

## Article

# Application of Machine Learning Models for Baseball Outcome Prediction

Tzu-Chien Lo <sup>1,†</sup>, Chen-Yin Lee <sup>2,†</sup>, Chien-Lin Chen <sup>3,\*</sup>, Tsung-Yu Hsieh <sup>1</sup>, Che-Hsiu Chen <sup>4</sup> and Yen-Kuang Lin <sup>5,\*</sup>

<sup>1</sup> Department of Physical Education, Fu Jen Catholic University, New Taipei City 24205, Taiwan; 093008@mail.fju.edu.tw (T.-C.L.); yu691207@gmail.com (T.-Y.H.)

<sup>2</sup> Center for Teacher Education, Fu Jen Catholic University, New Taipei City 24205, Taiwan; 158275@mail.fju.edu.tw

<sup>3</sup> Physical Education Office, Fu Jen Catholic University, New Taipei City 24205, Taiwan

<sup>4</sup> Department of Sports Performance, National Taiwan University of Sport, Taichung 40401, Taiwan; jakic1114@ntus.edu.tw

<sup>5</sup> Graduate Institute of Athletics and Coaching Science, National Taiwan Sport University, Taoyuan 33301, Taiwan

\* Correspondence: 1100506@ntsu.edu.tw (C.-L.C.); robbinlin@ntsu.edu.tw (Y.-K.L.)

† These authors contributed equally to this work as the first authors.

## Abstract

Data science has become an essential component in professional sports, particularly for predicting team performance and outcomes. This study aims to develop and evaluate machine learning models that accurately predict game outcomes in the Chinese Professional Baseball League (CPBL). Method: A total of 859 games from the 2021 to 2023 regular seasons were analyzed, using both traditional baseball statistics and advanced sabermetric indicators such as the Weighted Runs Created Plus (wRC+), Weighted Runs Above Average (wRAA), and Percentage of Leadoff Batters on Base (PLOB%). Five machine learning models—decision tree, logistic regression, Neural Network, Random Forest, and XGBoost—were constructed and assessed through a five-fold cross-validation. Evaluation metrics included accuracy, F1 scores, sensitivity, specificity, and the AUC-ROC. Results: Among the models, logistic regression and XGBoost achieved the highest performance, with an accuracy ranging from 0.89 to 0.93 and an AUC-ROC from 0.97 to 0.98. The feature importance and SHapley Additive exPlanations (SHAP) analysis revealed that the wRC+ and PLOB% were the most influential predictors, reflecting the offensive efficiency and pitching control. Conclusion: The results suggest that combining interpretable machine learning with sabermetrics provides valuable insights for coaches and analysts in professional baseball. Furthermore, incorporating performance weighting based on game context may further enhance model accuracy. This research demonstrates the potential of data-driven strategies in sports analytics and decision-making.

**Keywords:** Chinese Professional Baseball League; performance analysis; sport big data; weight index; explainable AI



Academic Editors: Marios Hadjicharalambous and Nikolaos Zaras

Received: 17 April 2025

Revised: 18 June 2025

Accepted: 18 June 2025

Published: 24 June 2025

**Citation:** Lo, T.-C.; Lee, C.-Y.; Chen, C.-L.; Hsieh, T.-Y.; Chen, C.-H.; Lin, Y.-K. Application of Machine Learning Models for Baseball Outcome Prediction. *Appl. Sci.* **2025**, *15*, 7081. <https://doi.org/10.3390/app15137081>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Big Data, or “massive data,” refers to datasets so large and complex that traditional analytical methods prove inadequate. Advanced technologies such as artificial intelligence, machine learning, and predictive analytics are necessary to derive meaningful insights. Big Data analytics has become a global trend, widely applied across industries for forecasting and decision-making [1]. In sports, it has revolutionized traditional strategies by leveraging

computational models to enhance athlete performance [2]. The film *Moneyball* exemplifies this shift, illustrating how data-driven approaches can challenge conventional wisdom and optimize team value [3].

Given the uncertainty and numerous variables in sports, predicting match outcomes has gained prominence in sports analytics [4]. While audiences often focus on final scores, fewer examine the intricate progression of games [5,6]. Play-by-play records objectively capture on-field actions, offering coaches and players valuable feedback. When systematically quantified, these statistics reveal causal links between gameplay elements and outcomes. Modern analytics companies have enhanced traditional metrics using mathematical transformations to better reflect game realities. For example, fielding independent pitching (FIP) isolates a pitcher's performance by eliminating the impact of relief pitchers, offering a more accurate evaluation than the earned run average (ERA). In basketball, the true shooting percentage (TS%) integrates two-point, three-point, and free-throw data to more comprehensively assess scoring efficiency [7,8]. Such advanced statistics often incorporate weighted formulas, enhancing their predictive power and practical value [4].

Unlike traditional statistical models, which emphasize causal inference, machine learning prioritizes predictive accuracy—even at the expense of interpretability. For instance, ref. [9] reported that logistic regression achieved only a 61.11% accuracy, while Artificial Neural Networks (ANNs) improved the performance to 72.22%. Similarly, ref. [10] tested regression, Bayesian classification, support vector machines (SVMs), and ANNs on five years of NBA game data, finding that the regression yielded the highest accuracy at 69.67%. These results suggest that logistic regression typically provides moderate prediction rates ranging from 60% to 70%.

The ANN has gained popularity due to its flexibility in modeling nonlinear patterns and minimal reliance on strict statistical assumptions. A study on college basketball reported an 89.42% accuracy using an ANN [11], while another analyzing Serbian league data achieved 80.96% [12]. In baseball, ref. [13] used a dataset with 26 batter and 34 pitcher variables to predict MLB outcomes and found that an ANN, with five-fold cross-validation, outperformed other models, reaching 93.91% accuracy. Decision tree models have also been applied, achieving a 78.2% accuracy with decision pruning and identifying key variables [14].

Further extending these findings, ref. [9] again applied both logistic regression and an ANN to forecast Yankees vs. Red Sox outcomes, reporting a 69.23% and 72.22% accuracy, respectively. Other research employed SVMs with 16 input features, noting that metrics like the slugging percentage and double plays aligned well with predictions from a Gaussian RBF kernel, achieving a 69% accuracy [15]. In another study, ref. [16] reduced a 60-variable MLB dataset to 15 key features, attaining a 60% prediction accuracy using an SVM. Building on this, ref. [13] analyzed 4858 MLB games from 2019, using both starting and total pitching data. Among 1D convolutional neural networks (1DCNNs), ANNs, and SVMs, ANNs delivered the best results—a 94% accuracy with all pitchers included, versus 92% using only starters. A comparative study by [6] evaluated multiple models before and after feature selection, finding the SVM to be the most accurate at 65.75%.

A major limitation of high-performing models like the ANN and SVM lies in their opacity. Explainable Artificial Intelligence (XAI) addresses this issue by offering tools to make “black-box” models more transparent. According to [17], XAI can be categorized into intrinsic interpretability, model-to-model interpretability, and post hoc interpretability. Its main objectives are to validate behavior, improve performances, and foster human understanding and trust in AI systems [18]. Balancing predictive accuracy with interpretability remains a central challenge. To support a post hoc analysis, this study employs SHAP (SHapley Additive exPlanations), a game-theoretic method that quantifies each feature's

contribution to individual predictions. SHAP enhances both the global and local interpretability by attributing prediction outcomes to specific input features based on Shapley values. This makes it particularly well-suited for decoding complex performance data in sports analytics.

Accordingly, this study investigates the performance outcomes of Chinese Professional Baseball League (CPBL) games by (1) identifying the most accurate among five machine learning models and (2) determining and ranking the most influential features using a SHAP interpretation of the top-performing model.

## 2. Method

The research workflow comprised five sequential stages: data collection, preprocessing, model training, performance evaluation, and interpretation (see Figure 1). Data were collected from the official CPBL website and Baseball Rebas Data Company, yielding 1738 regular-season games from 2021 to 2023. During preprocessing, tied games were excluded to ensure binary outcomes suitable for machine learning applications. Six machine learning algorithms were implemented and trained using 5-fold cross-validation: decision tree, logistic regression, Artificial Neural Network (ANN), Random Forest, LightGBM, and XGBoost. Performance was assessed using seven metrics: accuracy, F1 score, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUC). To enhance interpretability, feature importance analysis and SHapley Additive exPlanations (SHAP) were applied to quantify individual variable contributions to model predictions. This systematic approach ensured transparent model development and comprehensive evaluation.

This study analyzed game data collected from the official CPBL website and the Baseball Rebas Data Company, covering a total of 900 regular-season games from the 2021 to 2023 CPBL seasons. After excluding 31 games that ended in a tie, a final dataset comprising 859 games with definitive outcomes was obtained, resulting in 1738 game records. Table 1 summarizes the key variables utilized in the analysis. This study compiled a comprehensive dataset comprising both batting and pitching statistics from professional baseball games. A total of 21 offensive (batter-related) and 16 defensive (pitcher-related) performance variables were collected, as detailed in Table 1. The binary outcome variable indicating win or loss (W/L) served as the dependent variable for model prediction. All data were obtained from official game records to ensure accuracy, completeness, and consistency across observations.

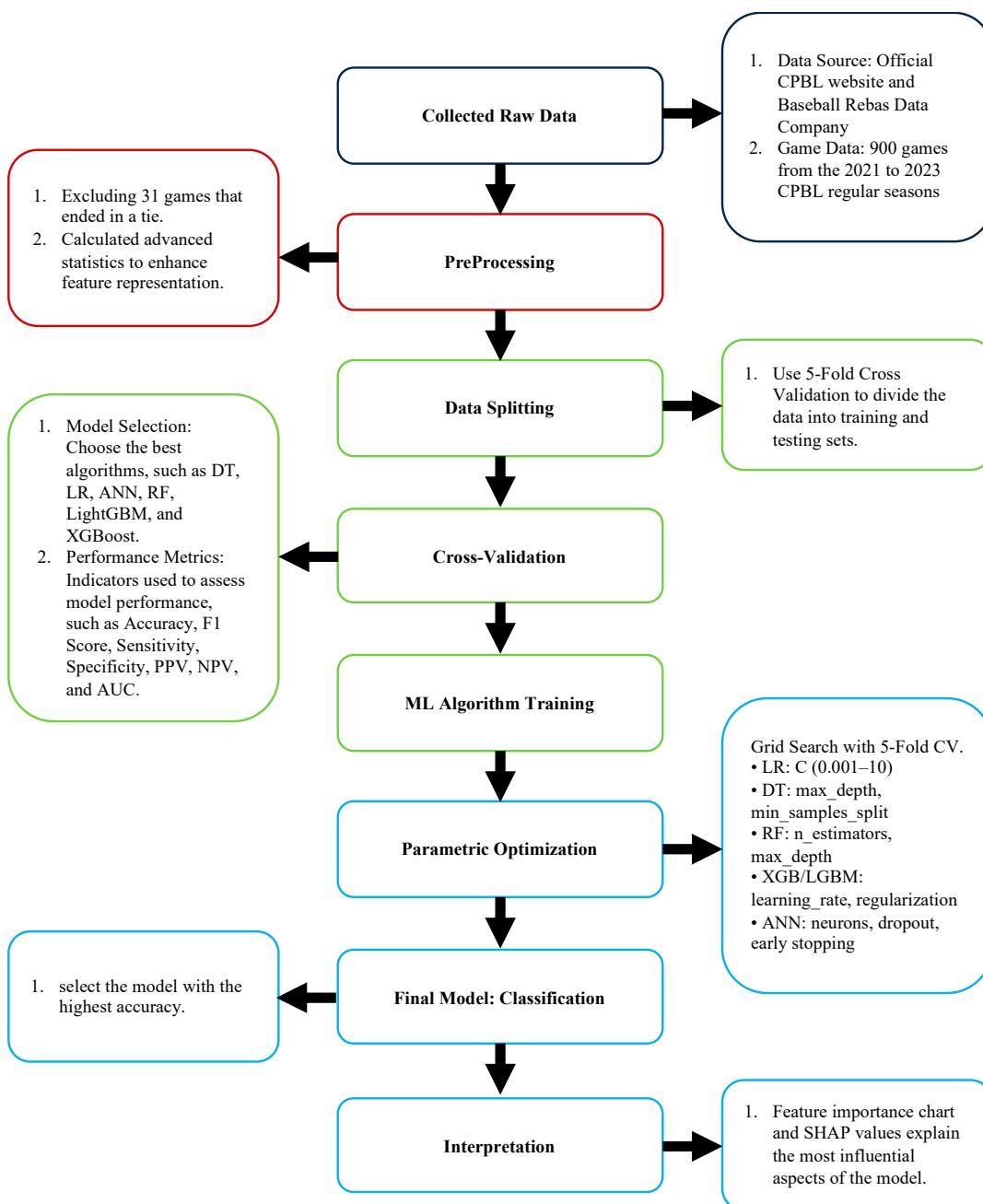
To ensure full transparency and reproducibility, the complete implementation code, including data preprocessing, model training, performance evaluation, and visualization, is publicly available as a Google Colab notebook at <https://colab.research.google.com/drive/1iDYo5dIUSGOva0m1e4t9C1ADef1u2PfI?usp=sharing>. (accessed on 17 June 2025).

**Table 1.** Collection of research variables.

Number	Batter	Abbreviation	Number	Pitcher	Abbreviation
B1	At bats	AB	P1	Batter Faced	BF
B2	Runs	R	P2	Pitch Count	NP
B3	Run batted in	RBI	P3	Strike	Strike
B4	Single	1B	P4	Hits	P_H
B5	Double	2B	P5	Home runs	P_HR
B6	Triple	3B	P6	Base on balls	P_BB
B7	Home runs	HR	P7	Strikeout	P_SO
B8	Hits	H	P8	Walks per nine innings	B/9
B9	Walks	BB	P9	Hits per nine innings	H/9
B10	Strike out	SO	P10	Strikeouts per nine innings	K/9
B11	Sacrifice hits	SH	P11	Strike percentage	S%
B12	Sacrifice flies	SF	P12	Batting average on balls in play	BABIP

**Table 1.** Cont.

Number	Batter	Abbreviation	Number	Pitcher	Abbreviation
B13	Batting average	AVG	P13	Walks plus hits per inning pitched	WHIP
B14	On-base percentage	OBP	P14	Ground ball fly ball ratio	G/F
B15	Slugging percentage	SLG	P15	Pitches per innings	NP/IP
B16	On-base plus slugging	OPS	P16	Left on base percentage	PLOB%
B17	Isolated power	IsoP			
B18	Ground ball fly ball ratio	G/F			
B19	Weighted on-base average	wOBA			
B20	weighted score expectation	wRAA			
B21	Weighted Runs Created Plus	wRC+			
				dependent variable Result W/L	

**Figure 1.** A block diagram of the machine learning-based research workflow.

## 2.1. Machine Learning Methods

In this study, Python3.13 was employed for data preprocessing, model training, performance evaluation, and model interpretability. A five-fold cross-validation strategy was adopted to develop and validate five machine learning models: decision tree, logistic regression, Artificial Neural Network, Random Forest, LightGBM, and XGBoost.

### Decision Tree (DT).

Decision trees are supervised learning models that split data into branches based on feature values, forming a tree-like structure. Each internal node represents a decision based on a feature, while each leaf node corresponds to a predicted outcome. This method is valued for its interpretability and simplicity, though it may overfit the data if not pruned properly. In this study, decision trees are used as a baseline due to their transparency and effectiveness in identifying influential variables, which has been demonstrated in materials classification tasks.

### Logistic Regression (LR).

Logistic regression is a fundamental classification technique used to model the probability of a binary outcome. It employs the logistic function to estimate the likelihood that an input belongs to a certain class, based on a linear combination of features. Despite its simplicity, it remains robust and interpretable, making it a reliable baseline for many classification problems, including those in materials science.

### Artificial Neural Network (ANN).

Artificial Neural Networks (ANNs) are computational models composed of interconnected layers of artificial neurons that process inputs through nonlinear transformations. They are particularly effective in capturing complex and nonlinear relationships in data, especially when sufficient training data are available. ANNs have demonstrated strong predictive performance in various machine learning applications, including those in materials informatics and other domains. In this study, we employed a feedforward neural network architecture for binary classification. The model consists of three layers: an input layer, one hidden layer, and an output layer. The hidden layer comprises 64 neurons with rectified linear unit (ReLU) activation functions, followed by a dropout layer with a dropout rate of 0.3 to mitigate overfitting. The model is trained using the cross-entropy loss function, which is appropriate for binary classification tasks.

### Random Forest (RF).

Random Forest is an ensemble learning method that builds a collection of decision trees on bootstrapped samples of the data and averages their predictions. It introduces additional randomness by selecting a subset of features at each split, which enhances generalization and reduces overfitting. Random Forests offer a balance between accuracy and interpretability and have shown strong performance in materials-related predictive modeling.

### Extreme Gradient Boosting (XGBoost).

XGBoost is a high-performance implementation of gradient boosting that builds decision trees sequentially, where each tree attempts to correct the errors of the previous ones. It incorporates regularization to control overfitting and leverages efficient computation strategies such as parallel processing and tree pruning. XGBoost is well-suited for structured data and has become a standard in high-stakes machine learning applications.

### Light Gradient Boosting Machine (LightGBM)

LightGBM is a gradient boosting framework that uses tree-based learning algorithms optimized for speed and memory efficiency. Unlike traditional boosting methods, it employs histogram-based algorithms and leaf-wise tree growth, which significantly accelerate

training while maintaining accuracy. LightGBM is particularly suitable for large-scale structured data and supports parallel and GPU learning, making it an effective choice for performance-critical applications.

### Hyperparameter Tuning

To enhance model performance and mitigate overfitting, we conducted hyperparameter optimization using grid search in conjunction with five-fold cross-validation. The search space for each model was designed based on domain expertise and commonly recommended values from prior studies (Table 2). For logistic regression, we applied L2 regularization (`penalty = 'l2'`) with inverse regularization strength  $C \in \{0.001, 0.01, 0.1, 1, 10\}$ , using the `lbfgs` solver and `max_iter = 1000`. The decision tree model explored `max_depth`  $\in \{3, 5, 7, 10\}$  and `min_samples_split`  $\in \{2, 5, 10, 20\}$ . The Random Forest classifier was tuned with `n_estimators`  $\in \{100, 200\}$ , `max_depth`  $\in \{5, 10, 15\}$ , and `min_samples_split`  $\in \{2, 10\}$ , with bootstrapping enabled. For XGBoost, the tuning space included `n_estimators`  $\in \{100, 200\}$ , `learning_rate`  $\in \{0.01, 0.1, 0.2\}$ , `max_depth`  $\in \{3, 6, 9\}$ , `subsample` and `colsample_bytree`  $\in \{0.6, 0.8, 1.0\}$ , and regularization parameters `reg_alpha` and `reg_lambda`  $\in \{0, 0.1, 1\}$ . Early stopping was applied based on validation performance. LightGBM adopted a similar tuning scheme with `num_leaves`  $\in \{31, 63\}$ , `max_depth`  $\in \{-1, 5, 10\}$ , and early stopping enabled to prevent overfitting. The Artificial Neural Network (ANN) comprised one hidden layer with 64 neurons using ReLU activation and a dropout rate of 0.3. It was optimized with the Adam optimizer, `learning_rate`  $\in \{0.001, 0.01\}$ , and `batch_size`  $\in \{32, 64\}$  and trained for up to 50 epochs with early stopping (`patience = 10`). All models were evaluated using the same five-fold cross-validation splits to ensure fair and consistent comparison.

**Table 2.** Hyperparameter settings for the six machine learning models used.

Model	Hyperparameters Used
Logistic Regression	<code>penalty = 'l2'</code> , <code>solver = 'lbfgs'</code> , <code>max_iter = 1000</code>
Decision Tree	<code>criterion = 'gini'</code> , <code>max_depth = 5</code> , <code>min_samples_split = 10</code>
Random Forest	<code>n_estimators = 100</code> , <code>max_depth = 10</code> , <code>min_samples_split = 10</code> , <code>bootstrap = True</code>
XGBoost	<code>n_estimators = 100</code> , <code>max_depth = 6</code> , <code>learning_rate = 0.1</code> , <code>subsample = 0.8</code> , <code>colsample_bytree = 0.8</code>
LightGBM	<code>n_estimators = 100</code> , <code>learning_rate = 0.1</code> , <code>num_leaves = 31</code> , <code>max_depth = -1</code>
ANN	1 hidden layer (64 units), ReLU activation, dropout = 0.3, optimizer = Adam, epochs = 50

### 2.2. Evaluated Machine Learning Model

The models' performance was evaluated using six key metrics: accuracy, F1 score, sensitivity, specificity, positive predictive value, negative predictive value, and the area under the receiver operating characteristic curve. The AUC-ROC (area under the receiver operating characteristic curve) is a key metric for evaluating the overall performance of a classification model. The performance of the machine learning models in this study was evaluated using six key metrics, each providing unique insights into the model's predictive capabilities. The formulas for these metrics are as follows:

Accuracy—Measures the overall correctness of the model's predictions:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}} \quad (1)$$

F1 score—Represents the harmonic mean of precision and recall, balancing false positives and false negatives:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

Sensitivity (recall or true positive rate, TPR)—Measures the model's ability to correctly identify positive cases:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

Specificity (true negative rate, TNR)—Measures the model's ability to correctly identify negative cases:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

Positive predictive value (PPV)—Indicates the proportion of predicted positive cases that are actually positive:

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

Negative predictive value (NPV)—Indicates the proportion of predicted negative cases that are actually negative:

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (6)$$

The area under the ROC curve (AUC-ROC) is a key metric for evaluating the discriminative performance of classification models. Values range from 0.5 (no discrimination) to 1.0 (perfect discrimination), with thresholds commonly interpreted as excellent (0.90–1.00), good (0.80–0.90), fair (0.70–0.80), poor (0.60–0.70), and failing (<0.60). AUC is especially informative in cases of class imbalance or varying misclassification costs.

In sports analytics, particularly baseball prediction models, a higher AUC indicates stronger predictive accuracy in distinguishing between wins and losses. Models with AUC near 0.5 lack discriminative power and require further refinement. The best-performing model was further analyzed using SHAP values and feature importance plots to enhance interpretability. Statistical analyses were conducted using Python to identify key performance indicators that influence game outcomes, offering strategic insights to improve competitive performance.

### 3. Result

Results presented in Table 3 indicated that the logistic regression and XGBoost consistently outperformed the other models in classification tasks. These models achieved high accuracy (0.89–0.93) and AUC values (0.97–0.98), reflecting an excellent discriminatory ability. Furthermore, they demonstrated well-balanced precision (0.90–0.94), recall (0.89–0.94), and F1 scores (0.90–0.92), indicating a reliable performance in identifying true positives while minimizing false positives. Their high NPVs (0.88–0.92) further underscored their robustness in reducing false negatives, a critical consideration in decision-making under uncertainty.

The Random Forest also exhibited stable and competitive performance across all evaluation criteria, with accuracy and AUC values close to those of the logistic regression and XGBoost. While its F1 scores were marginally lower, it offered a practical trade-off between predictive performance and interpretability. The ANN model demonstrated the highest sensitivity (up to 0.96), surpassing all other models in identifying wins (positive class). However, this came at the cost of a lower specificity (as low as 0.75) and precision, suggesting a tendency to overpredict positive outcomes. Additionally, the ANN incurred the longest average execution time (46.8 s), potentially limiting its practicality in real-time or resource-constrained environments.

**Table 3.** Comparison of machine learning models for baseball data prediction.

Model	Accuracy	F1 Score	Sensitivity (Recall)	Specificity	PPV (Precision)	NPV	ROC AUC	Running Time
Decision Tree	0.86	0.86	0.83	0.90	0.89	0.84	0.86	0.2
	0.84	0.84	0.82	0.86	0.86	0.83	0.84	
	0.85	0.85	0.82	0.89	0.88	0.83	0.85	
	0.86	0.86	0.90	0.82	0.83	0.89	0.86	
	0.85	0.86	0.89	0.82	0.83	0.88	0.85	
Mean	0.85	0.85	0.85	0.86	0.86	0.85	0.85	0.20
Logistic Regression	0.91	0.90	0.90	0.91	0.91	0.90	0.97	1.9
	0.91	0.91	0.89	0.93	0.92	0.89	0.97	
	0.93	<b>0.93</b>	0.92	0.94	0.94	0.92	<b>0.98</b>	
	0.93	<b>0.93</b>	0.94	0.91	0.92	0.94	<b>0.98</b>	
	0.89	0.89	0.87	0.90	0.90	0.88	0.97	
Mean	0.91	0.91	0.90	0.92	0.92	0.91	0.97	1.90
Neural Network	0.89	0.90	<b>0.96</b>	0.83	0.85	0.95	0.96	46.8
	0.89	0.88	0.86	0.91	0.91	0.86	0.94	
	0.86	0.85	0.79	0.94	0.93	0.82	0.96	
	0.88	0.88	0.86	0.90	0.90	0.86	0.96	
	0.83	0.84	0.91	0.75	0.78	0.90	0.94	
Mean	0.87	0.87	0.88	0.87	0.87	0.88	0.95	46.80
XGBoost	0.92	0.92	0.89	<b>0.95</b>	<b>0.95</b>	0.90	0.97	0.9
	0.91	0.91	0.91	0.91	0.91	0.91	0.97	
	0.90	0.90	0.89	0.91	0.91	0.89	<b>0.98</b>	
	0.92	0.92	0.95	0.88	0.89	<b>0.95</b>	<b>0.98</b>	
	0.88	0.88	0.87	0.88	0.88	0.87	0.95	
Mean	0.91	0.91	0.90	0.91	0.91	0.90	0.97	0.90
Random Forest	0.90	0.89	0.87	0.92	0.92	0.88	0.97	2.4
	0.89	0.89	0.88	0.91	0.91	0.88	0.96	
	0.90	0.90	0.89	0.92	0.92	0.89	0.97	
	0.90	0.91	0.91	0.90	0.90	0.91	0.97	
	0.89	0.89	0.90	0.88	0.89	0.89	0.95	
Mean	0.90	0.90	0.89	0.91	0.91	0.89	0.96	2.40
Light GBM	0.91	0.90	0.88	0.93	0.93	0.89	0.97	1.0
	0.91	0.91	0.90	0.92	0.92	0.90	0.97	
	0.90	0.90	0.90	0.91	0.91	0.90	0.97	
	0.91	0.91	0.93	0.89	0.90	0.93	0.98	
	0.87	0.87	0.89	0.84	0.85	0.89	0.95	
Mean	0.90	0.90	0.90	0.90	0.90	0.90	0.97	1.00

P.S: Bold values in the table highlight the best performance achieved among the compared models for each metric.

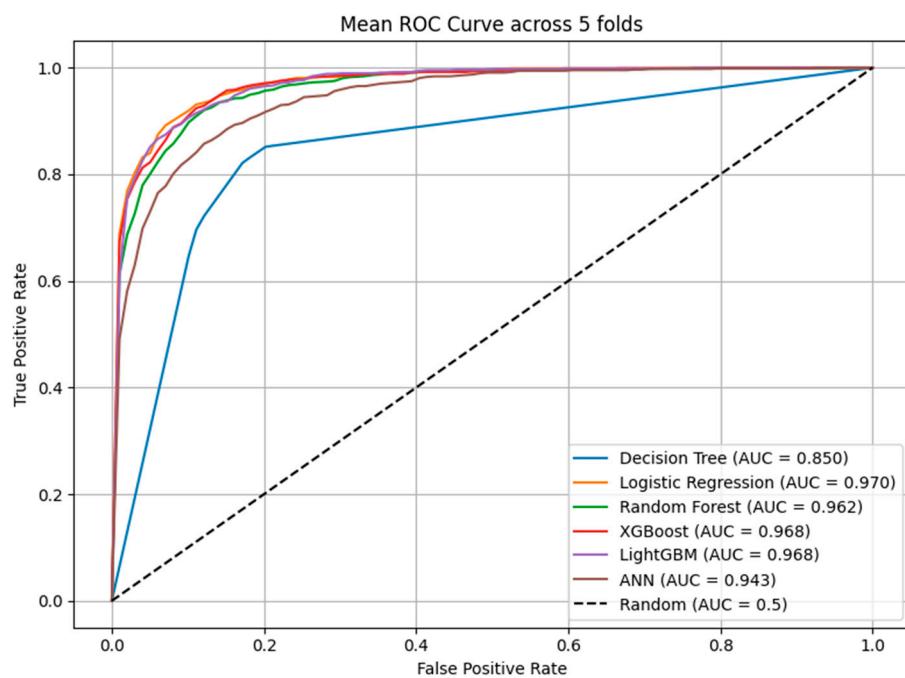
The decision tree model yielded the lowest overall performance, with AUC values near 0.85 and more variability in accuracy (0.84–0.86). Nonetheless, its minimal computation time (0.2 s) and high interpretability position it as a viable option when efficiency and transparency are prioritized over predictive precision.

Differences in model performance can be attributed to each algorithm's ability to capture nonlinear relationships and interactions among features. Gradient boosting models such as XGBoost and LightGBM are particularly well-suited for modeling such complexities, which likely accounts for their superior results. In contrast, simpler models like decision trees are less equipped to handle such intricacies, resulting in reduced accuracy.

In summary, logistic regression and XGBoost emerged as the most effective models for predicting CPBL game outcomes, offering a strong balance between predictive accuracy, model stability, and practical utility. These findings support their application in sports analytics and real-time decision-making in professional baseball settings.

Figure 2 presents the ROC curves for all six machine learning models, comparing their classification performance for game outcome predictions. The ROC curve plots the true

positive rate against the false positive rate at various threshold settings, with the area under the curve (AUC) serving as the primary performance metric.



**Figure 2.** AUC in different models.

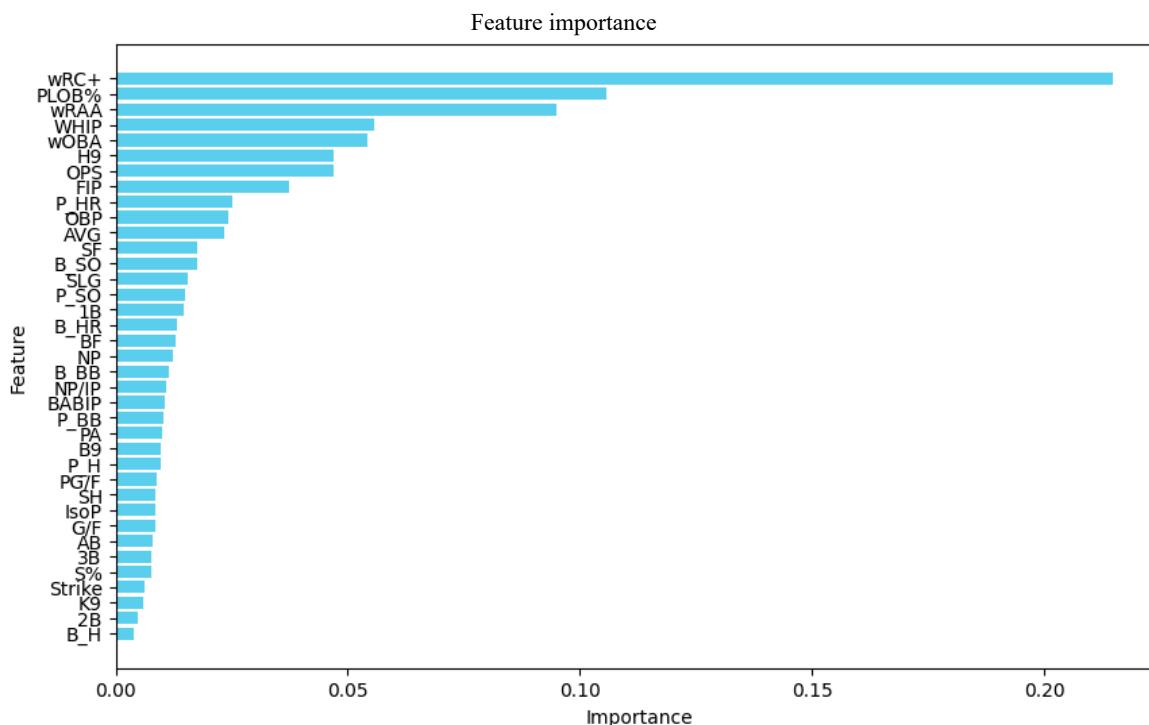
Logistic regression achieved the highest AUC (0.970), demonstrating an exceptional discriminative ability. XGBoost and LightGBM followed closely with AUC values of 0.968, reflecting the strong performance of gradient boosting algorithms in capturing complex nonlinear relationships. Random Forest exhibited a competitive performance (AUC = 0.962), confirming the effectiveness of ensemble methods. The Artificial Neural Network achieved an AUC of 0.943, indicating a solid but comparatively lower performance. The decision tree yielded the lowest AUC (0.850), suggesting a limited predictive capability relative to other models.

All models substantially outperformed random chance (represented by the diagonal dashed line, AUC = 0.5), confirming their practical utility for classification tasks. The results highlight the superior performance of ensemble and boosting methods for this particular prediction problem.

#### Feature Importance Analysis

Figure 3 displays the feature importance rankings derived from the XGBoost model, with corresponding values presented in the accompanying table (Table 4). The most influential predictor is wRC+ (Weighted Runs Created Plus), with an importance value of 0.21, demonstrating its critical role in evaluating offensive contributions. The PLOB% (percentage of left-on-base runners) and wRAA (Weighted Runs Above Average) rank among the top features, reflecting their importance in measuring team efficiency in run production and prevention. Pitching metrics constitute several high-ranking features, including the WHIP (walks + hits per inning pitched), FIP (fielding independent pitching), H9 (hits per nine innings), and P\_HR (percentage of home runs allowed). These results underscore the significance of limiting baserunners, hits, and home runs for game success. Traditional batting statistics—OPS (on-base plus slugging), OBP (on-base percentage), and AVG (batting average)—also demonstrate a considerable influence on predictions, confirming their continued relevance in modern baseball analytics. Lower-ranked features such as Strike Percentage (S%), K9 (strikeouts per nine innings), and B\_H (total hits by

batters) showed a relatively minimal impact on model predictions. The analysis reveals that game outcomes are primarily driven by offensive efficiency metrics and pitching control variables, highlighting the fundamental importance of run creation and prevention in determining baseball game results.

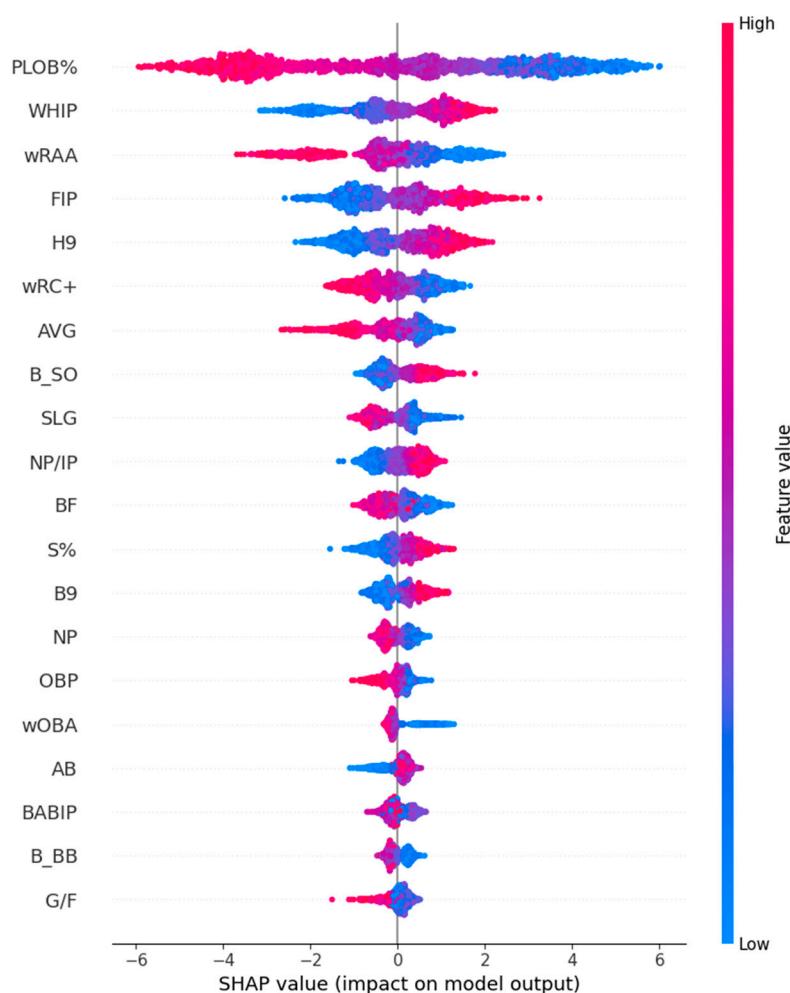


**Figure 3.** Feature importance analysis of XGBoost model.

**Table 4.** Feature importance table.

Rank	Feature	Importance	Rank	Feature	Importance
1	wRC+	0.21	20	B_BB	0.01
2	PLOB%	0.11	21	NP/IP	0.01
3	wRAA	0.10	22	BABIP	0.01
4	WHIP	0.06	23	P_BB	0.01
5	wOBA	0.05	24	PA	0.01
6	H9	0.05	25	B9	0.01
7	OPS	0.05	26	P_H	0.01
8	FIP	0.04	27	PG/F	0.01
9	P_HR	0.03	28	SH	0.01
10	OBP	0.02	29	IsoP	0.01
11	Avg	0.02	30	G/F	0.01
12	SF	0.02	31	AB	0.01
13	B_SO	0.02	32	3B	0.01
14	SLG	0.02	33	S%	0.01
15	P_SO	0.01	34	Strike	0.01
16	1B	0.01	35	K9	0.01
17	B_HR	0.01	36	2B	0.00
18	BF	0.01	37	B_H	0.00
19	NP	0.01			

Figure 4 presents the SHAP summary plot, visualizing feature contributions to XGBoost model predictions for game outcomes. Each dot represents an individual game instance, with the color indicating the feature magnitude (red for higher values, blue for lower values). The x-axis displays SHAP values, quantifying each feature's contribution to predictions, where positive values increase the win probability and negative values decrease it.



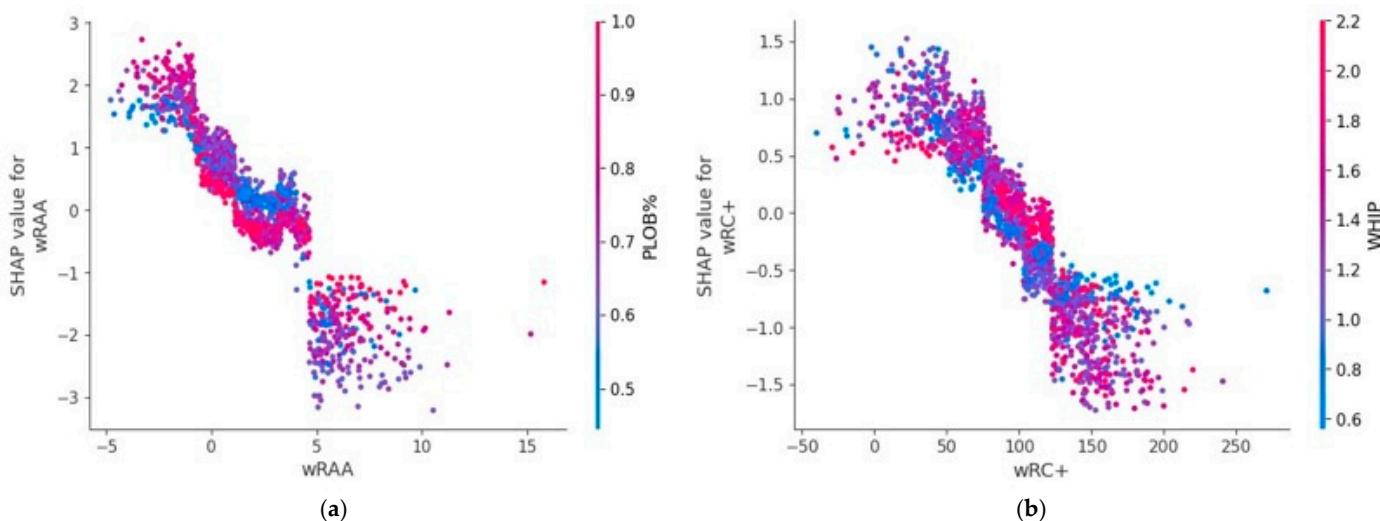
**Figure 4.** The SHAP summary plot: the feature importance and impact on the model output. Point colors represent feature values (red: high, blue: low), while the  $x$ -axis shows SHAP values indicating the direction and magnitude of each feature's effect. The spread of points reflects the variability of the feature influence across samples.

The PLOB% (left on base percentage) demonstrates the strongest influence, with higher values (red dots) generally producing positive SHAP contributions. The WHIP (walks and hits per inning pitched) and wRAA show substantial predictive significance, with a clear separation between high and low values driving predictions in opposite directions.

The traditional metrics AVG and wRC+ emerge as influential predictors, where a superior performance typically generates positive SHAP values, confirming their positive impact on game outcomes. Conversely, features such as the B\_BB (batting walks) and G/F (ground ball to fly ball ratio) exhibit minimal contributions, with SHAP values clustering near zero.

This analysis confirms the relative importance of key performance metrics in determining game outcomes, providing quantitative insights into the statistical factors that drive model predictions.

Figure 5a,b provide insights into the interaction effects between offensive and pitching metrics using SHAP dependence plots. Specifically, these visualizations illustrate how key offensive indicators (wRAA and wRC+) interact with pitching-related contextual features (PLOB% and WHIP, respectively) to influence the model's prediction of game outcomes.



**Figure 5.** (a) A SHAP dependence plot of the interaction between the wRAA and PLOB%. (b) A SHAP dependence plot of the interaction between the wRC+ and WHIP.

Figure 5a plots the SHAP values of wRAA (Weighted Runs Above Average) against its original values, with color encoding the PLOB% (Percentage of Left On Base). A strong negative association is observed between the wRAA and its SHAP value, suggesting diminishing marginal returns of offensive production on the predicted win probability. Notably, data points with a higher PLOB% (depicted in red) tend to have less negative SHAP values for wRAA, indicating that effective pitching (i.e., the ability to strand baserunners) can moderate or amplify the impact of offensive contributions. This interaction highlights that the offensive value alone is insufficient without the corresponding pitching performance.

Figure 5b similarly depicts the relationship between wRC+ (Weighted Runs Created Plus) and its SHAP values, with WHIP (Walks plus Hits per Inning Pitched) as the interaction feature. A comparable negative trend is present, where increased wRC+ correlates with lower marginal model contributions after a certain threshold. However, observations with lower WHIP values (indicated by blue) are associated with higher SHAP values for wRC+, reinforcing the idea that offensive efficiency is more predictive of winning outcomes when supported by disciplined and effective pitching.

Taken together, these results emphasize the interdependent nature of batting and pitching metrics in driving model predictions. The SHAP interaction analysis reveals that the explanatory power of offensive metrics is contextually modulated by the strength of the pitching staff, thus supporting a more holistic interpretation of baseball performance analytics within explainable machine learning frameworks.

#### 4. Discussion

This study proposes a research objective of identifying the best predictive model through the indicators of multiple evaluation models and then using the best model to determine the factors that influence competition outcomes. The accuracy, sensitivity, specificity, positive predictive value, and negative predictive value are all optional validation metrics for evaluating model performance. However, accuracy is typically used as the primary evaluation metric. The higher the accuracy, the greater the model's reliability. When predicting the outcomes of sports competitions, data from more than three years of matches is often used to ensure a sufficient training volume, thereby reducing prediction errors. This study utilized match data from 2021 to 2023 and found that both XGBoost and Random Forest achieved up to 95% prediction accuracies when using advanced metrics. XGBoost achieves a high accuracy through improvements to decision trees, where each

tree is interrelated, making it highly effective for predictions. On the other hand, Random Forest combines multiple decision trees for classification, and its predictive performance is not affected by the increased variability, making both models highly accurate [17]. Previous studies have shown that commonly used Artificial Neural Networks achieve an accuracy of about 75% [19–23]. Decision trees and support vector machines (SVMs) have been reported to achieve prediction accuracies of over 85% [24–26].

Recent research also supports the superior performance of XGBoost over neural networks in structured data settings. Ref. [27] compared XGBoost and multi-layer perceptron neural networks across five class-balanced datasets and found that XGBoost consistently outperformed neural networks in both accuracy and F1 scores. Particularly in datasets with overlapping or imperfect data distributions, XGBoost demonstrated a better generalization. The authors attributed this to XGBoost's ability to construct effective feature subspaces using tree-based ensembles, while neural networks struggled to optimize in high-dimensional discrete spaces. This finding aligns with our current results and offers empirical support for using ensemble methods like XGBoost in baseball outcome prediction tasks.

The first factor influencing game outcomes is wRC+. This advanced metric is derived from the wOBA (weighted on-base average), incorporating adjustments for park factors and league averages. The advantage of wRC+ lies in its ability to compare players across leagues and eras [28–30]. Previous studies have explored changes in wRC+ using Statcast batted ball data. The findings revealed that the relationship between the exit velocity and launch angle significantly impacts wRC+. Moreover, 70% of the data from Statcast influences changes in wRC+. This highlights that a batter's performance in terms of the exit velocity and launch angle are both critical factors affecting run production [31]. The impact of wRC+ on a team's success is significant because it quantifies a player's ability to generate runs, the fundamental factor in winning games. A higher wRC+ indicates that a player contributes more effectively to run production, leading to increased scoring opportunities.

The wRAA represents the expected run contribution, indicating how many more runs a hitter can generate for the team compared to the league average. This advanced metric has an impact of 7.7% on game outcomes. However, its value is influenced by the number of plate appearances. Since each time a batter reaches a base, the offensive team gains an additional opportunity to attack, it directly increases the chances of scoring runs. Therefore, during game strategy planning, coaches must focus on creating more plate appearances for their players. This can be achieved by improving the plate discipline to increase walk opportunities, implementing more aggressive offensive strategies, and reducing passive bunt tactics. By effectively increasing plate appearances, teams can maximize their expected run contributions, thereby enhancing their chances of winning. These four metrics are considered highly significant, each with an importance level exceeding 5%.

The wOBA is widely recognized as the best metric for evaluating a player's offensive performance. It accurately assigns linear weights to different offensive outcomes to determine their true value [32]. Among the factors influencing game outcomes, the wOBA accounts for 9.9%. Previous studies have suggested that the on-base percentage (OBP) is a critical determinant of scoring and winning, with its impact on runs being 2.33 times greater than the slugging percentage (SLG) [33].

However, OPS, which simply sums the OBP and SLG, fails to reflect the individual contributions of different offensive events. To address this issue, ref. [34] developed the wOBA by incorporating linear weights and league adjustment coefficients, ensuring that the statistic better captures the relative importance of different hitting outcomes. This metric effectively evaluates both slugging ability and on-base skills, making it a more comprehensive indicator of offensive performance.

Ref. [35] found that coaches should prioritize the wOBA as a core classification variable when assessing hitters. This statistic is not only a key characteristic of a hitter's contribution to team victories but also a reasonable indicator for organizations to evaluate a player's market value. To maximize the wOBA, hitters should focus on increasing the exit velocity, as research indicates that exit velocity provides a greater information gain than the launch angle. While the launch angle is relevant, exit velocity plays a more significant role in offensive success. Maintaining a high wOBA requires a combination of factors, such as increasing the exit velocity to improve contact quality, enhancing contact rates to reduce strikeouts, and avoiding slow baserunning that could limit extra-base hits, and minimizing swing-and-miss rates to ensure a consistent offensive output [36]. Another effective way to improve the wOBA is by increasing the number of plate appearances, which allows a player to accumulate more valuable offensive events and enhance their overall weighted on-base percentage.

By integrating on-base and slugging percentages, the wOBA assigns the appropriate value to each type of hit and on-base event, making it a more precise measure of offensive production than the OBP alone. Previous research has demonstrated a strong positive correlation between the wOBA and team runs scored in the Korea Baseball Organization (KBO) [36]. Furthermore, ref. [37] found that teams with high-wOBA catchers, first basemen, second basemen, and left fielders were more likely to reach the postseason. This suggests that the wOBA is not only a key performance indicator but also a valuable metric for player selection and team-building strategies. The LOB% refers to the percentage of baserunners that a pitcher allows but successfully prevents from scoring, calculated as the ratio of stranded runners to total baserunners (excluding home runs). A pitcher who effectively limits runners left on base demonstrates the ability to escape high-pressure situations and prevent opponents from capitalizing on scoring opportunities. Teams with elite pitchers often force the opposing offense to leave runners stranded, increasing pressure on the attacking team and limiting their ability to score runs.

Also, stranded runners' presence creates psychological pressure on the offensive team when they transition to defense, as they become more conscious of preventing further runs. Consequently, the LOB% is a crucial factor influencing game outcomes. A study on American college pitchers found that lower cognitive anxiety is associated with a higher PLOB% (pitcher's left on base percentage) [38]. This suggests that a pitcher's ability to suppress runs is linked not only to skill and experience but also to mental resilience. Effective pitchers who prevent runners from scoring also tend to have lower ERAs, while pitchers with a LOB% below the league average may be more affected by luck rather than skill. When faced with high-pressure situations, pitchers with strong mental fortitude and experience are more likely to escape jams, turning the pressure back onto the opposing team. If the offensive team fails to score, they may become overly focused on preventing runs while on defense, which can inadvertently lead to defensive mistakes and allow runs.

Therefore, a pitcher's ability to maintain a high LOB% is a critical determinant of success in competitive baseball. FIP is a crucial metric for evaluating a pitcher's performance, as it eliminates defensive factors and focuses solely on strikeouts, walks, and home runs—elements directly controlled by the pitcher. In this study's machine learning analysis, FIP was identified as a key variable influencing game outcomes, demonstrating its high predictive value for pitcher performance. Pitchers with lower FIP tend to deliver more consistent outings, reduce scoring risks, and ultimately contribute to higher team winning percentages [39]. Additionally, FIP complements other metrics such as the WHIP and PLOB%, providing a more comprehensive assessment of pitching ability. These findings further validate the predictive power of FIP in baseball analytics, and future research could

incorporate additional game contexts and deep learning techniques to enhance the accuracy of pitcher performance evaluations.

The SHAP analysis provided a clear interpretability framework, revealing that the wRC+ and PLOB% had the largest marginal contribution to the model's output. This implies that offensive productivity and the ability to suppress scoring opportunities are the two most decisive dimensions in predicting game outcomes. For coaches and analysts, this insight suggests that optimizing lineups based on high-wRC+ players, and strategically managing pitchers with a high PLOB% under pressure, may lead to significantly improved game outcomes.

This study utilizes machine learning methods to analyze baseball data and compares different models' predictive performance to evaluate the impact of key statistical indicators on game outcomes. The results show that XGBoost and logistic regression perform the best across multiple evaluation metrics, with an average accuracy exceeding 0.91 and an ROC AUC above 0.97, demonstrating their superior predictive capability. In contrast, while the XGBoost model offers high interpretability, its overall predictive performance is relatively lower. Additionally, the Neural Network model exhibits greater fluctuations in some evaluation metrics, indicating that its stability requires further improvement.

Further analysis of feature importance reveals that wRC+ is the most influential variable in the XGBoost model, emphasizing its crucial role in assessing a hitter's offensive capability. Consistent with the SHAP value analysis, the PLOB%, wRAA, and WHIP are also highly influential variables, highlighting their importance in predicting player performance. The SHAP values further indicate that pitching-related metrics, such as the PLOB%, WHIP, wRAA, and FIP, have a significant impact on model predictions, reinforcing their role in evaluating pitcher performance. Meanwhile, hitter-related statistics such as the wRC+ and wOBA also exhibit strong explanatory power regarding game outcomes, further validating the importance of these advanced hitting metrics.

This study addresses these gaps by (1) comparing multiple machine learning models—including XGBoost, Random Forest, and logistic regression—across various evaluation metrics; (2) applying SHAP values to identify key offensive and pitching indicators that influence game outcomes; and (3) validating the critical role of advanced metrics such as the wRC+, wRAA, wOBA, and LOB% in predicting performance. The key contributions of this research lie in integrating model interpretability with predictive performance to enhance practical applicability in professional baseball analytics. These contributions not only advance methodological approaches but also support data-informed decision-making in player development and game strategy.

## 5. Limitations

Despite promising results, this study has several limitations. First, the model performance depends heavily on data quality and completeness. Missing or noisy features can reduce both accuracy and interpretability, particularly for complex models like neural networks and XGBoost. Second, ensemble models, while more accurate, are often less transparent, which may limit their use in coaching contexts where explainability is important. Third, since the models were trained solely on CPBL data, their generalizability to other leagues is uncertain without retraining or transfer learning. Fourth, although cross-validation was applied to control overfitting, the lack of external validation reduces the robustness of the findings. For future research we suggest (1) incorporating real-time game or biometric data to improve feature quality; (2) using time-series or recurrent models to capture performance trends; and (3) conducting cross-league or cross-season evaluations to test model adaptability. Additionally, although this study focused on the main effects using SHAP, it did not explore feature interactions or temporal dependencies. Future

studies may address this by applying SHAP interaction values or tools like PDP and ICE plots to uncover non-additive relationships.

**Author Contributions:** Methodology, Y.-K.L.; Validation, C.-H.C.; Formal analysis, C.-L.C.; Data curation, C.-L.C.; Writing—original draft, T.-C.L.; Writing—review & editing, C.-Y.L. and Y.-K.L.; Visualization, Y.-K.L.; Supervision, T.-Y.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Sports Administration, Ministry of Education, Taiwan, under the 2024 Sports Science Project for Competitive Sports (Project title: “The effect of constraints led approach and variable practice effect in pitchers’ skills and Pitching Command”).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of National Taiwan Sport University (protocol code NTSUIRB-113-028, approved in [April 2025]).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author due to privacy or ethical restrictions.

**Acknowledgments:** The authors thank Rebas Sports Inc. for providing data.

**Conflicts of Interest:** The author declares that there are no conflicts of interest.

## References

1. Rodríguez-Mazahua, L.; Rodríguez-Enríquez, C.A.; Sánchez-Cervantes, J.L.; Cervantes, J.; García-Alcaraz, J.L.; Alor-Hernández, G. A general perspective of Big Data: Applications, tools, challenges and trends. *J. Supercomput.* **2016**, *72*, 3073–3113. [[CrossRef](#)]
2. Namatevs, I.; Aleksejeva, L.; Polaka, I. Neural network modelling for sports performance classification as a complex socio-technical system. *Inf. Technol. Manag. Sci.* **2016**, *19*, 45–52. [[CrossRef](#)]
3. Rein, R.; Memmert, D. Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus* **2016**, *5*, 1410–1434. [[CrossRef](#)] [[PubMed](#)]
4. Horvat, T.; Job, J. The use of machine learning in sport outcome prediction: A review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1380. [[CrossRef](#)]
5. Bunker, R.; Susnjak, T. The application of machine learning techniques for predicting match results in team sport: A review. *J. Artif. Intell. Res.* **2022**, *73*, 1285–1322. [[CrossRef](#)]
6. Ouyang, Y.; Li, X.; Zhou, W.; Hong, W.; Zheng, W.; Qi, F.; Peng, L. Integration of machine learning XGBoost and SHAP models for NBA game outcome prediction and quantitative analysis methodology. *PLoS ONE* **2024**, *19*, e0307478. [[CrossRef](#)]
7. Leicht, A.S.; Gomez, M.A.; Woods, C.T. Team performance indicators explain outcome during women’s basketball matches at the Olympic games. *Sports* **2017**, *5*, 96. [[CrossRef](#)]
8. Fahey-Gilmour, J.; Dawson, B.; Peeling, P.; Heasman, J.; Rogalski, B. Multifactorial analysis of factors influencing elite Australian football match outcomes: A machine learning approach. *Int. J. Comput. Sci. Sport* **2019**, *18*, 100–124. [[CrossRef](#)]
9. Chen, C.-W.; Lee, T.-S.; Liang, Y.-C. Construction of the Winner Predictive Model in Major League Baseball Games: Use of the Artificial Neural Networks. *Sports Exerc. Res.* **2014**, *16*, 167–181. [[CrossRef](#)]
10. Cao, C. Sports Data Mining Technology Used in Basketball Outcome Prediction. Master’s Thesis, Dublin Institute of Technology, Dublin, Ireland, 2012. Available online: <https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1040&context=scschcomdis> (accessed on 11 April 2025).
11. Li, T.-H.; Li, C.-M. Using Artificial Neural Networks to Construct the Winner Prediction Model in Second Grade of Open-Women University Basketball Association of Taiwan. *Rev. Leis. Sport Health* **2015**, *6*, 2–14.
12. Ivanković, Z.; Racković, M.; Markoski, B.; Radosav, D.; Ivković, M. Analysis of basketball games using neural networks. In Proceedings of the 2010 11th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, 18–20 November 2010; IEEE: Piscataway, NJ, USA; pp. 251–256.
13. Huang, M.L.; Li, Y.Z. Use of machine learning and deep learning to predict the outcomes of major league baseball matches. *Appl. Sci.* **2021**, *11*, 4499. [[CrossRef](#)]

14. Lin, Y.-L.; Lin, Y.-S. Data Mining to Explore the Key Performance Factors in the National Basketball Association. *J. Tour. Leis. Manag.* **2014**, *2*, 95–103.
15. Tolbert, B.; Trafalis, T. Predicting major league baseball championship winners through data mining. *Athens J. Sports* **2016**, *3*, 239–252. [CrossRef]
16. Valero, C.S. Predicting Win-Loss outcomes in MLB regular season games—A comparative study using data mining methods. *Int. J. Comput. Sci. Sport* **2016**, *15*, 91–112. [CrossRef]
17. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable ai: A review of machine learning interpretability methods. *Entropy* **2020**, *23*, 18. [CrossRef] [PubMed]
18. Samek, W.; Wiegand, T.; Müller, K.R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, arXiv:1708.08296.
19. Chen, P.-C. Comparison of RandomForest and XGBoost in Forecasting Robustness. Ph.D. Thesis, Master’s Program in Applied Statistics, Department of Statistics, Tamkang University, New Taipei City, Taiwan, 2022.
20. Hubáček, O.; Šourek, G.; Železný, F. Exploiting sports–betting market using machine learning. *Int. J. Forecast.* **2019**, *35*, 783–796. [CrossRef]
21. Kahn, J. *Neural Network Prediction of NFL Football Games*; IEEE Computer Society: Washington, DC, USA, 2003; pp. 1194–1197. Available online: [https://www.researchgate.net/publication/266334198\\_Neural\\_Network\\_Prediction\\_of\\_NFL\\_Football\\_Games](https://www.researchgate.net/publication/266334198_Neural_Network_Prediction_of_NFL_Football_Games) (accessed on 30 March 2025).
22. Loeffelholz, B.; Bednar, E.; Bauer, K.W. Predicting NBA games using neural networks. *J. Quant. Anal. Sports* **2009**, *5*, 1–17. [CrossRef]
23. Purucker, M.C. Neural network quarterbacking. *IEEE Potentials* **1996**, *15*, 9–15. [CrossRef]
24. Delen, D.; Cogdell, D.; Kasap, N. A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *Int. J. Forecast.* **2012**, *28*, 543–552. [CrossRef]
25. Mustafa, R.U.; Nawaz, S.M.; Ikram Ullah Lali, M.; Zia, T.; Mehmood, W. Predicting the cricket match outcome using crowd opinions on social networks: A comparative study of machine learning method. *Malays. J. Comput. Sci.* **2017**, *30*, 63–76. [CrossRef]
26. Pai, P.F.; ChangLiao, L.H.; Lin, K.P. Analyzing basketball games by a support vector machines with Decision Tree model. *Neural Comput. Appl.* **2017**, *28*, 4159–4167. [CrossRef]
27. Wu, J.; Li, Y.; Ma, Y. Comparison of XGBoost and the Neural Network model on the class-balanced datasets. In Proceedings of the 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC), Greenville, SC, USA, 12–14 November 2021; pp. 457–461.
28. McShane, B.; Braunstein, A.; Piette, J.; Jensen, T. A hierarchical Bayesian variable selection approach to Major League Baseball hitting metrics. *J. Quant. Anal. Sports* **2011**, *7*, 2. [CrossRef]
29. Sanford, G.D. *What Raw Statistics Have the Greatest Effect on wRC+ in Major League Baseball in 2017*; University of Minnesota Digital Conservancy: Duluth, MN, USA, 2018. Available online: <http://hdl.handle.net/11299/199934> (accessed on 14 April 2025).
30. Tripin, J. An Idiot’s Guide to Advanced Statistics: wOBA and wRC+. 2017. Available online: <https://www.lookoutlanding.com/2017/3/7/14783982/an%E2%80%93idiots%E2%80%93guide%E2%80%93to%E2%80%93advanced%E2%80%93statistics%E2%80%93woba%E2%80%93and%E2%80%93wrc%E2%80%93sabermetrics> (accessed on 14 April 2025).
31. Kupiec, R. Can Statcast Variables Explain the Variation in Weighted Runs Created Plus? Master’s Thesis, DePauw University, Greencastle, IN, USA, 2020.
32. Weinberg, N. How to Evaluate a Hitter, Sabermetrically [Internet]. Available online: <https://www.beyondtheboxscore.com/2014/5/26/5743956/sabermetrics%E2%80%93stats%E2%80%93offense%E2%80%93learn%E2%80%93sabermetrics> (accessed on 28 March 2025).
33. Bradbury, J.C. *The Baseball Economist: The Real Game Exposed*; Dutton: New York, NY, USA, 2007. [CrossRef]
34. Tango, T.M.; Lichtman, M.G.; Dolphin, A.E. *The Book: Playing the Percentages in Baseball*; Potomac Books, Inc.: Williamsport, MD, USA, 2007.
35. Chiu, Y.-C. Machine Learning Applications for Baseball Coaching Decision-Making: Case Study of Utilizing the Batter-Pitcher Matchups and Forecasting the Hitters’ Value. Ph.D. Thesis, The In-Service Master’s Program in Multimedia Network Communication and Digital Learning, National Ilan University, Yilan, Taiwan, 2022.
36. Detweiler, H. Success in Professional Baseball: The Value of Above Average Position Players; 2014. Available online: <https://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=1444&context=honors> (accessed on 14 April 2025).
37. Lou, K.; Barnicle, S.; Zizzi, S.; Follmer, D.J. The Moderating Effect of Hardiness on the Relationship Between Trait Anxiety and Performance in Collegiate Baseball Players. *J. Adv. Sport Psychol. Res.* **2022**, *2*, 4–15. [CrossRef]

38. Huang, H.-C.; Chen, Y.-H.; Wang, C.-H.; Huang, Z.-H. Analysis of Pitchers' "Expected-Base-Loss" in the Chinese Professional Baseball League using Machine Learning Methods. *J. Taiwan Soc. Sport Manag.* **2024**, *24*, 251–288. [[CrossRef](#)]
39. Li, S.F.; Huang, M.L.; Li, Y.Z. Exploring and selecting features to predict the next outcomes of MLB games. *Entropy* **2022**, *24*, 288. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.