most minutes on for the season.

The tables below summarize the First, Second and Third All-NBA teams along with position and overall score, ordered by position:

Position	**All-NBA First Team**		**ALL-NBA Second Team**		**All-NBA Third Team**	
	Player	**Score**	**Player**	**Score**	**Player**	**Score**
F	Giannis Antetokounmpo	4.55	Anthony Davis	2.64	Jimmy Butler	0.25
F	LeBron James	4.46	Kawhi Leonard	2.15	Khris Middleton	0.49
C	Nikola Jokić	1.20	Rudy Gobert	0.82	Joel Embiid	0.60
G	James Harden	4.07	Damian Lillard	2.88	Russell Westbrook	0.86
G	Luka Dončić	3.72	Trae Young	2.38	Bradley Beal	0.83

Conclusion

Although as of the time of this writing, the All-NBA teams for the 2019-20 NBA season have not been announced, our predicted teams are a reasonable prediction. Pre-season prediction by analysts, and mid-season reports very closely match our predictions. For example, a prediction published by Forbes Sports Reporting on April 2020 perfectly matches our First Team when accounting for the fact that Anthony Davis played 65% of his time as in the forward position.

We could further expand our model to predict All-NBA Teams in advance by asking the question - Given a player's performance this season, what is the probability we will make the All-NBA team next season? This would allow us to predict a player's performance in advance given a player's previous season and considering a player's projection and natural progression/decline. However, this analysis would be much less accurate and much more complex than what we did here. It's hard to predict how players will progress or regress between seasons or how different team schemes and environments affect their production. Then is the unpredictable chance of injury. During a single season, a player's performance is much easier to predict so a retrospective look would be more accurate than a future prediction. A more realistic addition to our model would be team record. While the All-NBA team is an individual award given on player's performance, it is well known that team records affect the voting. Players doing well on winning teams are more likely to win these awards than players putting up empty stats while losing.