Predictive Factors of Team Success in the WNBA: A Comparative Analysis with the NBA

Carsten Savage, Margaret Reinhard, Shusaku Matsuda, Zubair Lakhia

Predictive Factors of Team Success in the WNBA: A Comparative Analysis with the NBA

Carsten Savage, Margaret Reinhard, Shusaku Matsuda, Zubair Lakhia

Abstract

While the WNBA has become increasingly popular, its viewership remains far surpassed by the NBA. This study employs Principal Component Analysis to identify common predictors of team success across the WNBA and NBA and evaluates the effectiveness of gradient boosting trees, OLS linear regression, and random forest regressor on the common predictors. Random forest regressor outperformed gradient boosting trees due to the small sample size of highly aggregated play-by-play data, while OLS outperformed all models on the NBA data but underperformed on the WNBA data. Critical predictors across both sets of models include defensive rebounds, field goal percentage, and total turnovers. Utilizing common predictors in both the NBA and WNBA models enables more effective comparison of the models but decreases model performance for one or both sets of models.

1 Introduction — Motivation

In professional sports, understanding the factors that drive team success is crucial for improving performance, shaping strategies, and predicting outcomes. In the Women's National Basketball Association (WNBA), competition is fierce, but resources are often more limited compared to other leagues [1]. Success in the WNBA can be measured through various metrics, including those traditionally used in the National Basketball Association (NBA) such as "Inside vs Outside," "Offensive Aggressiveness," or "Efficiency vs Midrange" [18].

2 Problem Definition

Given team performance data from the WNBA and NBA, this study's objective is to build predictive models that accurately forecast team success in terms of wins in a season based on individual player metrics (e.g., assists) and team-level statistics (e.g., total assists). The models aim to identify the most influential factors contributing to a team's number of wins. Furthermore, the objective extends to comparing these factors between the two leagues to determine whether their relative importance differs, providing insights into the distinct dynamics of success in the WNBA versus the NBA[20]. In jargon-free terms, this work aims to understand the key factors that contribute to winning seasons in the WNBA and NBA and compare their impact in each league.

3 Literature Survey

Currently, team success in the NBA and WNBA is typically analyzed separately. No comprehensive study has compared the two leagues using modern machine learning techniques to identify common and unique predictors of team or player success.

Various basketball metrics have been used to develop models predicting team performance. For instance, Melnick [14] found a strong correlation between team assists and win-loss records in the NBA, highlighting that teamwork and ball distribution are more critical to success than individual scoring alone. Engelmann [5] compared various player impact metrics using play-by-play data to analyze individual contributions. While this work serves as a survey of metrics, useful for identifying star players or MVP candidates, it does not explore predictive modeling. Additionally, these analyses are specific to NBA data and do not explore potential differences in play style that may exist in the WNBA.

Some studies rely on older data that may not accurately reflect the current nature of the game. Buyukcelebi et al. (2024) [3] analyzed changes in NBA defensive strategies from 2008 to 2019 by examining key variables such as rebounds, steals, and blocks. They concluded that the importance of defensive rebounding in team success has increased over the years. Wang et al. (2022) [20] explored NBA data from 1980 to 2019, applying Principal Component Analysis (PCA) and machine learning models to analyze game-play trends and predict team success. Their results highlighted significant changes in playing styles, such as the shift from inside scoring dominance in the 1980s to the prominence of 3-point shooting in the 2010s. These suggest that any predictive models must be continually updated with the most recent data.

Existing literature employs various analytical techniques to generate models from the available data. Building on foundational methods, Ke et al. [13] and Berri [2] used linear regression models to predict players' overall ratings and minutes played per game, respectively, based on independent variables like field goals made and free-throw percentages. Duman [4] used hierarchical clustering to group players into clusters based on their playing styles for each of the five traditional positions: Point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), and center (C), with the goal of identifying player types that are most compatible together. Gong et al. [9] employed a hierarchical Bayesian linear model to estimate plus-minus points at the position level in the NBA and found that players with versatile offensive skills and larger players who defend the paint are the most valuable contributors on offense and defense, respectively. Martin-Gonzalez [11] examined time intervals between baskets in NBA games using a Poisson distribution model, finding that while most scoring events align with this model, the dynamics shift in the last minute of close games, where events follow a Power Law distribution. This study is limited to scoring events, suggesting an opportunity to extend the analysis to include non-scoring events. Similarly, Toma [19] investigated high-pressure situations using ordinary least squares (OLS) regression to assess free-throw performance in close games during the final minutes. These papers all have relevance to this work in that the techniques and models utilized can be applied to recent WNBA data to predict player and team success.

A common shortcoming in most of the literature is the lack of interactive visualizations, which is implemented in this study. Some literature provide interactive visualizations, but they are primarily NBA-centric [6][7].

4 Methodology

The analysis spans the 12 most recent WNBA seasons and 6 most recent NBA seasons, building on existing literature that often rely on older or single-season data. The numbers of seasons were selected to have sample sizes of similar scale for each league's dataset. The datasets include both team statistics and shot type data extracted from play-by-play information. By comparing NBA and WNBA success factors, this paper identifies differences in playing styles and predictors of success. The predictive models build upon those used in previous studies, incorporating them into new, interactive visualizations. This section provides a detailed outline of these methods.

4.1 Data Preparation and Warehouse Setup

NBA and WNBA data were sourced from the NBA and wehoop APIs, respectively [16][8]. To ensure efficient data management, a data warehouse in Google BigQuery was built with data build tool (dbt). The pipeline produces clean, organized data tables, which include:

- Averages per game for each team for each season.
- Total season wins and the champion for each season.
- Play-by-play data for both the NBA and WNBA, grouped by broad shot types (e.g., "paint shots" include hook shots and layups).

The appendix shows the lineage graph of the dbt pipeline (Fig. 6). It takes raw data from both league APIs and organizes it into clean, analysis-ready tables in Google BigQuery.

4.2 Exploratory Data Analysis (EDA) and Variable Selection

EDA was conducted to identify key relationships between potential model variables. Correlation matrices were generated for all pairs of variables to identify any strong linear relationships, e.g., average points scored per game and average shots made. Mutual information scores for each variable against number of wins in a season were computed to check which variables showed a dependence with season wins. This metric helped determine which of the correlated variables in a pair to drop. Together, these measures helped to reduce multicollinearity in the models.

As a final step, principal component analysis was applied to further reduce dimensionality while retaining the most relevant features, similar to the approach in [20]. Cumulative variance explained by each principal component was plotted, and the first seven principal components were selected (Figs. 7; 8). Variable loadings were computed for each principal component, and variables with high absolute loading values in the first seven principal components across both the NBA and WNBA datasets were selected as the final independent variables (Figs. 9; 10). These variables contributed the

most to the first seven principal components, which explain the majority of the variance in the data.

4.3 Visualization Design and Development

The following visualizations make this study's findings accessible and engaging:

- **Interactive What-If Predictor:** Enables users to adjust team metrics such as assist rates or turnovers to see how these changes affect projected season wins, helping to understand the impact of individual statistics on team success (Fig. 1).
- **Radar Chart:** Visualizes differences between high-win and low-win teams across various metrics (Fig. 2).
- **Shot Coordinate Charts:** Compare shot patterns between the NBA and WNBA using play-by-play data (Fig. 3).
- **Shot Type Distribution Charts:** Allow users to effectively compare statistics of the top five most winning teams in the NBA and WNBA (Fig. 4).

The visualizations are contained in a single Dash application that is powered by Google Cloud Platform's (GCP) App Engine and able to support up to 8 users at a time.

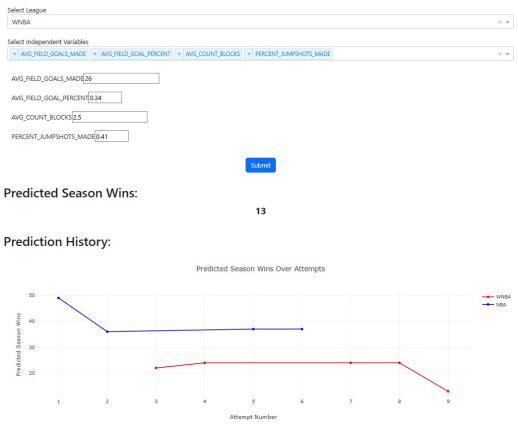

Figure 1: Interactive LightGBM-based What-If Predictor of season wins based on user-selected metrics and values

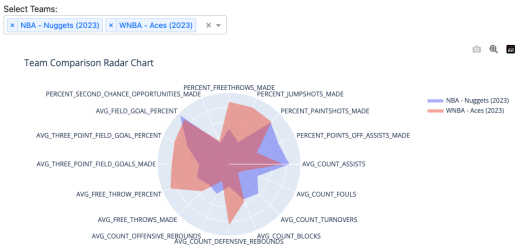

Figure 2: Radar chart to compare multiple teams across the NBA and WNBA

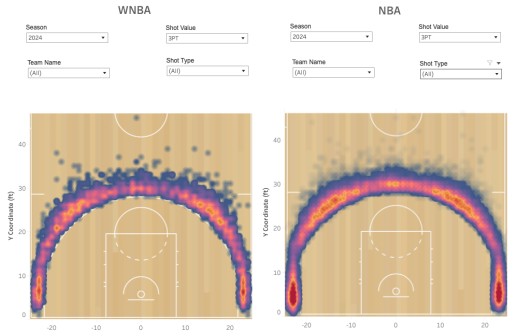

Figure 3: Shot chart of different shot types using shot coordinates

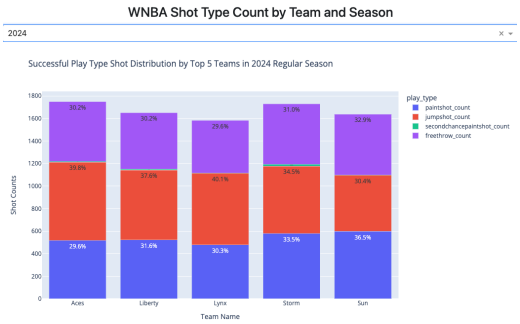

Figure 4: Shot Type Chart - WNBA

4.4 Model Development and Deployment

Multiple predictive models enable effective evaluation of team success in the NBA and WNBA. These include:

- **Linear Regression Model:** This OLS model can be represented by the following equation, in which β_0 is the intercept and μ is the unobservable:

$$
\begin{aligned}
\text{SEASON\_WINS} = \beta_0 & \\
& + \beta_1 \cdot \text{AVG\_ASSISTS} \\
& + \beta_2 \cdot \text{AVG\_DEF\_REBOUNDS} \\
& + \beta_3 \cdot \text{AVG\_FIELD\_GOAL\_PCT} \\
& + \beta_4 \cdot \text{AVG\_FOULS} \\
& + \beta_5 \cdot \text{AVG\_TOT\_TURNOVERS} \\
& + \beta_6 \cdot \text{PER\_JUMPSHOTS} \\
& + \beta_7 \cdot \text{PER\_PTS\_OFF\_ASSISTS} \\
& + \beta_8 \cdot \text{PER\_SECONDCHANCE} \\
& + \mu
\end{aligned}
$$

The random forest and LightGBM models employ the same independent variables as the OLS model. See appendix Fig. 25 for variable definitions.

- **Random Forest Regressor:** A random forest regressor model employing 1,000 decision trees with a maximum depth of 15.
- **LightGBM:** Deployed via Google Cloud Functions, this model offers fast and scalable predictions for season wins. The model employs gradient boosting decision trees with 20 leaves, a learning rate of 0.1, a maximum depth of 3, and up to 1,000 boosting rounds. While previous studies, such as Buyukcelebi et al. [3], have used Extreme Gradient Boosting (XGBoost), LightGBM is a more memory-efficient and faster alternative [12][21]. The Cloud Function takes POST requests with the dataset name, the names of independent variables for training and testing, and the what-if dictionary as the payload. It is invoked by the What-If Predictor in the Dash application. This interactive application allows users to adjust team metrics and immediately view the impact of these adjustments on projected season wins (Fig. 1).

These models offer valuable insights for fans by illuminating the effects of predictors on season wins. Cloud-based deployment ensures predictions are both scalable and fast.

4.5 Innovations

This work introduces several key innovations that expand upon existing research:

(1) **Comparative League Analysis:** Unlike previous studies that focus on single leagues or specific seasons, this analysis directly compares the NBA and WNBA, revealing both similarities and differences in playing styles and success predictors. Additionally, by grouping shot types (e.g., paint shots, three-pointers) from play-by-play data, this study offers a detailed analysis that enhances understanding, especially for women's basketball, where such metrics are less commonly explored.
(2) **Interactive What-If Season Wins Predictor:** This component of the Dash application empowers users to experiment with team metrics, such as assist rates or turnover rates, to observe projected impacts on regular season wins (Fig. 1). The application also leverages LightGBM which improves speed and memory efficiency compared to XGBoost models used in prior literature [3].
(3) **Radar Chart:** This chart allows users to effectively compare WNBA and NBA team performance by season. Previous literature have not shown WNBA and NBA data on the same overlapping radar chart (Fig. 2).
(4) **Shot Type Distribution Charts:** These charts allow users to effectively compare statistics of season winners in the NBA and WNBA by calculating and displaying the proportion of each shot type as a percentage (Fig. 4).
(5) **Cloud-Based Deployment:** The data and models are deployed on Google Cloud to ensure scalability and fast performance, allowing users to interact with large datasets seamlessly.

These innovations provide new, accessible insights into success factors in the NBA and WNBA.

5 Experiments and Evaluation

Experiments were conducted to assess the scalability, accuracy, and usability of the models and visualizations.

5.1 Scalability Evaluation

The models were tested to ensure that calculation speed remains efficient with additional data. The wall clock times for building the data pipeline and for response times in the What-if Season Wins Predictor were measured, comparing performance between a dataset with 6 WNBA seasons and an expanded dataset with 12 WNBA seasons.

The results showed a less-than-linear increase in total runtime, going from roughly 6 minutes to just over 9

minutes (a 58% increase) when the WNBA dataset was doubled to 12 seasons. Loading data into Google Cloud Storage is the longest step at approximately 60% of runtime. Retrieving data from the API is the most limiting step, as it is the only step that more than doubles in time with a doubling of the dataset (Fig. 13).

5.2 Prediction Accuracy

The models ultimately aimed to predict the number of regular season wins within a margin of ±30% in terms of relative error. Relative error is defined as mean absolute error (MAE) divided by the mean of the observed number of regular season wins. Accuracy metrics were compared across LightGBM, linear regression, and random forest models.

Table 1 summarizes the performance metrics for each model by league across all seasons.

Model	League	MSE	MAE	R^2
LightGBM	NBA	86.39	7.55	0.39
Linear Regression	NBA	58.36	6.27	0.49
Random Forest	NBA	73.61	7.03	0.35
LightGBM	WNBA	31.65	4.34	0.55
Linear Regression	WNBA	35.52	4.89	0.19
Random Forest	WNBA	28.04	4.27	0.36

Table 1: Comparison of model performance metrics

The NBA OLS regression model outperformed the LightGBM and random forest regressor models with the lowest MAE, lowest mean squared error (MSE), and the highest R^2 of the models. The OLS model's predicted season wins for the NBA was on average 6.27 wins different from the observed season wins value, and the model was able to explain 49% of the variance in season wins. Meanwhile, LightGBM and random forest regressor performed similarly on the WNBA data.

The NBA random forest's predictions were approximately 0.5 wins closer to the observed wins values on average than LightGBM's (Table 1). Random forest models are able to successfully handle challenges arising from small sample sizes [10]. Hyperparameter tuning may not improve model performance with a smaller dataset [10], and the performance of LightGBM and other more cutting-edge ensemble methods typically depends on hyperparameter tuning [15].

For both the NBA and WNBA LightGBM models, the variables AVG_FIELD_GOAL_PCT, AVG_TOT_TURNOVERS, and AVG_DEF_REBOUNDS were found to be the most significant by a large margin (Fig. 11; 12). The WNBA model relied much more heavily on PER_SECONDCHANCE than the NBA model (Fig. 12). This may indicate the WNBA's emphasis on ball movement and defensive discipline, contrasting with the NBA, which showed a stronger reliance on offensive efficiency metrics. These findings underscore the different play styles and strategies that drive success in each league.

As demonstrated by the NBA LightGBM model's feature importance chart (Fig. 11) and first and last tree diagrams (Figs. 20; 22), variables such as AVG_FIELD_GOAL_PCT and AVG_DEF_REBOUNDS are critical for accurately predicting season wins. While the WNBA model exemplifies a broader range of split values across its features, both models have diverse feature utilization (Figs. 14; 20; 22).

The WNBA LightGBM model's MAE and MSE are substantially lower than the NBA model's. The NBA model initially had a more challenging learning process, and it was able to minimize MAE to 7.55 season wins (Fig. 16) compared to the WNBA model's 4.34 wins (Fig. 17). An important consideration, however, is that average season wins in the underlying NBA data is higher than in the WNBA data (Fig. 24). Relative error demonstrates the size of the prediction's error relative to the underlying data. With mean NBA season wins as 38.72 and mean WNBA season wins as 18.52 across seasons, relative error percentages for the NBA and WNBA are 19.49% and 23.43%, respectively.

5.3 Usability Evaluation

The usability of the visualizations was assessed through a survey of 30 users (Fig. 26). Success was defined as at least 50% of survey participants reporting the following:

(1) The project is easy to navigate.
(2) The visualizations provide meaningful insights into the differences between the WNBA and NBA, highlighting the unique aspects of the WNBA.

87% of users found the application easy to navigate, and 93% highlighted that the visualizations provided meaningful insights into WNBA and NBA differences (Fig. 5).

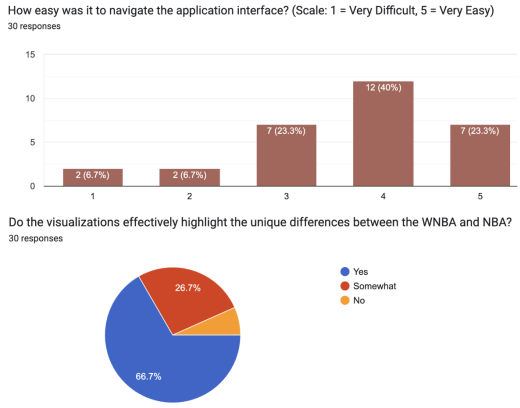

Figure 5: User feedback on application usability and insights

Additionally, users were asked to share what they learned and provide general feedback. Below are some examples of interesting insights derived from the visualizations:

(1) The radar chart highlights that the 2023 WNBA and NBA champions exhibited contrasting play styles. The WNBA champion Sparks had minimal turnovers and fouls, while the NBA champion Nuggets excelled in field goal percentage. Both teams shared high assist numbers, however, emphasizing the universal importance of generating efficient shots through teamwork.
(2) The stacked bar charts compare the top 5 NBA and WNBA teams with the most wins each season, breaking down their made shots into four play types: Paint shots, jump shots, second-chance paint shots, and free throws. Overall, the top 5 WNBA teams rely less on free throws and jump shots than their NBA counterparts, which could appeal to viewers who prefer a more physical style of play focused on scoring in the paint.
(3) Based on the shot chart, both leagues show a higher proportion of three point shots made from the corners near the end of the court versus other locations along the three-point arc. The proportion of corner threes is significantly higher in the NBA compared to the WNBA. This difference could possibly be attributed to differences in court lines and distances. Corner threes are nearly 2 ft closer to the basket in the NBA as compared to other shots along the three-point arc while the difference in the WNBA is only about 0.5 ft.
(4) Additionally, based on the shot chart, WNBA players take more mid-range two-point shots between the paint and three-point arc while NBA two-point shots tend to be closer to the rim. This highlights an overall shift in the NBA of focusing field goal attempts on the three-point line and shots near the basket which often have a higher expected value than mid-range two-point shots when field goal percentage is factored in.
(5) Using the What-If Scenario Tool: By inputting the 2024 Boston Celtics' team statistics—44 field goals made per game, an average field goal percentage of 49%, 6.5 blocks per game, and 46% of made shots coming from jump shots—the predictor estimates 46 season wins. However, if the average field goal percentage is reduced to the Grizzlies' 2024 regular season value of 43%, the prediction drops to 35 season wins.

5.4 Opportunities for Improvement

While the application was highly effective, user feedback identified several areas where further enhancements could improve clarity and functionality.

(1) When using the What-If Predictor, one user inputted 0 average field goals, but the tool predicted 30 season wins. This confused the user. There are no observations with 0 field goals made or 0 season wins in the data, so LightGBM cannot infer season wins based on this input data. The most logical remedy would be to implement logic returning 0 season wins if inference data is unrealistically low.
(2) Users suggested that the Shot Chart include the percentage of shots made out of shots taken.
(3) The radar chart was noted to have too many variables, making it difficult to interpret. Users suggested adding the ability to select or deselect variables to focus on fewer variables.
(4) Users requested that the What-If Predictor save input values between interactions to make it easier to experiment with multiple scenarios.

6 Conclusion and Discussion

Common predictors of team success were identified and compared across aggregated NBA and WNBA data. Mutual info scores and Principal Component Analysis were integral in identifying the finalized predictors.

Gradient boosting trees, OLS linear regression, and random forest regressor were employed to predict season wins. LightGBM underperformed compared to the random forest regressor due to limited sample size. OLS outperformed the competing models on the NBA data but underperformed on the WNBA data. Across both sets of models, defensive rebounds, field goal percentage, and total turnovers consistently ranked as critical predictors. While employing common predictors for both the NBA and WNBA facilitates a more effective comparison of the models, model performance consequently decreases for one or both sets of models.

Future work could focus on collecting data from additional seasons. With the rise of advanced data like WNBA 3D-tracking [17], future studies could leverage these data to analyze player movement, spatial strategies, and team dynamics. This technology offers the potential to enhance models, improving predictive accuracy and provide further game insights.

All members contributed a similar amount of effort.

References

[1] Nola Agha and David Berri. 2023. Demand for Basketball: A Comparison of the WNBA and NBA. *International Journal of Sport Finance* 18, 10 (2023), 35–44.

[2] David J Berri and Anthony C Krautmann. 2013. Understanding the WNBA on and off the court. In *Handbook on the economics of women in sports*. Edward Elgar Publishing, 132–155.

[3] Hakan Buyukcelebi, Fatma Nese Sahin, Mahmut Acak, Hüseyin Şahin Uysal, Cengizhan Sari, Dilara Erkan, Semra Yatak, and Raci Karayigit. 2024. Changes in Defensive Variables Determining Success in the NBA over the Last 10 Years. *Applied Sciences* 14, 15 (2024), 6696. https://doi.org/10.3390/app14156696

[4] Eyüp Anıl Duman, Bahar Sennaroğlu, and Gülfem Tuzkaya. 2024. A cluster analysis of basketball players for each of the five traditionally defined positions. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology* 238, 1 (2024), 55–75. https://doi.org/10.1177/17543371211062064

[5] Jeremias Engelmann. 2016. Possession-Based Player Performance Analysis in Basketball (Adjusted +/− and Related Concepts). In *Handbook of Statistical Methods and Analyses in Sports* (1st ed.), Jim Albert, Mark E. Glickman, Tim Hoshmandpour, and Tim Swartz (Eds.). Chapman and Hall/CRC, 14.

[6] Yu Fu and John Stasko. 2022. Supporting Data-Driven Basketball Journalism through Interactive Visualization. In *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3491102.3502078

[7] Yu Fu and John Stasko. 2023. HoopInSight: Analyzing and Comparing Basketball Shooting Performance Through Visualization. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2023), 858–868. https://doi.org/10.1109/TVCG.2023.3326910

[8] Saiem Gilani and Geoff Hutchinson. 2024. *wehoop: The Sports-Dataverse's R Package for Women's Basketball Data.* https://doi.org/10.32614/cran.package.wehoop R package version 2.1.0.

[9] Hua Gong and Su Chen. 2024. Estimating positional plus-minus in the NBA. *Journal of Quantitative Analysis in Sports* 20, 3 (2024), 193–217. https://doi.org/10.1515/jqas-2022-0120

[10] Sunwoo Han, Brian D. Williamson, and Youyi Fong. 2021. Improving random forest predictions in small datasets from two-phase sampling designs. *BMC Medical Informatics and Decision Making* 21, 1 (2021), 322. https://doi.org/10.1186/s12911-021-01688-3

[11] Juan Manuel García-Manso Enrique Arriaza Juan Manuel Martín-González, Yves de Saá Guerra and Teresa Valverde-Estévez. 2016. The Poisson model limits in NBA basketball: Complexity in team sports. *Physica A: Statistical Mechanics and its Applications* 464 (2016), 182–190. https://doi.org/10.1016/j.physa.2016.07.028

[12] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*. Curran Associates Inc., Long Beach, CA, USA, 3146–3154. https://github.com/Microsoft/LightGBM

[13] Yuhao Ke, Ranran Bian, and Rohitash Chandra. 2024. A unified machine learning framework for basketball team roster construction: NBA and WNBA. *Applied Soft Computing* 153 (2024), 111298.

[14] Merrill J. Melnick. 2001. Relationship between Team Assists and Win-Loss Record in the National Basketball Association. *Perceptual and Motor Skills* 92, 2 (2001), 595–602. https://doi.org/10.2466/pms.2001.92.2.595

[15] K.-T. Nguyen, T.-N. Tran, and H.-T. Nguyen. 2024. Research on the Influence of Hyperparameters on the LightGBM Model in Load Forecasting. *Engineering, Technology & Applied Science Research* 14, 5 (Oct. 2024), 17005–17010. https://doi.org/10.48084/etasr.8266

[16] Swar Patel. 2024. *nba_api.* https://pypi.org/project/nba_api/

[17] Associated Press. 2024. *WNBA brings new 3D tracking technology to enhance game analysis.* https://apnews.com/article/wnba-3d-tracking-c7c7c51c94fde0f4ed6cb2d12eef84e8

[18] Chris A Richardson. 2019. Evolution of a Player: Transitions in NBA Player Classifications. *Social Science Research Network* 3515711 (December 2019). https://doi.org/10.2139/ssrn.3515711

[19] Mattie Toma. 2015. Missed Shots at the Free-Throw Line: Analyzing the Determinants of Choking Under Pressure. *Journal of Sports Economics* 18, 6 (2015). https://doi.org/10.1177/1527002515593779

[20] Yuanchen Wang, Weibo Liu, and Xiaohui Liu. 2022. Explainable AI techniques with application to NBA gameplay prediction. *Neurocomputing* 483 (2022), 59–71. https://doi.org/10.1016/j.neucom.2022.01.098

[21] Dongyang Zhang and Yicheng Gong. 2020. The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure. *IEEE Access* 8 (2020), 220040–220049. https://doi.org/10.1109/ACCESS.2020.3040823

A Appendix

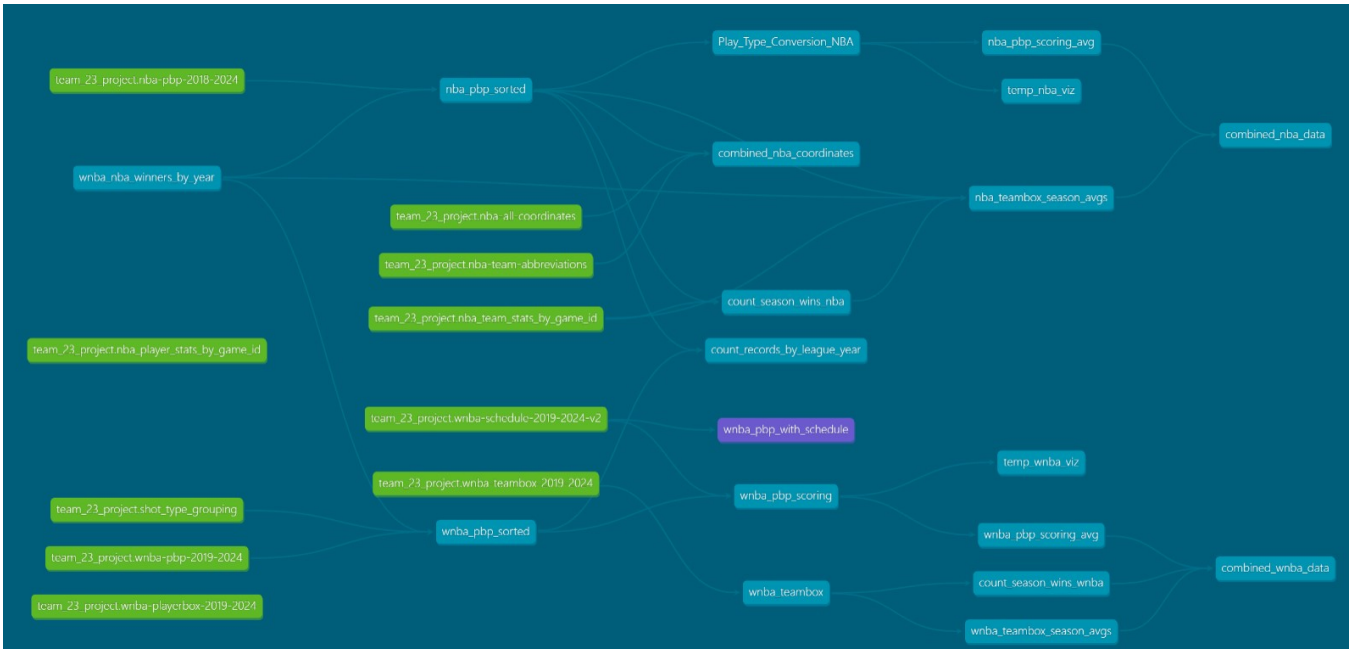

Figure 6: dbt Lineage Graph. The datasets employed by the models are the rightmost combined_nba_data and combined_wnba_data.

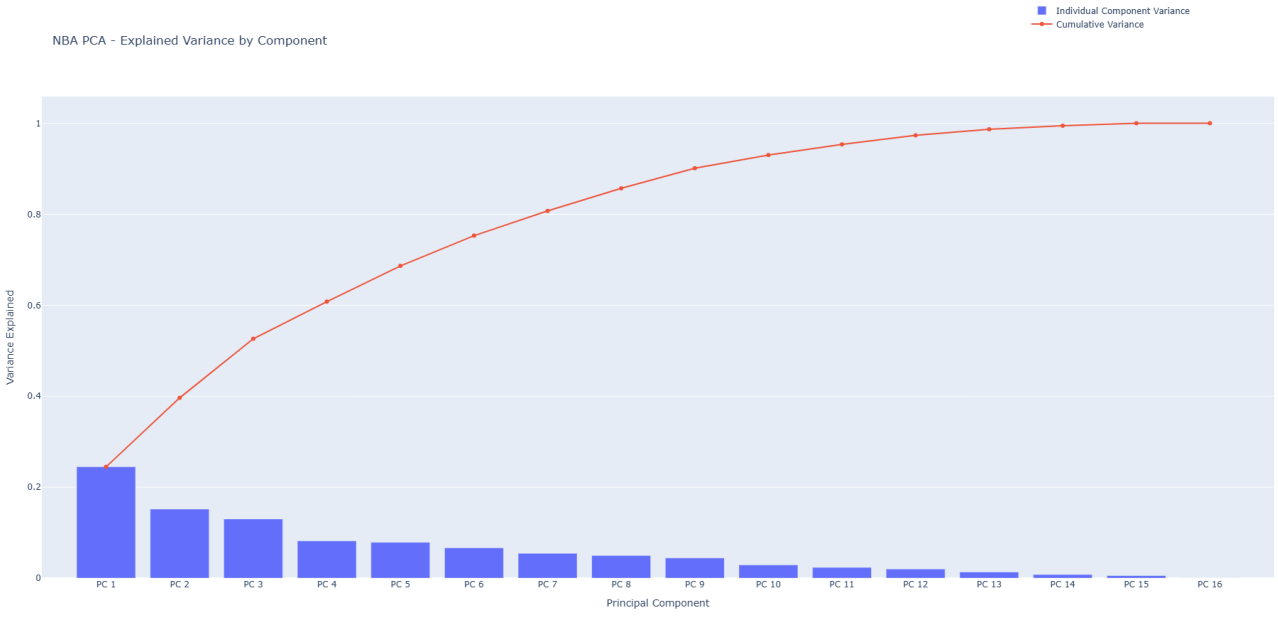

Figure 7: Explained Variance by Component - NBA

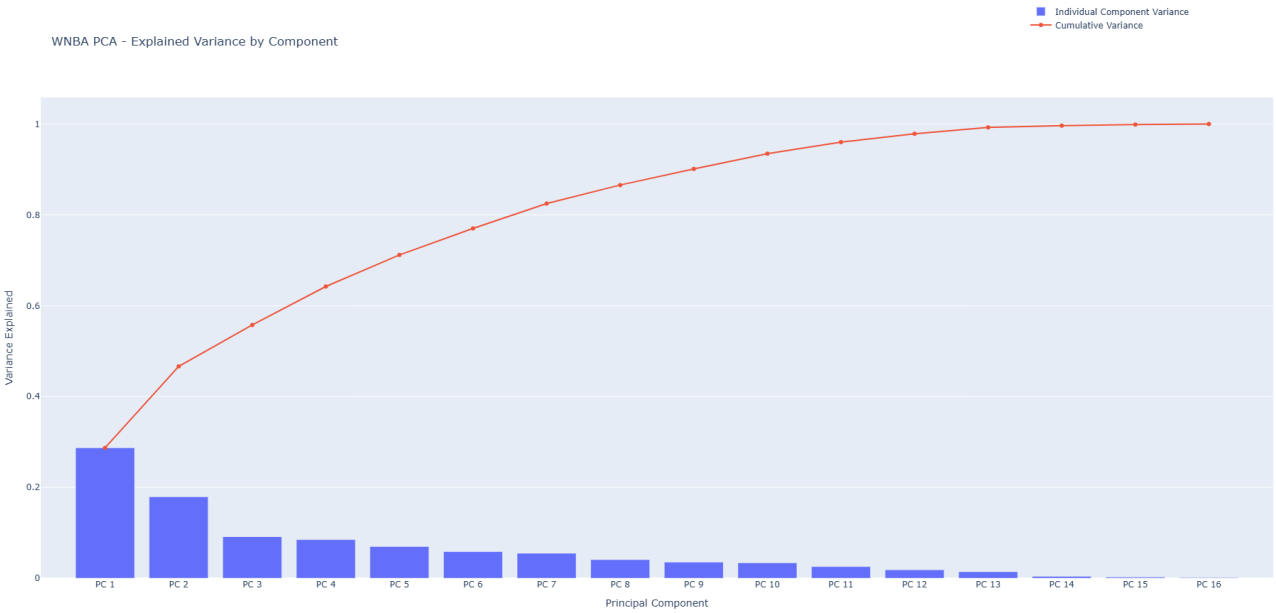

Figure 8: Explained Variance by Component - WNBA

		Principal Component No.						
League	Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
NBA	AVG_ASSISTS	0.204	0.276	0.448	-0.146	0.073	-0.143	-0.247
NBA	AVG_BLOCKS	0.065	-0.098	0.319	-0.037	0.480	0.325	0.119
NBA	AVG_DEF_REBOUNDS	0.211	-0.245	-0.049	-0.383	0.449	-0.208	0.311
NBA	AVG_FIELD_GOAL_PCT	0.310	0.044	0.381	-0.220	-0.253	0.079	-0.062
NBA	AVG_FOULS	-0.303	-0.310	0.235	-0.050	-0.064	-0.183	-0.126
NBA	AVG_PLUS_MINUS_POINTS	0.368	-0.301	0.221	0.078	0.059	0.068	0.028
NBA	AVG_PTS_AGAINST	-0.218	0.334	0.126	-0.360	-0.207	-0.082	-0.089
NBA	AVG_STEALS	-0.095	-0.050	0.343	0.560	0.292	-0.074	-0.136
NBA	AVG_THREE_PT_FIELD_GOAL_PCT	0.361	-0.111	0.159	-0.101	-0.306	0.011	-0.120
NBA	AVG_TOT_TURNOVERS	-0.258	0.042	0.164	-0.456	0.237	-0.260	0.070
NBA	per_freethrows	-0.285	-0.393	0.228	-0.021	-0.152	-0.010	-0.258
NBA	per_jumpshots	0.382	0.003	-0.348	-0.062	0.097	-0.122	-0.152
NBA	per_paintshots	-0.173	0.417	0.201	0.122	0.036	0.096	0.484
NBA	per_pts_off_assists	0.227	0.443	0.078	0.169	0.150	-0.111	-0.220
NBA	per_secondchance	-0.079	0.028	-0.034	-0.236	0.059	0.818	-0.151
NBA	VAR_FIELD_GOAL_PCT	0.136	-0.109	0.200	0.094	-0.391	0.007	0.605

Figure 9: Loadings by Principal Component - NBA

		Principal Component No.						
League	Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
WNBA	AVG_ASSISTS	-0.417	0.204	0.001	0.233	-0.020	0.025	-0.011
WNBA	AVG_BLOCKS	-0.157	-0.110	-0.292	0.092	0.401	0.017	0.608
WNBA	AVG_DEF_REBOUNDS	-0.258	0.031	-0.429	0.192	-0.114	-0.043	-0.103
WNBA	AVG_FIELD_GOAL_PCT	-0.359	0.233	0.078	0.186	-0.375	0.004	0.137
WNBA	AVG_FOULS	0.381	-0.023	0.169	0.187	0.084	0.048	0.119
WNBA	AVG_PTS_AGAINST	0.243	-0.070	-0.003	0.515	-0.215	-0.145	0.243
WNBA	AVG_STEALS	-0.105	0.226	0.281	-0.553	-0.031	0.130	0.154
WNBA	AVG_THREE_PT_FIELD_GOAL_PCT	-0.339	-0.055	0.194	0.178	-0.300	0.107	-0.263
WNBA	AVG_TOT_TURNOVERS	0.228	0.095	0.150	0.409	0.235	0.391	-0.301
WNBA	GAME_COUNT	-0.232	0.134	-0.240	0.015	0.373	-0.264	-0.087
WNBA	per_freethrows	0.176	-0.032	-0.562	-0.122	-0.168	0.094	-0.243
WNBA	per_jumpshots	-0.187	-0.488	0.354	0.034	0.053	-0.211	0.040
WNBA	per_paintshots	0.016	0.550	0.081	0.103	0.050	0.328	0.320
WNBA	per_pts_off_assists	-0.274	-0.077	0.125	0.102	0.537	0.205	-0.287
WNBA	per_secondchance	0.160	0.438	0.007	-0.056	0.121	-0.309	-0.301
WNBA	VAR_FIELD_GOAL_PCT	-0.054	-0.255	-0.191	-0.140	-0.098	0.652	0.014

Figure 10: Loadings by Principal Component - WNBA

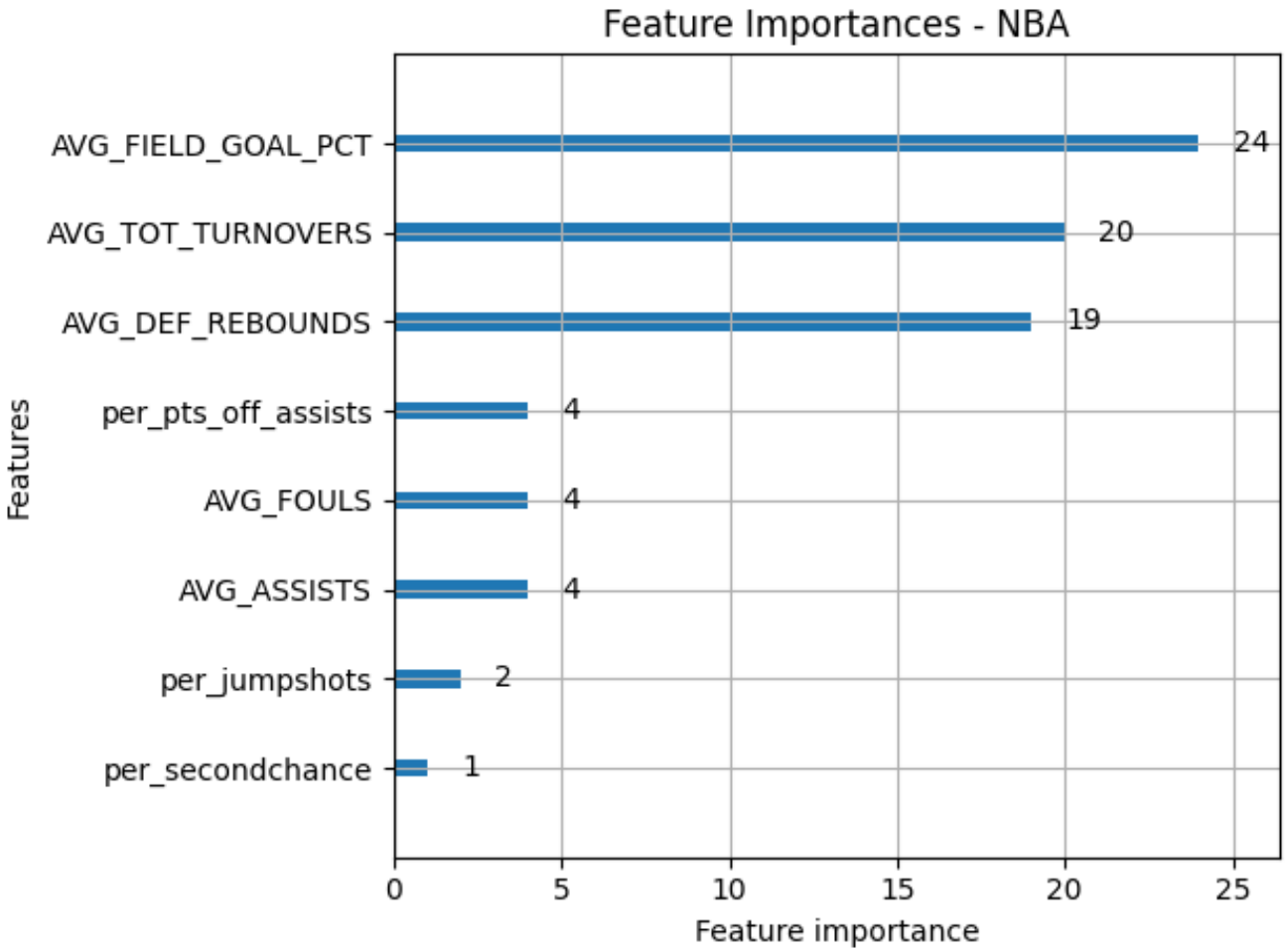

Figure 11: Feature Importances - NBA

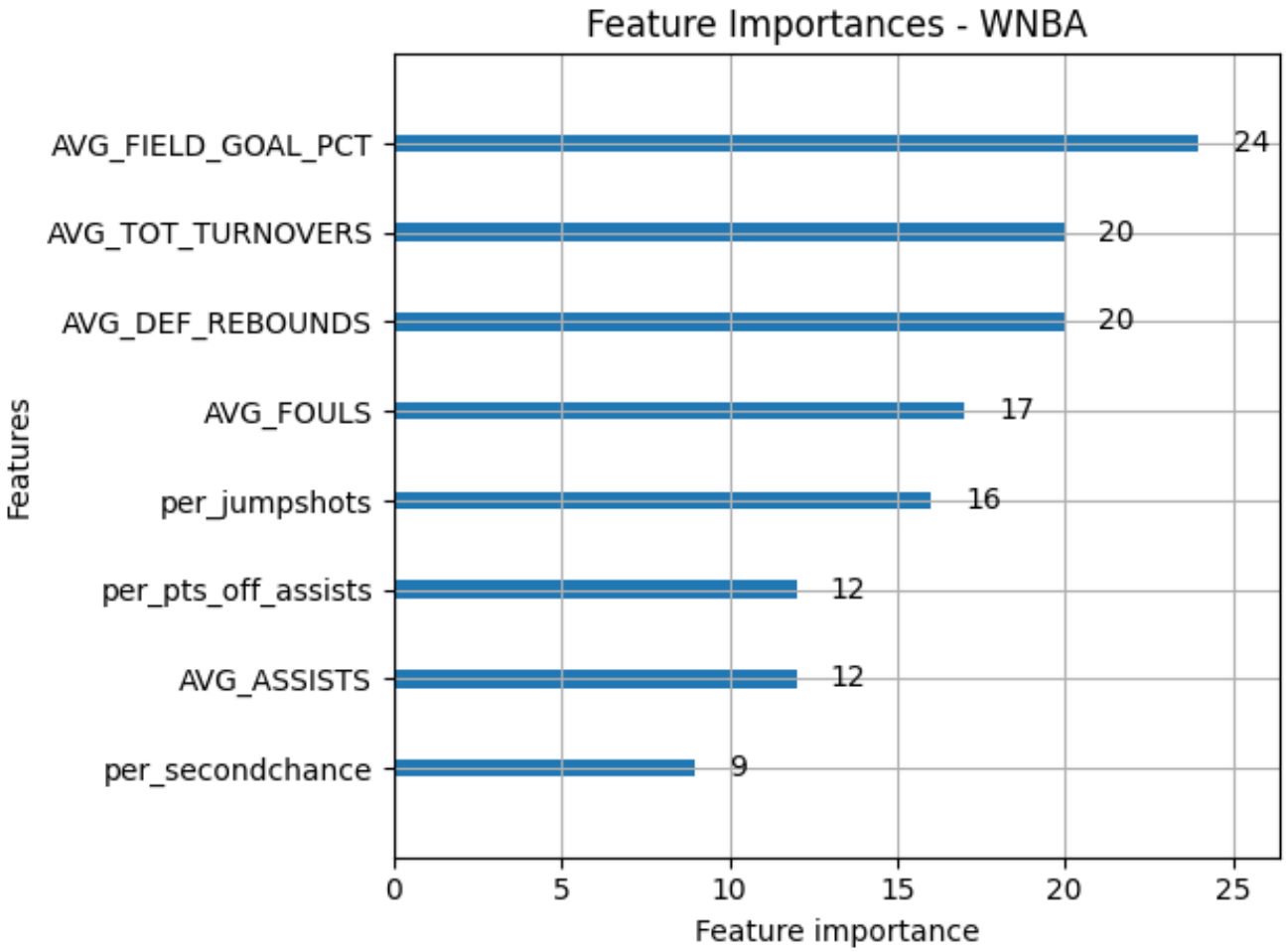

Figure 12: Feature Importances - WNBA

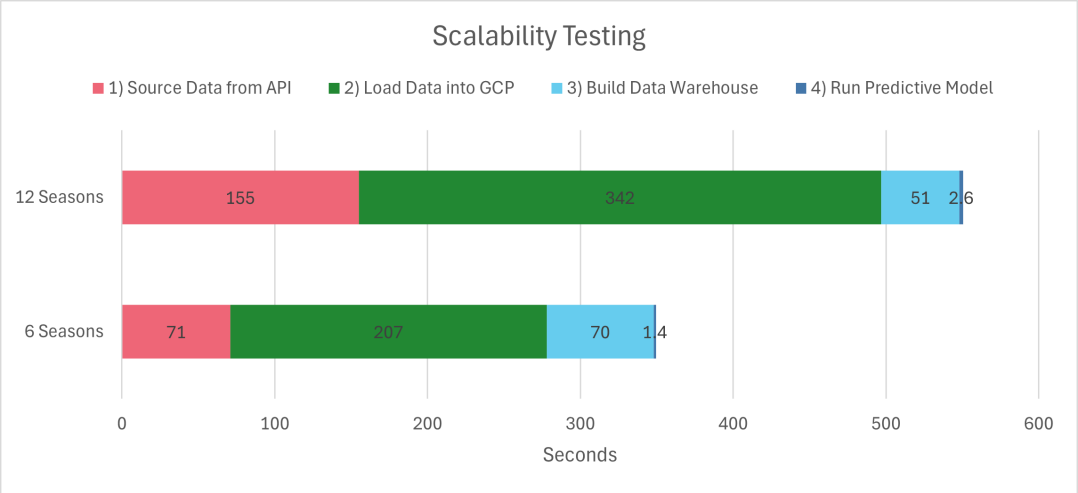

Figure 13: Scalability Test - Computation Time vs. Dataset Size

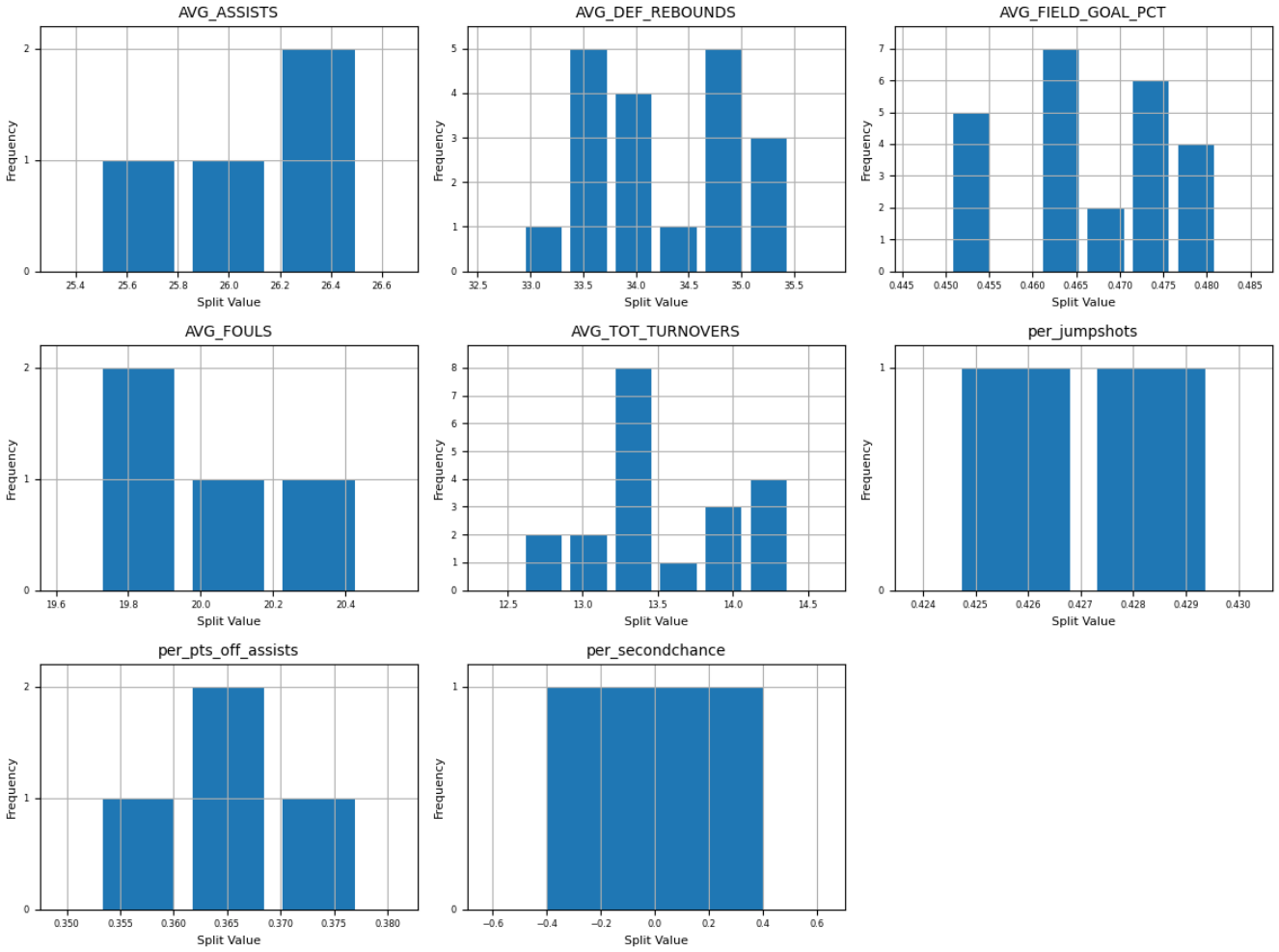

Figure 14: Split Value Histogram - NBA

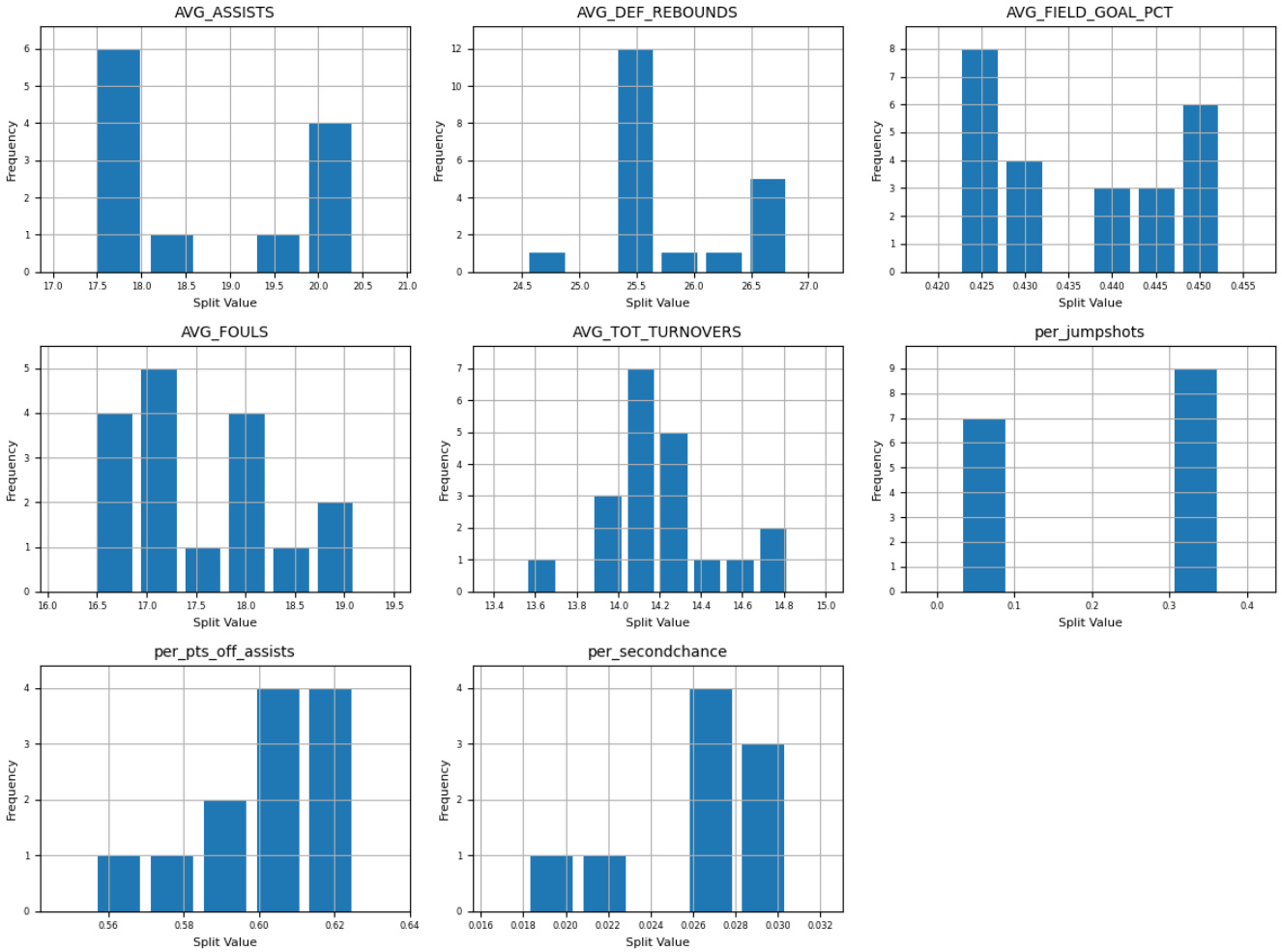

Figure 15: Split Value Histogram - WNBA

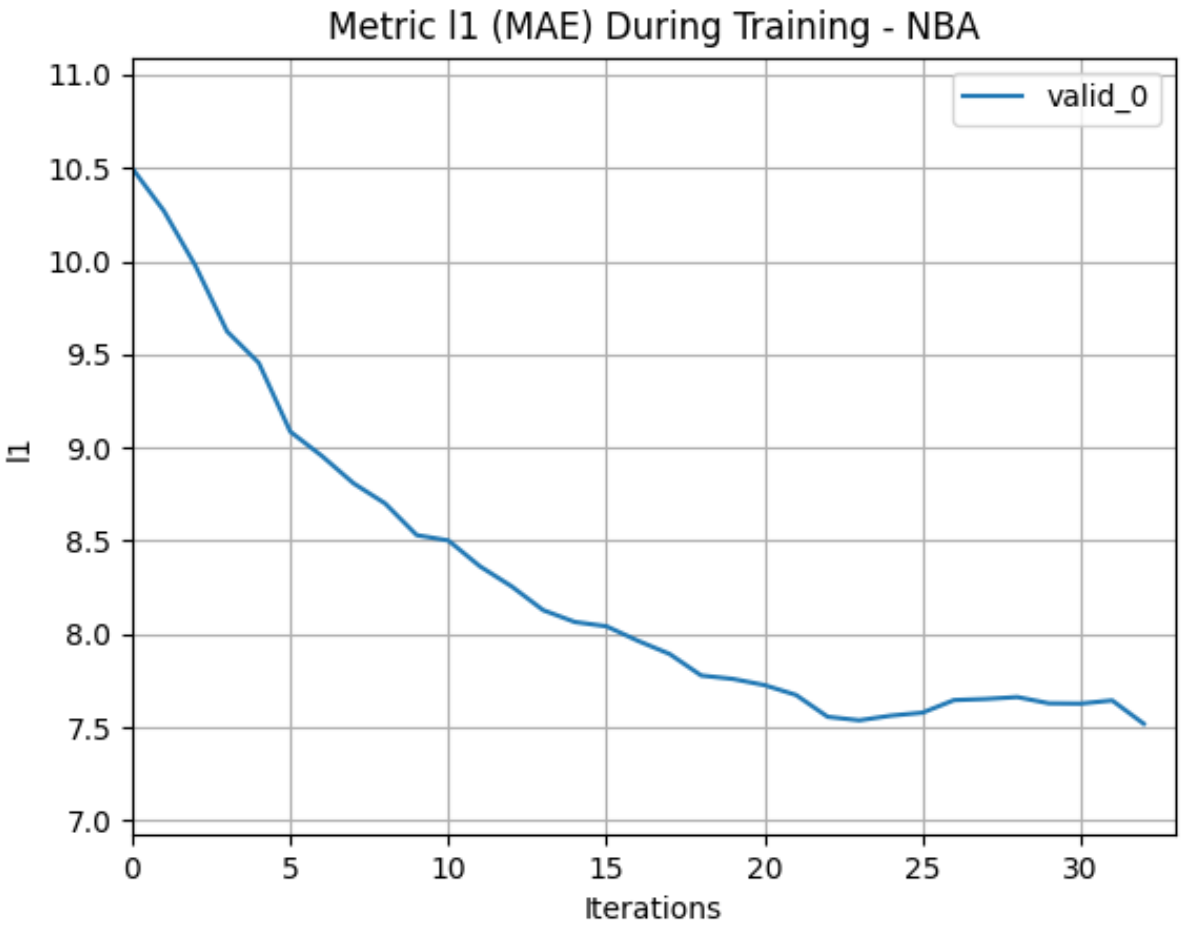

Figure 16: Metric l1 (MAE) During Training - NBA

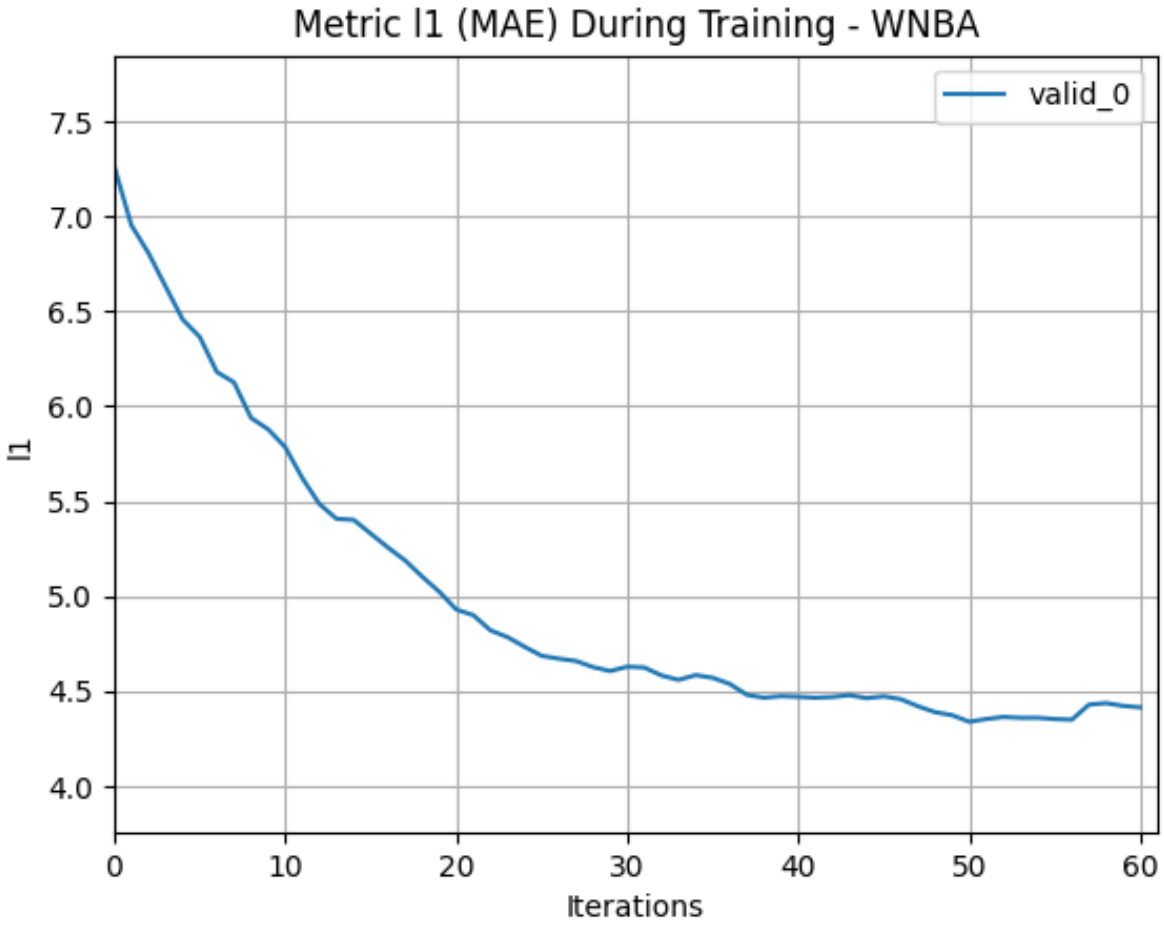

Figure 17: Metric l1 (MAE) During Training - WNBA

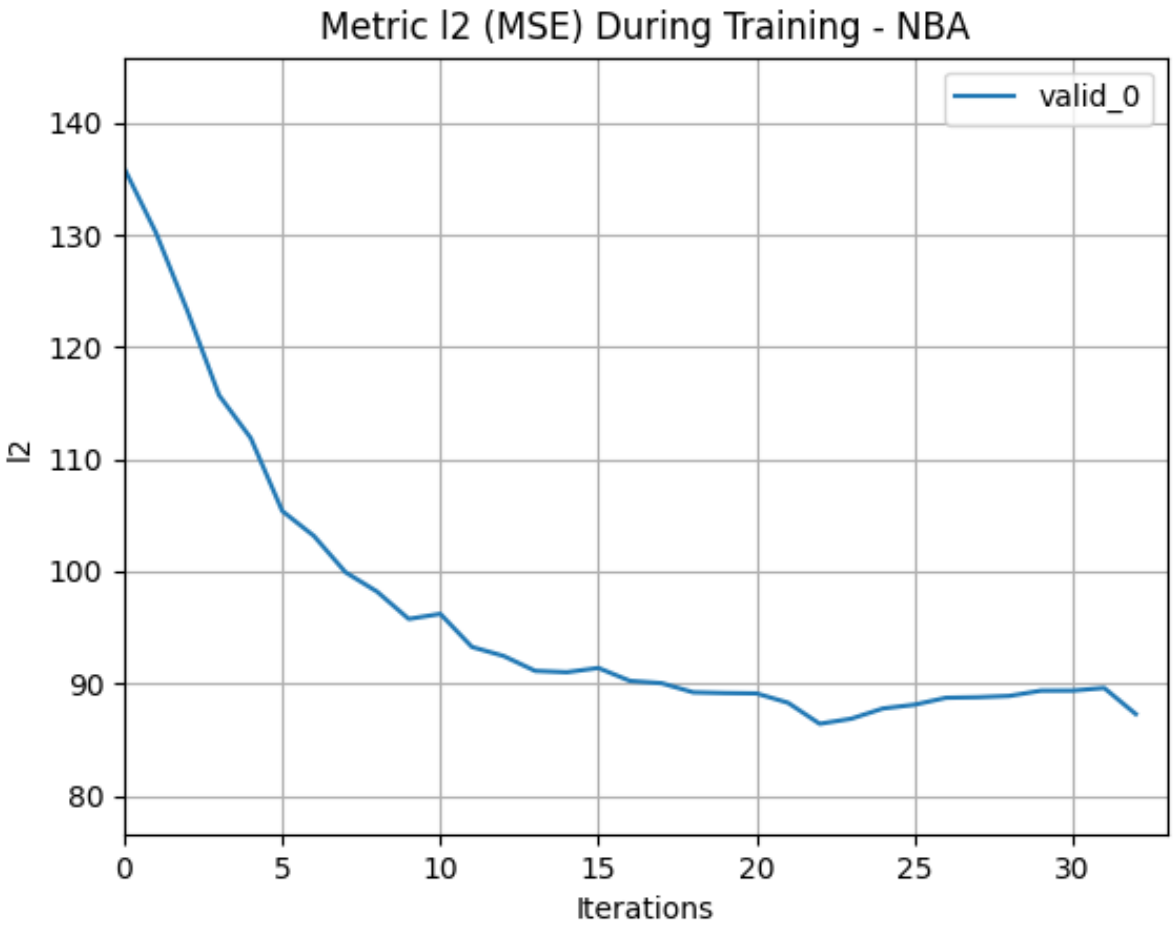

Figure 18: Metric l2 (MSE) During Training - NBA

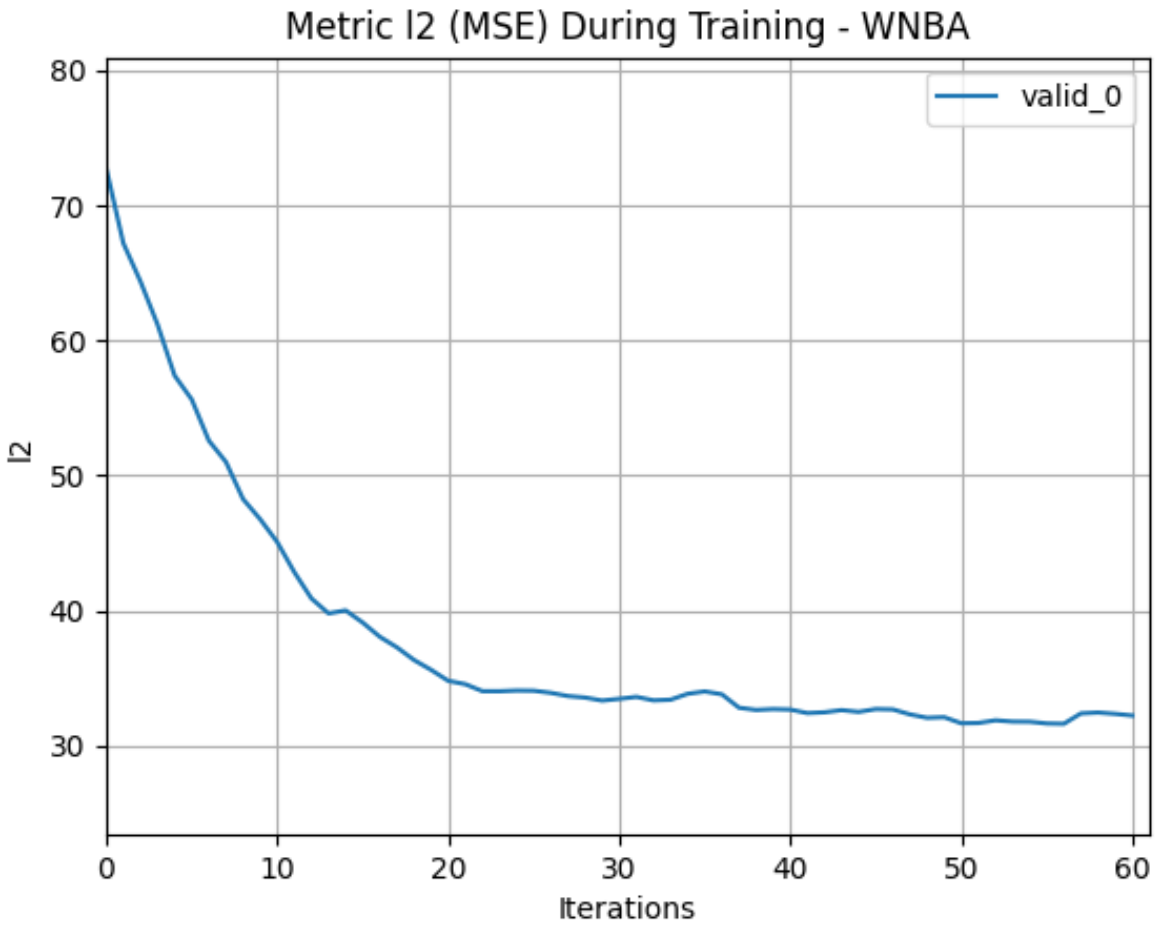

Figure 19: Metric l2 (MSE) During Training - WNBA

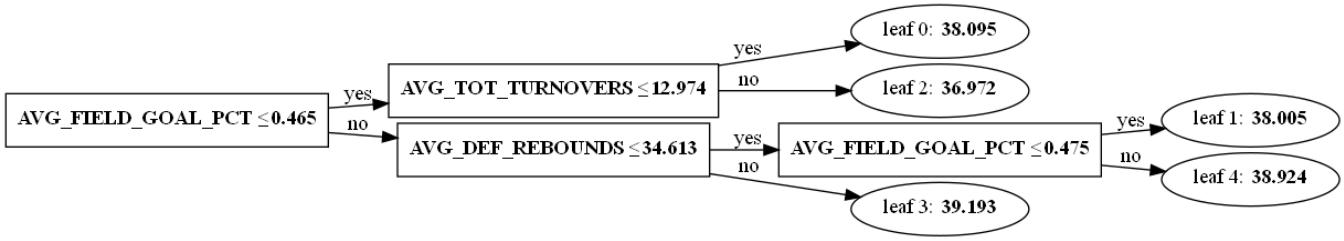

Figure 20: Example of First Decision Tree - NBA

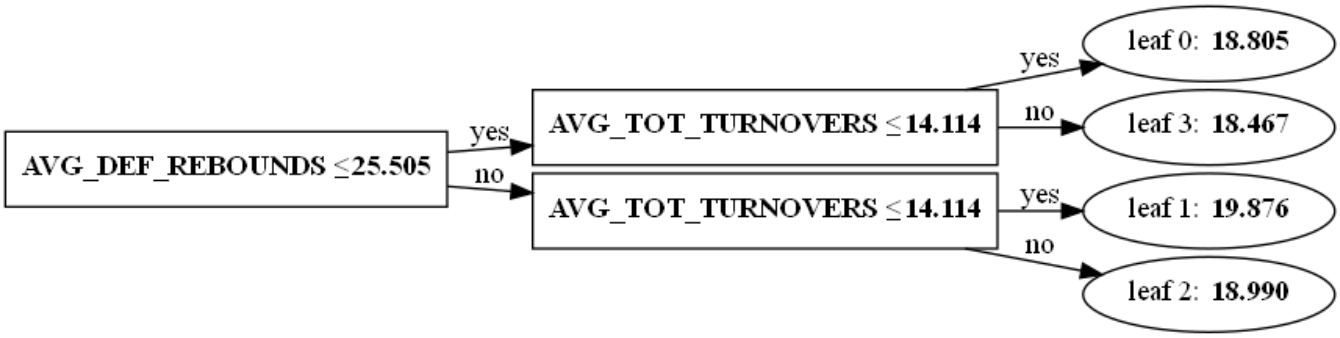

Figure 21: Example of First Decision Tree - WNBA

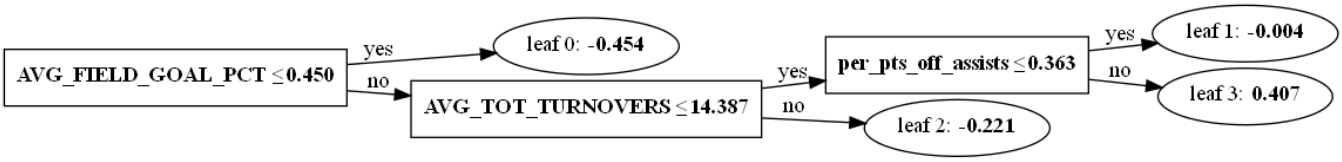

Figure 22: Example of Last Decision Tree - NBA

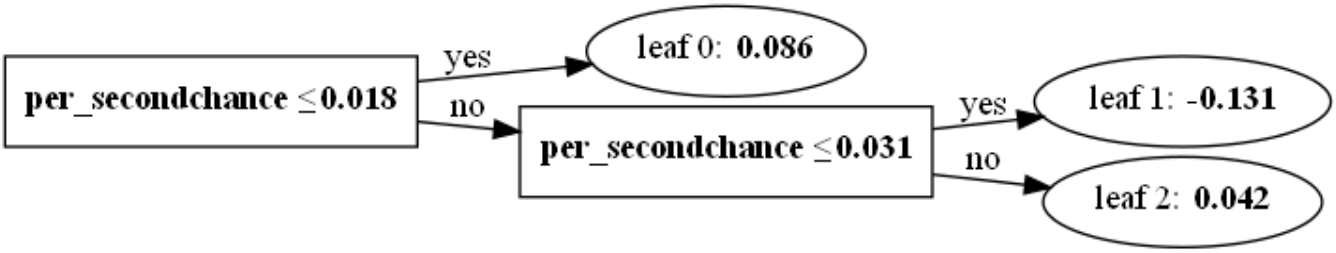

Figure 23: Example of Last Decision Tree - WNBA

league	measure	2019	2020	2021	2022	2023	2024
NBA	mean	41.00	32.37	36.00	41.00	41.00	41.00
NBA	std	12.03	10.42	10.01	11.57	10.02	13.41
NBA	min	17.00	15.00	17.00	20.00	17.00	14.00
NBA	25%	33.00	23.25	31.00	33.25	35.50	31.25
NBA	50%	41.50	30.00	37.00	43.50	42.00	46.50
NBA	75%	49.75	40.75	42.00	50.50	46.50	49.00
NBA	max	60.00	53.00	52.00	64.00	58.00	64.00
WNBA	mean	18.33	12.25	17.50	20.00	21.75	20.08
WNBA	std	7.49	6.30	7.55	8.82	10.40	8.41
WNBA	min	8.00	2.00	6.00	5.00	9.00	8.00
WNBA	25%	12.25	7.75	12.00	14.00	16.00	13.75
WNBA	50%	18.50	13.00	18.00	18.00	19.00	19.50
WNBA	75%	23.00	15.00	24.25	26.25	25.50	27.25
WNBA	max	32.00	24.00	27.00	35.00	42.00	32.00

Figure 24: Summary Statistics - Season Wins by League and Year

Data Dictionary

- COUNT_GAMES_PLAYED: The total number of games played in the regular season.
- AVG_FIELD_GOALS_MADE: The average number of field goals made per game.
- AVG_FIELD_GOALS_ATTEMPTED: The average number of field goal attempts per game.
- AVG_FIELD_GOAL_PERCENT: The average field goal shooting percentage per game.
- VARIANCE_FIELD_GOAL_PERCENT: The variance in field goal shooting percentage across games.
- AVG_THREE_POINT_FIELD_GOALS_MADE: The average number of three-point field goals made per game.
- AVG_THREE_POINT_FIELD_GOALS_ATTEMPTED: The average number of three-point field goals attempted per game.
- AVG_THREE_POINT_FIELD_GOAL_PERCENT: The average shooting percentage for three-point field goals per game.
- AVG_FREE_THROWS_MADE: The average number of free throws made per game.
- AVG_FREE_THROWS_ATTEMPTED: The average number of free throws attempted per game.
- AVG_FREE_THROW_PERCENT: The average free throw shooting percentage per game.
- AVG_COUNT_OFFENSIVE_REBOUNDS: The average number of offensive rebounds per game.
- AVG_COUNT_DEFENSIVE_REBOUNDS: The average number of defensive rebounds per game.
- AVG_COUNT_REBOUNDS: The average total rebounds (offensive and defensive) per game.
- AVG_COUNT_ASSISTS: The average number of assists per game.
- AVG_COUNT_STEALS: The average number of steals per game.
- AVG_COUNT_BLOCKS: The average number of blocks per game.
- AVG_COUNT_TURNOVERS: The average number of turnovers per game.
- AVG_COUNT_FOULS: The average number of personal fouls committed per game.
- AVG_POINTS_SCORED: The average number of points scored per game.
- AVG_PLUS_MINUS_POINTS: The average plus-minus score per game.
- AVG_POINTS_AGAINST: The average number of points scored by the opposing team per game.
- PERCENT_POINTS_OFF_ASSISTS_MADE: The percentage of total points scored that were assisted.
- PERCENT_PAINTSHOTS_MADE: The percentage of total shots made that were from the paint.
- PERCENT_JUMPSHOTS_MADE: The percentage of total shots made that were jump shots.
- PERCENT_FREETHROWS_MADE: The percentage of total points that came from free throws.
- PERCENT_SECOND_CHANCE_OPPORTUNITIES_MADE: The percentage of total shots made that were second-chance opportunities (Miss and try scoring again)

Figure 25: Dictionary of Available Independent Variables

Figure 26: Visualization Application Instances during User Feedback Period