International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(20s) 2026 pp. 1799-1810
https://doi.org/10.70917/ijcisim-2026-6106
Article

Leakage-Safe and Explainable Machine Learning for Predicting NBA Team Outcomes and Performance Deviations

Vir Agarwal

Class 12th CBSE, G.D. Goenka International School , Surat, India, Its4viragarwal@gmail.com

Abstract: Two classification tasks on NBA team standings data are examined: predicting whether a team qualifies for the playoffs from its mid-season record, and predicting whether a team's season will underperform, match, or overperform the level suggested by its previous-season record. Seven classical machine learning models (Logistic Regression, Random Forest, Extra Trees, Gradient Boosting, XGBoost, LightGBM, and CatBoost) are trained and compared under a leakage-safe framework in which features are restricted to information available at the prediction point and the data are split chronologically, with the earliest 80% used for training and the most recent 20% for testing. For playoff qualification, Random Forest is the best-performing model, reaching a test accuracy of 91.7% against a majority-class baseline of 53.3%, although accuracy is considerably lower for teams close to the eighth-seed boundary. For deviation from expected performance, Logistic Regression, tied with Gradient Boosting, is the best-performing model at 70.0% accuracy against the same 53.3% baseline. SHAP analysis identifies conference rank at mid-season as the most influential feature for playoff qualification, followed by win percentage, and the difference between early-season and previous-season win percentage as the most influential feature for deviation from expected performance.
Keywords: Sports Analytics, Machine Learning, Data Leakage Prevention, SHAP, Playoff Prediction

1. Introduction

In recent years, the field of sports analytics has grown from a peripheral curiosity into a central pillar of professional sports management, influencing decisions ranging from in-game strategy to franchise-level roster construction [1]. The National Basketball Association (NBA), in particular, has emerged as one of the most data-rich environments in professional sports, with detailed records maintained for every team, every game, and every season across its history. This growing availability of structured historical data has created an opportunity for quantitative methods to move beyond simple record-keeping and toward genuine predictive analysis. However, the mere existence of data does not guarantee useful insight; the manner in which that data is accessed, structured, and interpreted determines whether it becomes a source of understanding or a source of misleading conclusions. Poorly handled sports data can just as easily produce a model that appears accurate while actually reflecting a flawed or circular design, underscoring the importance of a disciplined analytical approach.

This tension between the abundance of raw sports data and the difficulty of extracting reliable meaning from it motivates the central problem addressed in this study. Raw historical records, on their own, are descriptive rather than predictive; they document what has already happened without offering a structured way to anticipate what is likely to happen next. Without a rigorous analytical framework, an analyst is left either to rely on subjective intuition, which does not generalize well across teams or seasons, or to construct models that inadvertently make use of information that would not actually be available at the moment a real prediction is required. There is, therefore, a genuine need for an approach that can convert historical, tabular sports records into forward-looking predictions, while remaining transparent about what information is being used and why a given prediction was made.

To address this need, the present study adopts a classical machine learning approach suited to structured, tabular data, rather than the deep learning architectures more commonly associated with unstructured data such as images or natural language. Season-level and standings-level basketball statistics are naturally organized into rows and columns, making tree-based and regression-based classifiers a more appropriate and interpretable choice than more complex architectures. The general procedure followed in this study involves acquiring historical NBA standings data, engineering features that are strictly limited to information available at a defined point in time, training and comparing multiple classical machine learning models on identical data splits, and subsequently applying explainability techniques to examine the internal reasoning of the best-performing models, rather than treating them as opaque predictive systems.

In summary, this paper proposes two related but distinct classification tasks built on NBA team standings data: a determination of whether a team will ultimately qualify for the playoffs, and a determination of whether a team's season will diverge meaningfully from the level of performance suggested by its own recent history. Both tasks are formulated in a manner that respects the chronological structure of a basketball season, ensuring that each prediction reflects only the information that would genuinely be available at the time such a prediction would be made.

The remainder of this paper is organized as follows. The proposed methodology emphasizes three aspects that distinguish it from a purely descriptive analysis of the same data: a strict separation between information available before a prediction point and information available only afterward, a comparative evaluation across multiple model families rather than reliance on a single algorithm, and the use of model explainability techniques to verify that each model's predictions are grounded in basketball-relevant reasoning rather than incidental statistical artifacts. Together, these elements form the basis for the implementation, results, and discussion presented in the sections that follow.

2. Literature Review

This section situates the present study within existing research on the application of machine learning and quantitative methods to sports outcome prediction. For each work reviewed, the discussion follows a consistent structure: an overview of the study, its methodology, its principal findings, and the specific respect in which the present study differs from it.

A. Arur and Shanbhag [3]

Arur and Shanbhag [3] present a study predicting Indian Premier League (IPL) cricket match outcomes using a range of classical machine learning classifiers applied to historical match-level data. The study employs Gradient Boosting, Decision Tree, Random Forest, and Light Gradient Boosting Machine classifiers trained on pre-match attributes such as the competing teams, venue, and toss outcome, following categorical encoding and hyperparameter tuning via randomized search with three-fold cross-validation. Team names were standardized to account for franchise renaming, and rows with missing outcome data were removed prior to model training. The Light Gradient Boosting Machine model achieved the strongest performance among the classifiers evaluated, with player-of-match history and the identities of the competing teams emerging as the most influential predictors according to feature-importance analysis. The authors report that explainable, tree-based ensemble methods are well suited to this classification task, while noting that the model has not been externally validated beyond the dataset examined. Unlike the present study, this work does not employ a chronologically ordered train-test partition or a mid-point forecasting snapshot, and relies on feature-importance rather than SHAP-based explainability [6], whereas the present study explicitly enforces temporal separation between training and testing data and applies SHAP to interpret model behavior at defined prediction points within a season.

B. Ouyang et al. [4]

Ouyang et al. [4] investigate the use of XGBoost combined with SHAP-based explainability to predict NBA game outcomes at multiple points during live gameplay, using statistics accumulated as the game progresses. The study compares XGBoost against several classifiers, including Logistic Regression, Support Vector Machines,

Random Forest, and LightGBM, evaluating each at successive stages of a game as additional quarters are completed. SHAP is applied to identify which accumulated in-game statistics, such as field-goal percentage and turnovers, most strongly influence the model's evolving predictions as the game progresses. XGBoost is identified as the best-performing model across accuracy, precision, and area-under-curve metrics at every stage of the game examined, with predictive reliability increasing substantially as more of the game is completed. SHAP analysis indicates that field-goal percentage and defensive rebounding are consistently influential, while the relative importance of assists and offensive rebounding shifts between the early and late stages of a game. This study predicts outcomes using statistics accumulated during the very game being predicted, whereas the present study restricts itself exclusively to information available before a defined prediction point, at either the mid-season or early-season level, precluding any use of in-progress or same-event outcome data.

C. Thabtah et al. [5]

Thabtah, Zhang, and Abdelhamid [5] examine the prediction of NBA game results using historical team statistics and a comparison of several classical machine learning classifiers. The study applies Naive Bayes, Artificial Neural Network, and Decision Tree algorithms to historical game data, systematically varying the feature set used for training to assess the impact of individual statistical categories on predictive performance, with particular attention to identifying which combination of features yields the most reliable classifier. Defensive rebounds are identified as the single most influential feature for predicting game outcomes, with the inclusion of additional features such as three-point percentage, free throws, and total rebounds improving classification accuracy over a baseline feature set. A decision-tree-based approach is reported as the most effective of the models compared. This study relies on box-score statistics drawn from completed historical games without addressing season-level forecasting or the specific leakage risks associated with predicting a team's eventual standing, both of which are central concerns explicitly designed around in the present study.

3. Implementation

A. Dataset

The dataset used in this study is a publicly available collection of NBA data hosted on Kaggle under the name "nba-games" [2]. It contains authentic records of NBA games and team standings. The primary file used for this study is ranking.csv, which records, for every team and for every date within every season, the number of games played, the number of wins and losses, the corresponding win percentage, and the team's separate home and road records. Table 1 describes the specific columns extracted from this file for use in the analysis.

Column Name	Description
TEAM_ID	A numerical identifier corresponding to a specific team
SEASON_ID	The NBA season to which the record belongs
STANDINGSDATE	The date on which the standings snapshot was recorded
CONFERENCE	The team's conference, Eastern or Western
G	The number of games played by the team up to this point
W / L	The number of wins and losses recorded to this point
W_PCT	Win percentage, calculated as wins divided by games played
HOME_RECORD / ROAD_RECORD	Win-loss record at home and away, expressed in the form '20-5'

TABLE 1. COLUMNS UTILIZED FROM THE NBA STANDINGS DATASET

This single source file was used to construct two independent classification models, each addressing a distinct question regarding an NBA team's season:

1. Will the team qualify for the playoffs?

2. Will the team's season outcome diverge substantially from the level of performance suggested by its record in the preceding season?

B. Prevention of Data Leakage

The central methodological constraint governing both models is that a model must be restricted to information that would genuinely be available prior to the outcome it is attempting to predict. Where a model is permitted to access information derived from the outcome itself, such as a team's final win total when predicting whether that same team made the playoffs, the resulting model exhibits what is termed data leakage. A model affected by data leakage typically performs exceptionally well under evaluation while providing no genuine predictive capability in a real-world setting, since the information it depends upon would not actually be available at the point a prediction is required.

To guard against this, both models are restricted to a snapshot of the season captured at an intermediate point in time, rather than the season's final outcome:

The Playoff Qualification model is restricted to each team's record after a fixed number of games played, corresponding to approximately the midpoint of a standard NBA season, and is tasked with determining whether that team ultimately finishes within the top eight positions of its conference.

The Beat Expectations model is restricted to a team's performance across a small number of games at the start of the current season, combined with that team's final record from the previous season, and is tasked with classifying the team's eventual season outcome as substantially better than expected, substantially worse than expected, or consistent with expectations. A substantial divergence is defined as a change in win percentage, relative to the prior season, that exceeds a fixed threshold.

An additional feature was incorporated into each model to capture longer-term team tendencies without violating the leakage constraint. For the Playoff Qualification model, this feature reflects the frequency with which a given team qualified for the playoffs in prior seasons. For the Beat Expectations model, this feature is the team's final win percentage from the immediately preceding season. Both quantities are derived exclusively from completed prior seasons and therefore introduce no risk of future information influencing a given prediction.

C. Chronological Data Partitioning

Evaluating a predictive model requires testing it against data it has not previously encountered. A conventional approach involves randomly partitioning the dataset into training and testing subsets; however, this approach is not appropriate for time-dependent sporting data, as it would permit information from a later season to inform predictions concerning an earlier one, a scenario that does not reflect any realistic use of the model. Instead, all team-season records were ordered chronologically, with the earliest eighty percent allocated to training and the most recent twenty percent reserved for testing. This partitioning method imposes a stricter and more representative evaluation standard.

D. Model Selection and Configuration

Rather than relying on a single algorithm, seven established machine learning classifiers were trained on identical data and compared under identical conditions: Logistic Regression, Random Forest, Extra Trees, Gradient Boosting, XGBoost, LightGBM, and CatBoost. Table 2 summarizes the primary configuration parameters applied to each model.

Model	Primary Configuration
Logistic Regression	max_iter = 1000
Random Forest	300 estimators, maximum depth = 6
Extra Trees	300 estimators, maximum depth = 6

Gradient Boosting	200 estimators, maximum depth = 3, learning rate = 0.05
XGBoost	300 estimators, maximum depth = 4, learning rate = 0.05
LightGBM	300 estimators, maximum depth = 4, learning rate = 0.05
CatBoost	300 estimators, maximum depth = 4, learning rate = 0.05

TABLE 2. CONFIGURATION PARAMETERS APPLIED TO EACH MODEL

Categorical attributes, including team identifiers and conference labels, were converted into numerical form through label encoding prior to model training, consistent with standard practice for classical machine learning approaches applied to tabular data.

4. Iv. Results

A. Model 1: Playoff Qualification

Following the construction of the midpoint-season snapshot for each team and season, four hundred and forty-nine team-seasons were available for training, with one hundred and twenty more recent team-seasons reserved for testing. Model performance was evaluated using five standard classification metrics: Accuracy, the overall proportion of correct predictions; Precision, the proportion of teams predicted to qualify for the playoffs that actually did so; Recall, the proportion of teams that actually qualified for the playoffs that were correctly identified; F1-score, a harmonic combination of Precision and Recall; and ROC-AUC, a measure ranging from 0.5, corresponding to random guessing, to 1.0, corresponding to perfect discrimination between playoff and non-playoff teams.

Random Forest was identified as the best-performing model. Its accuracy of 91.7% (110 of 120 test team-seasons) substantially exceeded that of a naive baseline in which the majority class is predicted for every case (53.3%), a comparison that is essential, since a model is only meaningfully useful if it exceeds this baseline by a considerable margin. Fig. 1 presents accuracy and ROC-AUC across all seven models, together with the baseline for reference.

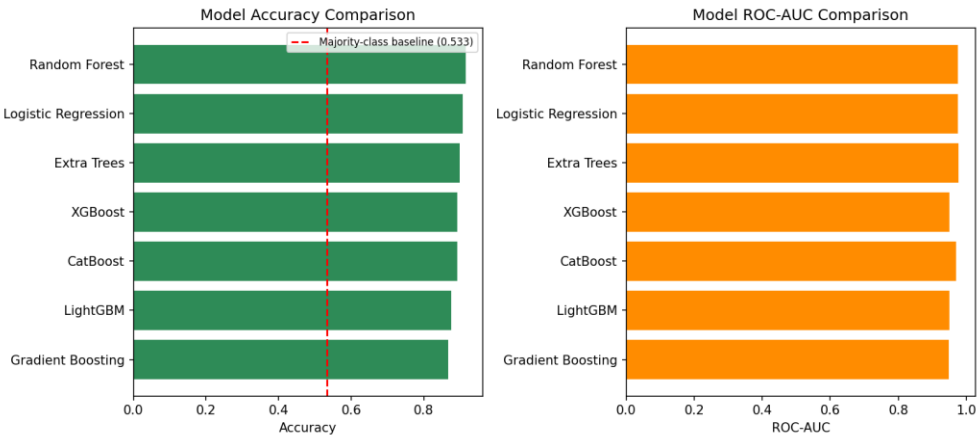


FIG. 1. ACCURACY (LEFT) AND ROC-AUC (RIGHT) ACROSS ALL SEVEN PLAYOFF QUALIFICATION MODELS, RELATIVE TO THE BASELINE

Fig. 2 presents the confusion matrix for the Random Forest model, the best-performing configuration identified in this study. A confusion matrix summarizes, for each true outcome, how frequently the model produced each possible prediction, thereby distinguishing correct classifications from the two possible types of error.

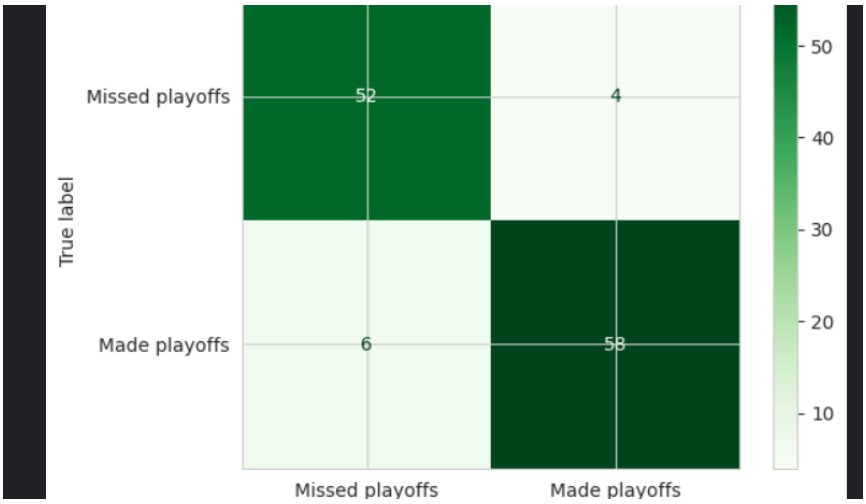


FIG. 2. CONFUSION MATRIX FOR THE RANDOM FOREST MODEL

While the overall accuracy achieved by this model is high, it is necessary to consider whether this reflects genuine predictive capability or simply the fact that most teams are not positioned near the playoff boundary at the season's midpoint. A team with a strongly winning record at the midpoint of the season is an unambiguous playoff qualifier, just as a team with a strongly losing record is an unambiguous non-qualifier; neither case constitutes a meaningful test of the model. The genuinely informative cases are those teams positioned near the eighth and final playoff position within their conference. Accuracy was therefore additionally examined as a function of each team's proximity to this boundary, with results presented in Fig. 3.

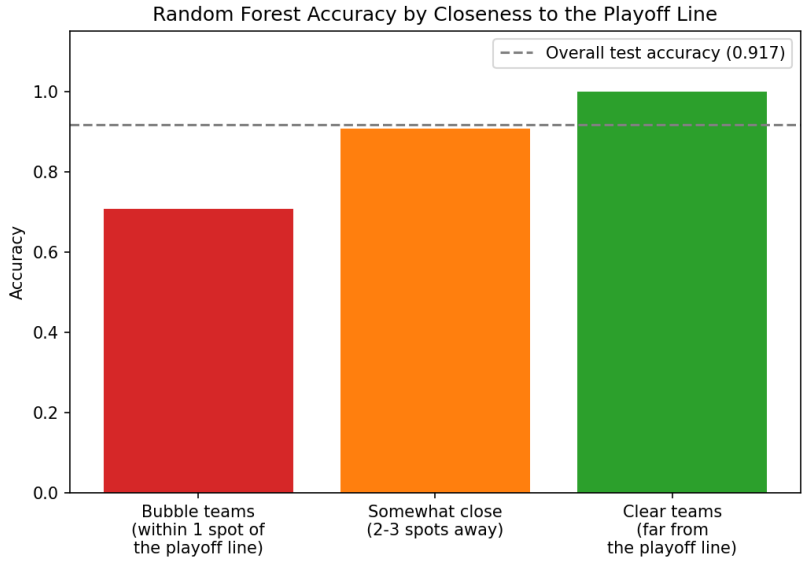


FIG. 3. RANDOM FOREST ACCURACY AS A FUNCTION OF PROXIMITY TO THE PLAYOFF BOUNDARY

As illustrated in Fig. 3, accuracy for teams situated immediately adjacent to the playoff boundary was considerably lower than the overall figure, while accuracy for teams clearly separated from the boundary approached, and in some cases reached, perfect classification. Both the aggregate accuracy figure and this boundary-proximity analysis are valid; they simply address different questions. The aggregate figure represents the appropriate overall measure of model accuracy, whereas the boundary-proximity analysis more accurately reflects the model's capability

in the specific scenario of greatest practical interest: determining which teams narrowly qualify for or narrowly miss the playoffs.

To examine the internal reasoning underlying the Random Forest model's predictions, SHAP (SHapley Additive exPlanations) [6], [7] was applied. SHAP quantifies, for each input feature, the average magnitude and direction of that feature's influence on the model's output across the test set, thereby providing a principled means of interpreting an otherwise opaque model. Fig. 4 presents the resulting feature importance ranking.

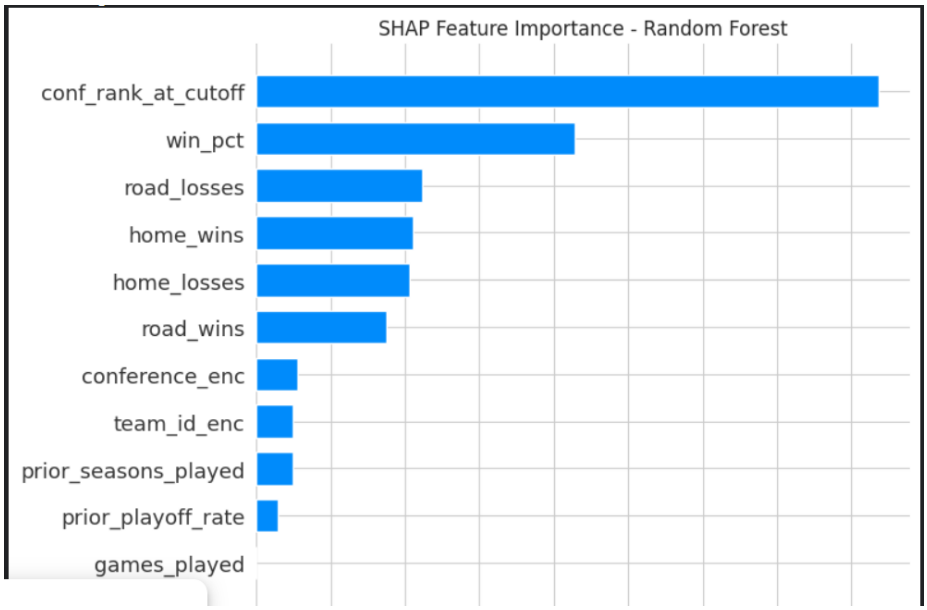


FIG. 4. SHAP FEATURE IMPORTANCE FOR THE RANDOM FOREST PLAYOFF QUALIFICATION MODEL

The team's conference rank at the midpoint of the season emerged as the dominant factor governing model predictions, followed by overall win percentage and the separate home and road win-loss records. This outcome is consistent with basic domain expectations, as a team's position in the standings partway through a season constitutes the most direct available indicator of its likely position at the season's conclusion.

B. Model 2: Deviation from Expected Performance

The second model addresses a materially different and more relative question, namely whether a team is performing in a manner consistent with, or divergent from, its own recent history, rather than whether the team is objectively strong or weak. Four hundred and seventy-eight team-seasons were available for training, with one hundred and twenty reserved for testing. This task involves three possible outcomes rather than two: Underperformed, As Expected, and Overperformed. Because the classification task is no longer binary, Precision, Recall, and F1-score were computed using macro-averaging, whereby each of the three outcome classes is weighted equally regardless of its relative frequency in the dataset.

Logistic Regression, the least complex of the seven models evaluated, achieved the highest accuracy (70.0%), sharing this distinction with Gradient Boosting. This result exceeds the majority-class baseline of 53.3%, though by a narrower margin than was observed for the Playoff Qualification model, a difference consistent with the relative difficulty of the underlying task: forecasting a season-long deviation in performance from only a small number of early-season games is inherently more uncertain than characterizing a team's standing from a substantially larger sample of games. Fig. 5 presents a comparative evaluation of all seven models.

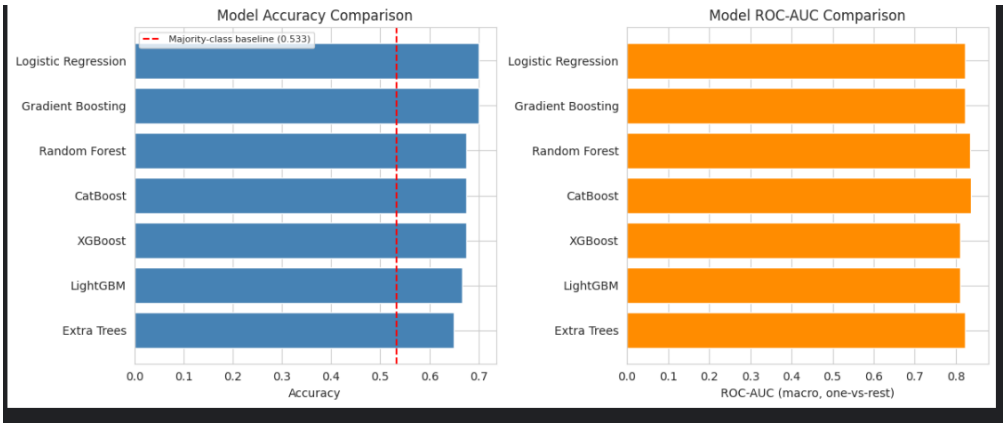


FIG. 5. ACCURACY (LEFT) AND ROC-AUC (RIGHT) ACROSS ALL SEVEN 'BEAT EXPECTATIONS' MODELS

Fig. 6 presents the confusion matrix for the best-performing model. As three outcome classes are involved, the confusion matrix takes the form of a three-by-three grid rather than the two-by-two grid used for the binary classification task. The model was most reliable when classifying team-seasons as consistent with expectations, this being the most frequently occurring outcome; its capacity to correctly identify the two divergent outcomes, while more limited, remained substantially above what would be expected under random assignment.

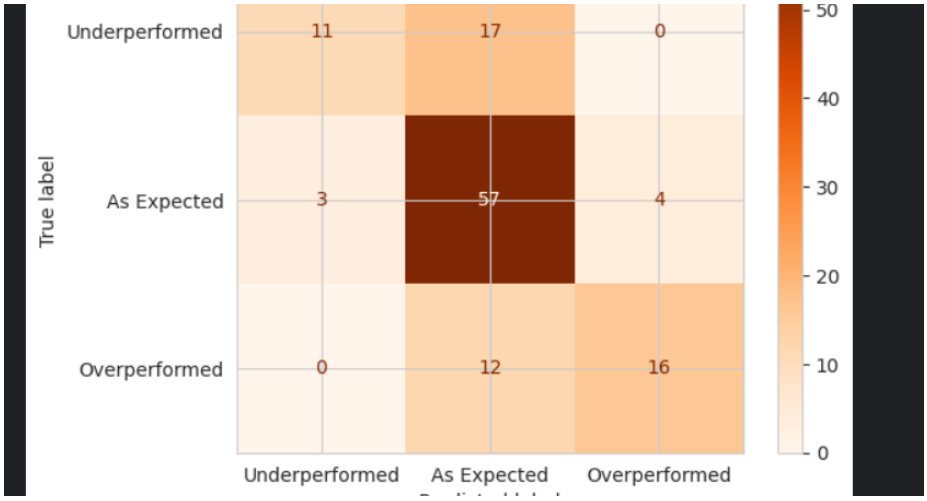


FIG. 6. CONFUSION MATRIX FOR THE LOGISTIC REGRESSION MODEL

Consistent with the analysis applied to the Playoff Qualification model, accuracy was examined as a function of the magnitude of each team's actual deviation from its expected performance, distinguishing team-seasons whose outcome fell close to the classification threshold from those whose outcome was more pronounced. Fig. 7 presents this analysis.

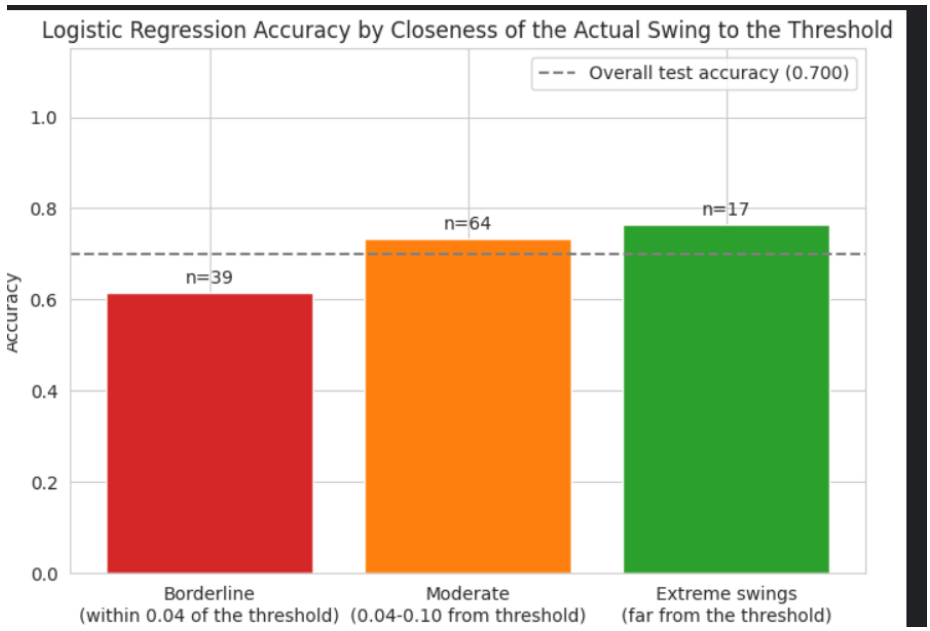


FIG. 7. LOGISTIC REGRESSION ACCURACY AS A FUNCTION OF PROXIMITY TO THE CLASSIFICATION THRESHOLD

Team-seasons whose actual deviation fell close to the classification threshold were associated with markedly reduced accuracy, only modestly exceeding random assignment, whereas team-seasons with more moderate deviations showed a comparatively small improvement, and those with the most pronounced deviations showed a further, though still limited, improvement. This gradual pattern contrasts with the considerably steeper improvement observed for the Playoff Qualification model. This distinction is consistent with the nature of each underlying quantity: a season-long deviation accumulates gradually across many games and therefore rarely becomes as unambiguous in retrospect as a team's standing at a fixed midpoint in the season. As with the first model, the aggregate accuracy figure and the threshold-proximity figure are both valid, and jointly provide a more complete characterization of model performance than either figure would in isolation.

SHAP analysis [6], [7] was again applied, in this instance using the XGBoost model, to identify the features most strongly associated with a predicted Overperformed outcome. Fig. 8 and Fig. 9 present the corresponding results.

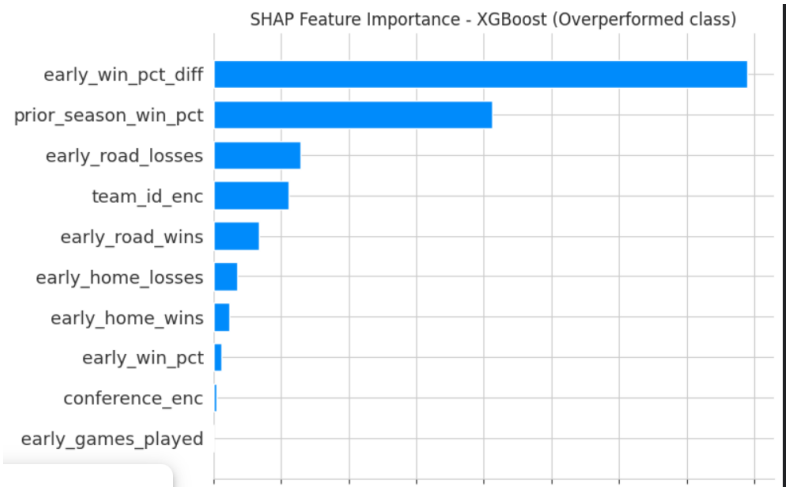


FIG. 8. SHAP FEATURE IMPORTANCE FOR PREDICTION OF AN OVERPERFORMED OUTCOME

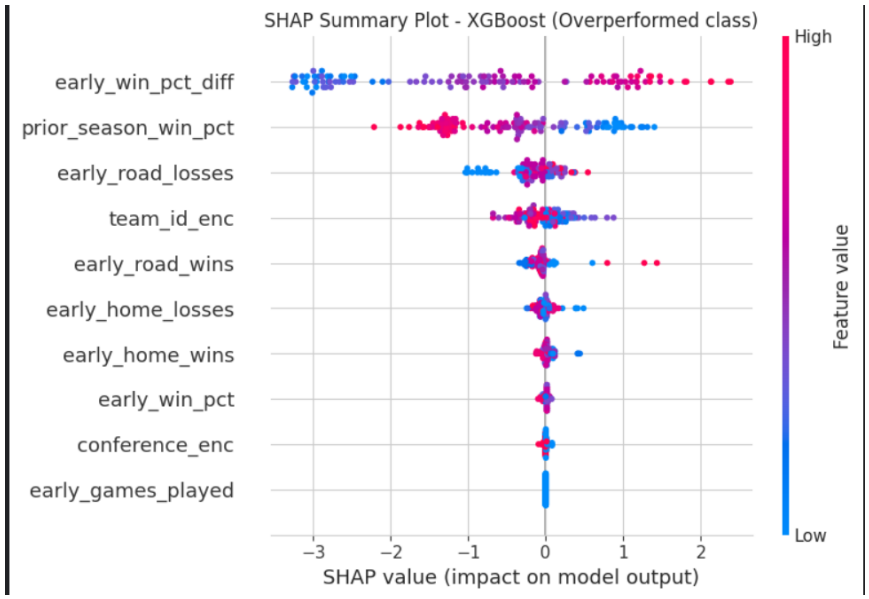


FIG. 9. SHAP SUMMARY PLOT FOR PREDICTION OF AN OVERPERFORMED OUTCOME

The most influential feature was the difference between a team's early-season win percentage and its win percentage in the preceding season, representing the degree of early-season improvement or decline. As shown in Fig. 9, elevated values of this feature are associated with a positive contribution toward an Overperformed classification, a relationship consistent with the underlying basketball rationale. The second most influential feature was the team's win percentage in the preceding season considered independently.

5. Discussion

A. Summary

The results obtained in this study indicate that both classification tasks benefit meaningfully from the leakage-safe, chronologically validated methodology adopted throughout this work. Random Forest emerged as the best-performing model for the task of playoff qualification, while Logistic Regression and Gradient Boosting emerged jointly as the strongest performers for the task concerned with deviation from expected performance. Considered together, these results indicate that a team's standing at a defined point in the season provides a considerably more reliable predictive signal than short-term changes in early-season form, and that the relative difficulty of a forecasting task is closely tied to how much of the season has already elapsed at the point a prediction is made. Rather than treating the two tasks as interchangeable applications of the same modeling approach, the results suggest that they occupy meaningfully different positions along a spectrum of predictive difficulty, with playoff qualification approaching a comparatively tractable classification problem and deviation from expected performance remaining a substantially harder, more uncertain one. This distinction, rather than the methodology alone, is the principal finding to emerge from comparing the two tasks directly.

B. SHAP-Based Discussion

Although the quantitative evaluation metrics reported in this study establish that both models perform meaningfully above their respective baselines, a purely quantitative assessment does not, by itself, establish whether a model has learned a basketball-relevant pattern or has instead identified an incidental statistical regularity within the training data. To address this, SHAP (SHapley Additive exPlanations) analysis [6], [7] was applied to the best-performing tree-based configuration for each task, converting the internal, quantitative computations of each model into a qualitative account of which factors most influenced its predictions. For the playoff qualification task, the analysis indicated that a team's standing within its conference at the point of prediction was the dominant influence

on the model's output, ahead of raw win percentage and home or road performance splits, a hierarchy that closely mirrors the reasoning a human analyst would apply to the same problem. For the task concerned with deviation from expected performance, the analysis indicated that the degree of early-season improvement or decline, measured relative to the team's performance in the preceding season, was the dominant influence, followed by the prior season's win percentage considered independently. In both cases, the qualitative explanations produced through SHAP corroborate the quantitative results obtained through standard evaluation metrics, providing a degree of confidence that the models are reasoning in a manner consistent with domain expectations rather than exploiting arbitrary correlations present in the training data.

C. Strengths and Limitations

The principal strength of the methodology adopted in this study lies in its explicit prevention of data leakage and its use of a chronological, rather than random, partitioning of the data, both of which ensure that reported performance reflects a realistic forecasting scenario rather than an inflated, retrospective fit. The comparative evaluation of seven distinct model families, rather than a single algorithm, further strengthens the reliability of the conclusions drawn, as the relative performance of these models differed meaningfully between the two tasks under consideration. Nevertheless, certain limitations should be acknowledged. The analysis presented here relies on a single dataset source, and the reported models have not been validated through cross-validation across independent or externally sourced datasets, which would be necessary to establish that the observed performance generalizes beyond the specific seasons examined. Additionally, the feature set used in this study is confined to standings-derived statistics and does not incorporate player-level, injury-related, or other contextual information that may plausibly influence a team's trajectory. Future work extending this methodology to multiple independent datasets, or incorporating a broader range of contextual features, would help to establish the extent to which the present findings generalize beyond the data examined here.

6. Conclusion

Across both classification tasks examined in this study, Random Forest was identified as the best-performing model for playoff qualification, while Logistic Regression, alongside Gradient Boosting, was identified as the best-performing model for predicting deviation from expected performance.

The most significant inference to be drawn from this study is that a team's standing at a fixed point in time carries substantially more predictive value than short-term deviations from historical form, indicating that sustained, accumulated performance remains a considerably stronger signal than early-season fluctuation for the purpose of season-outcome forecasting.

7. Acknowledgement

The author acknowledges the use of artificial intelligence (AI) tools in the preparation of this manuscript, specifically Claude Sonnet 5, a large language model developed by Anthropic. This AI tool was used in four specific capacities: generating explanatory content accompanying the analysis; constructing and formatting the tables presented throughout this paper; refining and improving the clarity of the written language; and helping to build and implement several of the machine learning models discussed in this study. All experimental design choices, interpretations, and conclusions remain those of the author. The author additionally acknowledges the creators and maintainers of the "nba-games" dataset made publicly available on Kaggle, without which this study would not have been possible.

REFERENCES

1. Wang, Henry, Arnab Sarker, and Anette Hosoi. "The Effect of Basketball Analytics Investment on National Basketball Association (NBA) Team Performance." Journal of Sports Economics 26, no. 6 (2025): 668–688. https://doi.org/10.1177/15270025251328264.
2. nathanlauga. "NBA Games." Kaggle. Accessed 2026. https://www.kaggle.com/datasets/nathanlauga/nba-games.

3. Arur, Vrushank Mahesh, and Aakash Dhananjay Shanbhag. "Indian Premier League Cricket Match Winner Predictor 2025: A Sports Analytics Study." In Proceedings of the 2024 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 1–6. Male, Maldives: IEEE, 2024.
4. Ouyang, Yida, Xiang Li, Wenjie Zhou, Wenjie Hong, Wei Zheng, Fei Qi, and Lifang Peng. "Integration of Machine Learning XGBoost and SHAP Models for NBA Game Outcome Prediction and Quantitative Analysis Methodology." PLOS ONE 19, no. 7 (2024): e0307478. https://doi.org/10.1371/journal.pone.0307478.
5. Thabtah, Fadi, Li Zhang, and Neda Abdelhamid. "NBA Game Result Prediction Using Feature Analysis and Machine Learning." Annals of Data Science 6, no. 1 (2019): 103–116. https://doi.org/10.1007/s40745-018-00189-x.
6. Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." In Advances in Neural Information Processing Systems 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, 4765–4774. Red Hook, NY: Curran Associates, Inc., 2017.
7. K. Patel, and A. Shanbhag, “Exploring ML for Predictive Maintenance Using Imbalance Correction Techniques and SHAP,” in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pp. 1–10, 2022.