Predicting NBA Game Results with Machine Learning

Corey Hemenway

Master of Applied Economics

Washington State University

Motivation

NBA front offices constantly look to improve upon their decision-making processes, ultimately to win more games. Game prediction on its own yields minimal real value to front offices; however, pairing powerful machine learning techniques with model explanation tools can provide useful information on what attributes most affect game outcomes. Front offices can then focus on the most important attributes associated with winning, whether they be improved through coaching strategy, player development, or player signings (acquisition of natural skill). Another common pitfall seen in NBA game prediction models relates to "cherry-picked" model performance. Machine learning models provide a systematic approach to analyzing game outcomes from existing data. This concept applies to machine learning by picking a portion of the data that may be easier to predict when testing a model.

Research Question

The objective of this study is to determine the previous five game metrics that contribute most to increasing winning probability as well as give examples of actions, such as player or coaching investments, that a front office may take based on the most important metrics and how they are assumed to be proven upon. This study uses random forest and gradient-boosting machine learning models to predict NBA game outcomes with accuracy that matches or outperforms existing models and incorporates SHAP (Shapley Additive Explanations) values to show how each feature contributes to predicting whether the home team won or lost. These insights can help coaching staff and front offices identify the most influential factors in game outcomes, as well as explore ways to improve their team and coaching staff.

Background

This section intends to give some background on the relevant empirical strategies utilized in previous NBA machine learning models as well as highlight some improvements in both the models and their applications. A previous study predicted NBA outcomes with both the random forest and gradient boosting models, among other machine learning models, and utilized SHAP values[1] to interpret results[2]. The study found the gradient boosting model to have the highest accuracy; however, the study's goal was to predict game outcomes with data from the game itself and only utilized data across three seasons. The study also found means to reduce the number of variables while still maintaining model performance by utilizing a correlation matrix.

Another study provided NBA teams with valuable measures of the determinants of winning or losing games but also used statistics derived from the games themselves[3]. The study utilized a larger dataset, spanning twenty seasons, and found the random forest model to have the highest accuracy. Their results found that field goal percentages are most important when looking at same-game predictions, which intuitively makes sense as the metric directly correlates to scoring efficiency across both teams.

This study utilizes statistics derived from a team's previous five games, including valuable percentages and performance metrics, as well as team-level controls that highlight the team's record and other important traits[4]. This study expects the significance of variables to vary when compared to the studies analyzing same-game statistics, as solely utilizing pregame statistics addresses how a team's strength and current level of play across different metrics translate to their performance in their next game. This strategy provides value in game prediction, but most importantly, it identifies which team-specific metrics can be optimized before the game takes place, as opposed to indicating what metrics during the game are most indicative of results. In-game metrics are a byproduct of front-office decision-making before the

---

[1] SHAP values explain model results by measuring each feature's impact in prediction for each specific game. Hence, each game has unique SHAP values for each feature. SHAP values are additive in that their computations are computed independently, giving a direct and interpretable impact of each feature.
[2] Zhao, Kai, Chunjie Du, and Guangxin Tan. 2023. "Enhancing Basketball Game Outcome Prediction through Fused Graph Convolutional Networks and Random Forest Algorithm"
[3] Junwen Wang. 2023. Predictive Analysis of NBA Game Outcomes through Machine Learning. In The 6th International Conference on Machine Learning and Machine Intelligence
[4] See appendix for information on how early season games are predicted and the problems with quantifying team strength across seasons

game took place, which is what this study intends to identify and optimize. The first study utilizes only three years of data, while the second utilizes data from before the league saw a boom in three-point attempts starting around 2010. This study utilizes fourteen years of data, starting after this last major change in playstyle.

Data Description

The data used for this project is box score data for every game played from the 2010-11 to 2023-24 seasons[5]. The 2019-2020 regular season games post-COVID-19 (games played in the bubble) are removed to counteract any difficulties the models may have in quantifying an advantage given to the home or away team, as the home team title was given based on that season's schedule but had no effect on game outcome given the neutral site. Each row contains the rolling average for all statistics over the last five games for both teams playing in the game and their opponents (so each statistic has four different iterations); assist-to-turnover ratio, effective field goal percentage, and offensive rebound percentage are just a few of the metrics derived from these box scores[6]. A variable comparing team wins is also used.

Table 1 presents the summary statistics for the basis variables used in the final model. The final model may not utilize all these variables and may introduce interaction effects between different variable combinations[7]. For example, an interaction between WinComp and the home team indicator may help capture the benefit of being the home team under different circumstances (heavily favored, underdogs, or even matchups). The target variable indicating a home team win has a mean of 0.57, which means that from the 11,191 observed games, the home team won about 57% of the time. The home team win rate, which is 57% in this dataset, is a common benchmark for model accuracy, as a 57% accurate model could be created by simply choosing the home team every game.

[5] NocturneBear (Vitalii Korolyk) · GitHub
[6] See appendix for details on some of the statistics derivations
[7] See appendix for more information on variable selection and correlation

Table 1: Summary Statistics of Basis Variables[8]

Variable	mean	std	min	max	mean in loss	mean in win
Coached Skills						
Assist-to-Turnover Ratio L5	1.73	0.36	0.74	3.68	1.69	1.77
Free Throws Made L5	17.35	3.12	8.2	34	17.31	17.40
Offensive Rebounds L5	10.26	2	4	18	10.32	10.21
Offensive Rebound Percent L5	0.23	0.04	0.11	0.37	0.23	0.23
Three Pointers Made L5	10.74	2.9	2.6	22	10.43	11.02
Natural Skills						
Field Goal Percentage L5	0.46	0.03	0.36	0.56	0.46	0.47
Three Point Percentage L5	0.36	0.04	0.18	0.54	0.35	0.36
Steals L5	7.65	1.52	3	15	7.58	7.71
Effective Field Goal L5	0.52	0.04	0.39	0.67	0.51	0.53
Other						
Home Games - Games Won L5	0.05	1.59	-5	5	-0.19	0.27
Points Allowed L5	107.36	7.97	77.8	133.6	107.41	107.32
Home Team Indicator	0.5	0.5	0	1	0.43	0.57
Plus Minus L5	0.31	7.78	-31.4	29.2	-1.21	1.73
Points Scored L5	107.81	8.4	77.2	135.8	106.88	108.69
Games Won L5	2.55	1.29	0	5	2.32	2.76
WinComp	0.01	0.13	-0.55	0.7	-0.04	0.06

In addition to comparing the top contributors SHAP values across models, this study utilizes SHAP plots to further identify what traits to look for in roster and coaching staff construction; this can be achieved by analyzing distributions of important features, whether they are enhanced through improvements in a team's natural player skills or coached skills as highlighted in **Table 1**. The variables classified as coached skills are variables that are primarily affected by changes in coaching strategy, while natural skills are skills developed by the player itself or related to specific physical traits of a player; of course, some players may naturally be better suited to improve specific coached skill variables. For example, a coach may implement a run-and-gun playstyle, think 2018 Houston Rockets, focusing more on three pointers made, offensive rebounds, and free throws. Steals are primarily a natural skill, seen in coordinated, fast, or long players, but may be increased with an aggressive defensive playstyle. The remaining shooting metrics in the natural skill category are assumed to be primarily affected by a team's

[8] L5 indicates the statistic being the rolling average over the last five games

natural scoring ability, although alterations in offensive schemes may have minor effects on these numbers.

The ‘mean in loss’ and ‘mean in win’ values for the rolling average variables (ones with an L5 indicator in **Table 1**) indicate that variable’s average value over the last five games for teams that won or lost. Comparing these values helps give a general idea of what variables translate most from previous performance to present performance in both wins and losses. These values indicate that winning teams over their last five games tend to have made more free throws and threes, had a lower assist-to-turnover ratio, more steals and points scored, and marginally better shooting percentages; losing teams saw more offensive rebounds over their last five games. Intuitively, it may be that losing teams are more likely to be weaker over their last five games, which may go hand in hand with lower shooting percentages, which leads to more offensive rebound opportunities. Variables that differ most across game outcomes are likely to be included in the final model.

Methodology

Random forest is a machine learning method that uses multiple decision trees to make predictions by combining their results, like asking multiple experts’ opinions on a game outcome, with the team receiving the majority vote being the predicted winner. Gradient boosting is another technique that improves prediction accuracy by adding trees one at a time, with each new tree learning from the mistakes made by the previous ones. Random forest trains trees independently helping the model apply better to new data even with outliers, while gradient boosting corrects errors of previous trees, leading to a higher accuracy, but potentially worse applications to new data.

Both the random forest and gradient boosting models saw the best average performance under the set of variables highlighted in **Table 2**. These variables are then entered into the gradient boosting and random forest models to achieve the resulting accuracy scores and SHAP values, the magnitude of each feature’s contribution. The five rolling average variables (indicated with an L5) are each seen twice in the final model, with a separate iteration for each team in the matchup.

Table 2: Best Average Accuracy Model's Final Features

Variable	Definition
WinComp and Home Interaction	Interaction term between WinComp and Home team indicator
WinComp	Compares season win totals
Effective Field Goal Percentage L5	Adjusted field goal percentage to account for three pointers' value
Steals L5	Steals per game over the last 5 games
Plus/Minus L5	Plus minus over the last 5 games
Assist-to-Turnover Ratio L5	Number of assists divided by turnovers in last 5 games
Offensive Rebound Percentage L5	Percent of offensive rebounds secured over the last 5 games

This final subset of variables uses only twelve features in total; this allows for consistency across different subsets of test data. Every variable seen in **Table 2** directly relates to different key aspects of deciding game outcomes. Effective field goal percentage and plus/minus capture a team's ability to score and defend, while assist-to-turnover ratio, offensive rebound percentage, and steals highlight a team's ability to maximize their possessions and minimize their opponent's. WinComp and the WinComp and home team interaction term give important details on team strength and how it relates to any home team advantages. The home team indicator itself was dropped as it did not add any additional accuracy that the interaction term did not capture. Correlated variables may introduce redundancy into the model; the basis box-score variables used to derive the final twelve variables highlighted in **Table 2** saw no concerningly high correlations [9].

Unnecessary features may provide some additional model accuracy but may only be useful in applications to specific games and sample sets. For example, a team's recent proclivity for grabbing offensive rebounds may have rightfully been the deciding factor for the model in one or two specific games, maybe between teams that have been shooting below average in recent games, hence the value in offensive rebounds. But in turn, the model may find patterns between game outcomes and offensive rebounding metrics and apply them to games where the outcomes may not have been affected by a team's recent dominance in offensive rebounding. This idea is known as overfitting and can lead to poor performance when using the model on test data[10].

[9] See Appendix for the correlation matrix
[10] See appendix for additional details about how parameters may inflate performance

Results

Table 3 presents the average metrics across five iterations for both model types. Precision indicates the percentage of accurate predictions for a certain group, whereas accuracy looks at the percentage of accurate predictions for the entire model, recall indicates how well a model captures all the relevant 'positives' for a category (a home team win/loss), and F1-score attempts to combine both recall and precision into a single metric. Precision, recall, F1-score, and accuracy are varying indicators of model performance and may be emphasized during the model-building process to help achieve specific goals. For example, a team that performs well at home but may be looking to improve their performance on the road may look to build a model with the highest precision or recall score for the home team lost category, with varying focus given to either metric depending on how the front offices value false negatives and false positives. A team focused on identifying all the features (even features important in correctly predicting only a few game outcomes) that contribute most to away team wins may attempt to minimize false negatives, which in this case are represented by the model failing to predict the away team winning when they in fact did win. This focus may result in a model utilizing features specific to away team wins but may have poor application to game predictions when home teams win.

Both the random forest and gradient boosting results seen in **Table 3** see much higher metrics when predicting games where the home team won. This study's goal is to correctly predict as many games as possible and to analyze what features aided most in this process. Intuitively, the discrepancy between the 'home team lost' and 'home team won' categories' metrics indicates the power of home court advantage but also reiterates the fact that home team wins are easier to predict, ultimately yielding features better equipped to predict home team wins. Again, differences in front office's goals may yield different sets of features and improved metrics for the 'home team lost' categories, despite decreases in overall model performance.

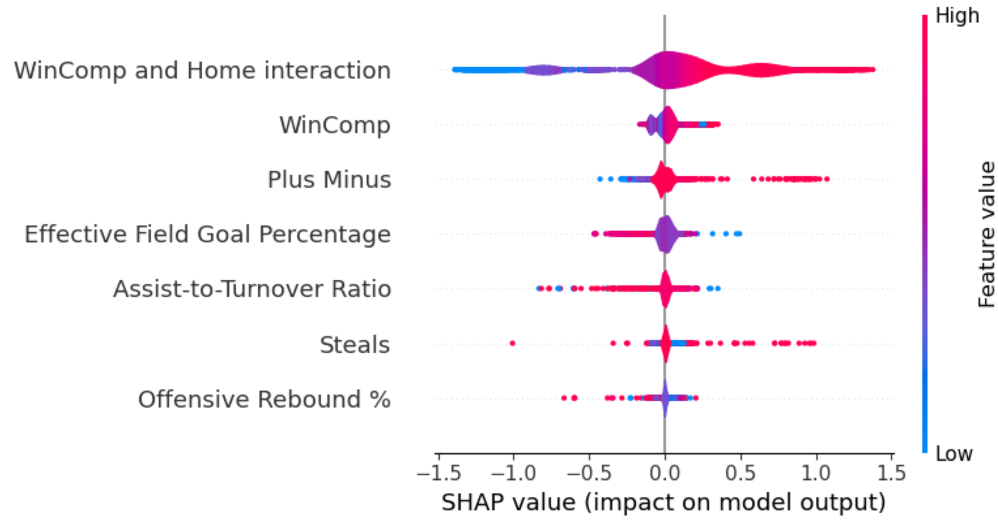
Table 3: Random Forest and Gradient Boosting Models Results

Avg Metrics	Random Forest: Home Team Lost	Random Forest: Home Team Won	Gradient Boosting: Home Team Lost	Gradient Boosting: Home Team Won
Avg Precision	0.64	0.73	0.67	0.72
Avg Recall	0.6	0.75	0.61	0.75
Avg F1-Score	0.63	0.74	0.64	0.74
Avg Support	960	1277	968	1271
Avg Accuracy	*	69.10%	*	70.50%

* Accuracy is only predicted for the target variable 'home team win'

Figure 1 highlights the distribution of SHAP values across the features, where the top distributions indicate the most important features. The distributions indicate the spread of values over the relevant metrics; for example, the interaction term clustering towards the center indicates that most games occur between two fairly even teams, indicated by the feature value coloring. Low and high values, indicated with blue and red, respectively, indicate lopsided contests or a large win differential. Note that those lopsided contests generally see larger magnitude SHAP values, indicating differences in team strength or wins are a large factor in game outcome prediction. The bee-swarm plot in **Figure 1** as well as the SHAP plots in **Figure 2**, show different ways to interpret SHAP distributions. The distribution plots in **Figure 2** highlight the level of effect the feature has, seen through the magnitude of SHAP values, for differing levels of the feature. SHAP values closer to zero for a specified range of that feature indicates the feature having minimal effect on game outcome.

Figure 1: Bee swarm Plot of Final Features SHAP Value Distributions



Model Comparison and Applications

SHAP values indicate the contribution to the model's decision-making process by each feature. The SHAP plots in **Figure 2 and 4**, showing the distributions for the top five contributors to the gradient boosting model, highlight these potentially optimal SHAP values. As seen in **Table 4**, the top unique contributors—the WinComp and home interaction term, plus/minus, assist-to-turnover ratio, effective field goal percentage, and steals— do happen to align with the random forest model's top contributors; this is telling of how consistently important these top contributors are in predicting game outcomes. The range of SHAP values across the final features gives some indication of the magnitudes of importance; WinComp and its interaction with the home team indicator saw the largest values, while offensive rebounding, steals, assist-to-turnover ratio, and effective field goal percentage saw marginal effects compared to the top contributors. This study focuses on the magnitude of SHAP values as opposed to their negative or positive values, as a road team's negative SHAP value will correspond to the home team winning probability decreasing and their probability of winning increasing (as a home team win is the target variable for prediction)[11].

[11] See appendix for information on how SHAP values may vary based on team classification

Table 4: Feature Ranking Comparison

Variable	Random Forest Ranking	Gradient Boosting Ranking	Range of SHAP Values
WinComp and Home Interaction	1	1	-1.5 to 1.5
WinComp	2	2	-1.5 to 1.5
Plus/Minus L5	3	3	-0.8 to 0.8
Effective Field Goal Percentage L5	4	4	-0.2 to 0.2
Assist-to-Turnover Ratio L5	5	5	-0.6 to 0.2
Steals L5	7	6	-0.2 to 0.2
Offensive Rebound Percentage L5	6	7	-0.2 to 0.2

The dependence plots also show how altering these metrics may affect the magnitude of SHAP values and therefore a team's winning chances. The assist-to-turnover rate is an example of a variable that can be altered by both changes in roster construction, by signing players with a higher assist-to-turnover ratio, and changes in coaching strategy. Its plot, seen in **Figure 2,** shows that teams with a higher assist-to-turnover ratio, say over 2.5, start to see gradual increases in its effect on game outcome. While teams with a lower ratio, say under 1.25, start to see a stark decline in its effect on game outcome.

Effective field goal percentage was deemed useful in game prediction, albeit not to the magnitude of home team indication and general team strength indicators; however, the metric is more straightforward in how it may be improved upon. Effective field goal percentages around 57% see a similar magnitude of effect on winning chances when compared to higher levels, minimal effect between 45% and 57%, and a lower magnitude of effect compared to the effect after the 57% optimal level for values below 45%. It may seem counterintuitive to identify an optimal value as increasing shooting percentages would intuitively seem to never yield any downsides. However, the value seen in highlighting these optimal levels is in identifying where diminishing returns exist in these attributes, allowing front offices to allocate cap space to improve metrics with a larger effect on game outcomes.

For example, in the 2024-25 season, Derrick White and Cameron Johnson had similar effective field goal percentages at 58.1% and 58% respectively. However, seen in **Figure 3**, Derrick White has a higher percentage of games under the lower bound threshold, indicating more detrimental shooting games despite their similar percentages. This may indicate that all else equal Cameron Johnson's effective field goal percentage is undervalued compared to Derrick White's, especially looking at their salaries of $20,543,478 and $28,100,000 respectively.

As seen in **Figure 2**, the optimal level of steals is seen to be between 8-10 per game over a team's last five games. A team tasked with improving their steals metric may look to sign athletic, long, or coordinated players, as steals are assumed to be more of a natural skill. However, an argument could be made that a coach's defensive scheme plays a role in the steals metric, which may give means to make upgrades in both roster and coaching staff. Like previous features, front offices can allocate cap space accordingly based on their performance compared to the optimal level of steals. A team lacking in steals but experiencing diminishing returns in another important feature like effective field goal percentage would see value in converting some of the cap space allocated to shooting into players with traits correlated with steals.

For example, in the 2024-25 season, the Suns rank second worst in terms of steals per game with only seven per game but rank fourth in effective field goal percentage at 56.4%. Bradley Beal has been an efficient scorer this year, but I would expect the Suns to see improvement if some of Beal's hefty $50 million salary was allocated towards a player or players that improved their steals per game and kept their effective field goal percentage around that optimal level, seeing that Beal only averages one steal per game. Scrappier guards like last offseason free agent Kris Dunn or Scottie Pippen Jr. (who make considerably less and would have to be packaged with a larger contract) would make immediate impacts in that department, averaging 1.7 and 1.3 steals respectively.

Figure 2: SHAP distributions for coached and natural skills

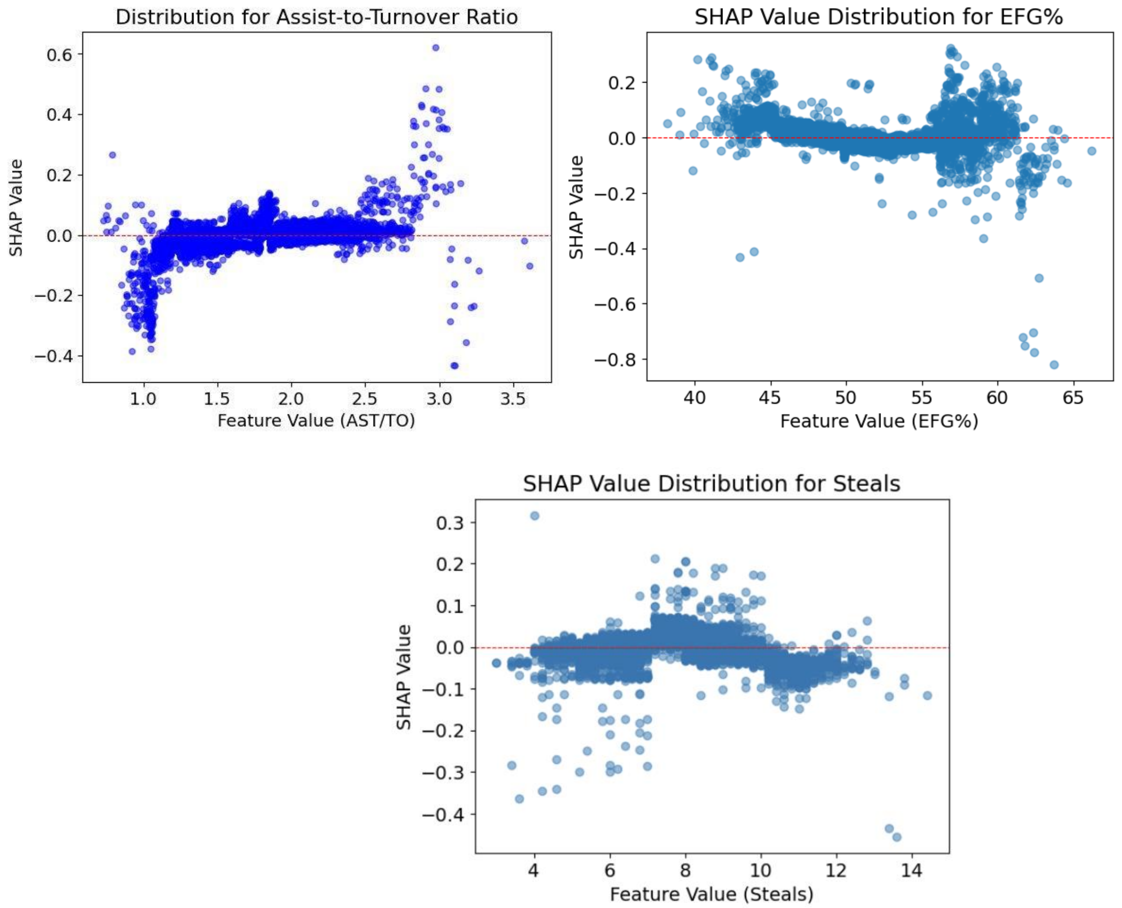
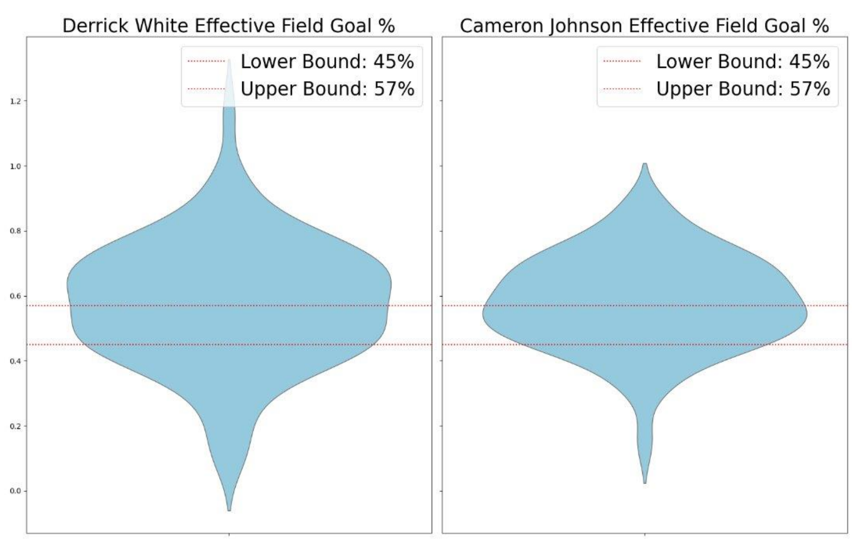
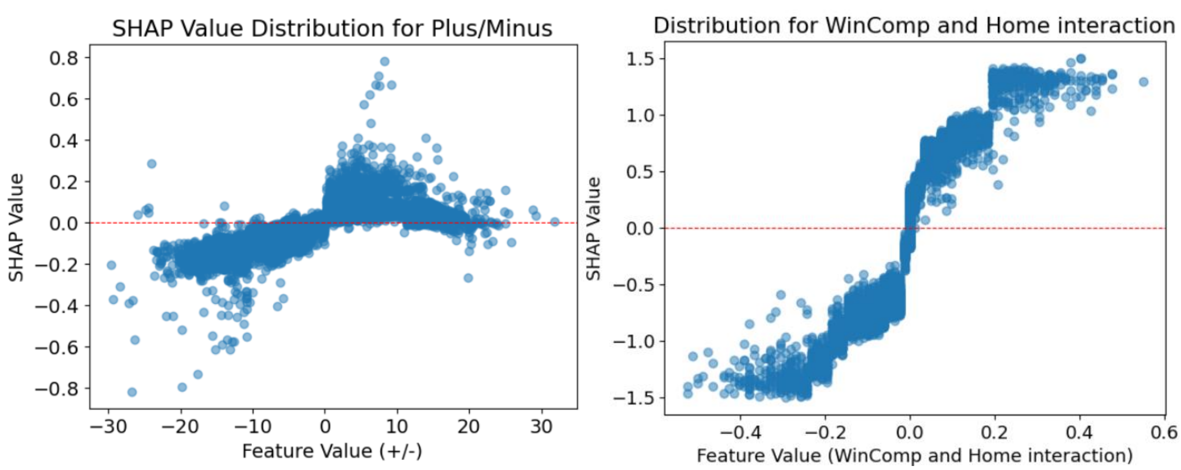


Figure 3: Effective Field Goal Distributions for Derrick White and Cameron Johnson



The WinComp and home team interaction plot and the plus/minus plot seen in **Figure 4** are straightforward in their interpretation, with an increase in the WinComp and home indicator interaction term (indicating a strong home team) corresponding with a higher SHAP value or contribution towards predicting game outcome. The same relationship is seen in the plus/minus feature; albeit not as strong, an increase in the previous five games' plus/minus corresponds with a larger contribution towards game prediction. The applicability of plus/minus in how it relates to game outcomes is questionable, as of course a team's goal is always to maximize winning margins, and a team with higher winning margins is seen as stronger. The lack of continued correlation between an increasing plus/minus and game outcomes may be indicative of momentum and confidence over the last five games, indicated by a high plus/minus, not translating to the outcome of their next game. These features were deemed important in model decision-making; however, it is difficult to link either to a specific coached or natural skillset that can be improved upon, as the features either relay information about general team strength or home team indication (which is out of a team's control).

Figure 4: SHAP Value Distribution Plots for Top Five Contributors of GBC



References

Junwen Wang. 2023. Predictive Analysis of NBA Game Outcomes through Machine Learning. In The 6th International Conference on Machine Learning and Machine Intelligence (MLMI2023), October 27–29, 2023, Chongqing, China. ACM,NewYork,NY,USA,10pages.https://doi.org/10.1145/3635638.3635646

Zhao, Kai, Chunjie Du, and Guangxin Tan. 2023. "Enhancing Basketball Game Outcome Prediction through Fused Graph Convolutional Networks and Random Forest Algorithm" *Entropy* 25, no. 5: 765. https://doi.org/10.3390/e25050765

NocturneBear (Vitalii Korolyk) · GitHub

Silver. 2022 “How We Calculate NBA Elo Ratings” | FiveThirtyEight

Appendix

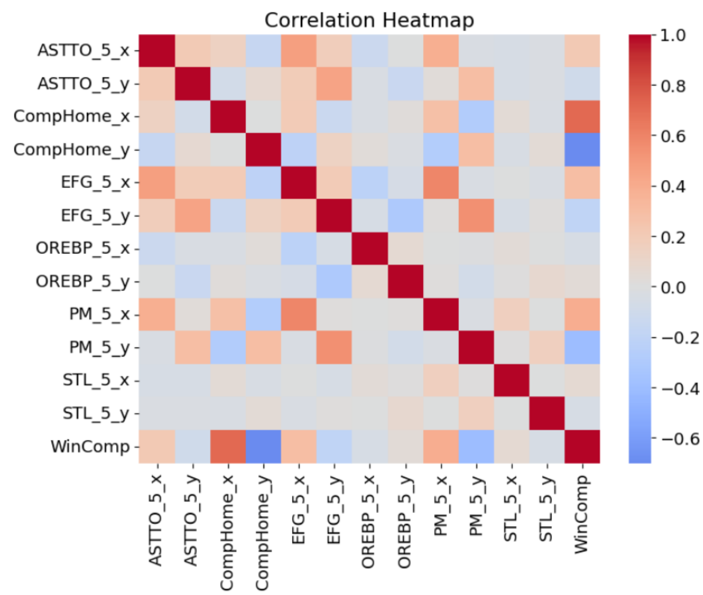
This appendix intends to highlight important technical details regarding the model, its variables, and performance.

4. A downside of using the previous five game performance metrics is the lack of data for games early in the season. Models that run across seasons have been implemented. The FiveThirtyEight model utilizes data on offseason transactions to quantify how previous season success will relate to current success. However, attempting to quantify the effect of offseason transactions may introduce unnecessary unreliability into the model. Due to the large sample size of 2,460 games per season across fourteen seasons and the goal of encapsulating the relationship between short-term performance and game outcomes. This paper only analyzes data after five games have elapsed at the start of the season. The start-of-season data is still used to help create baseline performance metrics.

6. Assist- to-turnover ratio, a metric used by Zhao (2023) and commonly seen across basketball data, is created by dividing the assist column by the turnover column. A rolling total of wins in the last five games is created by summing the win/loss column over the last five games. The offensive rebound percentage is created by dividing the total offensive rebounds by offensive rebounds plus the opponent's defensive rebounds. Metrics like effective field goal percentage (EFG%) and true shooting percentage (TS%) are created using a combination of free throws, field goals, and three-pointers made or attempted and their respective percentages.

7. Many of these box score statistics and statistics created from them may not be used in the final model after feature selection, especially due to their potentially high correlation with other columns that may be a better indicator of game outcome, but it is still important to test and see if they improve the final model's accuracy.

9. Correlation Matrix Corresponding to **Table 2's** Features, Indicating No Concerningly High Correlated Relationships



10. Tuning hyperparameters to tailor only one specific random state or test dataset can also lead to overfitting. In this paper's case, alterations to test size, which is the proportion of data used in testing the model, may see improvements in model accuracy. However, it is possible that the increase in accuracy is solely due to the newly introduced test data points being easily predicted, i.e., home team winners with a high win comparison metric, indicating a lopsided contest. Like feature selection, other hyperparameters utilized in the random forest or gradient-boosting models, like maximum depth or learning rate, may only see increases in accuracy under a specific test dataset. Adjustments in those hyperparameters may even reduce model accuracy under different conditions.

11. As discussed, each game has data from two teams. However, classifying the teams as home or away teams before the modeling process does not allow the model to leverage the insights extracted from potential increases in winning probability due to home team advantages. Hence, the need for no team-wide classifications before the modelling process and a home team variable indicating the home team. Classifying the points before modelling saw a drastic decrease in performance when compared to including the binary indication. However, this does cause issues in SHAP value interpretation as the target variable is a home team win; so, a negative SHAP value may belong to an away team,

decreasing the probability of a home team win but in turn increasing the probability of their own victory. For these reasons, this paper assumes negative SHAP values for highly efficient metrics and positive SHAP values for low- valued metrics belong to away teams.