World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Predicting rugby world cup outcomes with machine learning: A comparative model evaluation

Ifeanyi Innocent Ugwu *

University of South Wales.

World Journal of Advanced Research and Reviews, 2026, 31(01), 233–241

Publication history: Received on 27 May 2026; revised on 01 July 2026; accepted on 03 July 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.31.1.1832

Abstract

Predicting the outcome of rugby union matches has traditionally relied on expert judgement and historical observation, an approach that is slow, difficult to scale across large fixture lists, and susceptible to human bias. This study investigates whether supervised machine learning can provide a more objective and scalable alternative, using the 2023 Rugby World Cup qualification-era fixtures as the evaluation context. A dataset of international rugby results spanning 1871 to 2024 was compiled, cleaned, and enriched through feature engineering, including team ranking points, recent-form indices, and match-experience counts. Four classification algorithms; Logistic Regression, Support Vector Machine with a radial basis function kernel, Random Forest, and AdaBoost, were trained on an 80/20 train-test split and evaluated using accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), with hyperparameters tuned via grid and randomised search. Random Forest achieved the highest overall accuracy (72%) and the strongest discrimination for home victories (AUC = 0.88), followed by AdaBoost (70%), and Logistic Regression and SVM (69% each). All models struggled to predict draws, a consequence of severe class imbalance in the underlying data. The findings confirm that ensemble tree-based methods offer a modest but consistent advantage over linear and margin-based classifiers for rugby outcome prediction, while highlighting class imbalance and limited feature richness as the principal barriers to further improvement. The paper concludes with recommendations for resampling strategies, external validation, and the inclusion of player- and weather-level features in future work.

Keywords: Rugby World Cup; Sports Analytics; Machine Learning; Match Outcome Prediction; Random Forest; Class Imbalance

1. Introduction

The Rugby World Cup is among the most closely followed events in international sport, drawing sustained attention from fans, broadcasters, and analysts across its Pool Phase, qualifying play-offs, and knockout rounds (Swart, 2017). Forecasting who will win a given fixture, how far a team will progress, and by what margin has long been a source of debate among commentators and supporters. Historically, such forecasts have depended on expert opinion and manual analysis of past form, an approach that is time-consuming, difficult to apply consistently across a large number of fixtures, and vulnerable to the biases of individual analysts (Smith & Jones, 2019).

The growth of data analytics and machine learning (ML) has opened an alternative path: data-driven models capable of processing large volumes of historical results quickly and consistently. Such techniques have already been applied successfully in football and basketball, but rugby, and the Rugby World Cup specifically, remains comparatively under-studied (Bunker & Susnjak, 2022; Horvat & Job, 2020). This study addresses that gap by building and comparing several supervised ML classifiers trained on 153 years of international rugby results (1871–2024) to predict match outcomes for the 2023 Rugby World Cup.

* Corresponding author: Ifeanyi Innocent Ugwu; Email:enearow0361@gmail.com

1.1. Problem Statement and Motivation

Traditional, expert-driven approaches to forecasting match outcomes do not scale well to modern tournament calendars, in which dozens of fixtures may need to be assessed within a short window. They are also prone to inconsistency, since predictions rest on the subjective judgement of the analyst rather than on a reproducible, quantitative process (Doe & Lee, 2021; Brown et al., 2020). While ML-based sports prediction has matured considerably in football and other high-profile sports, its application to the Rugby World Cup specifically has received little dedicated attention in the literature. This creates both a practical gap; fans, bettors, and team analysts lack a rugby-specific predictive tool — and an academic one, since the comparative performance of standard classification algorithms on rugby data has not been systematically documented (Johnson & Williams, 2022).

Aim and Objectives

The aim of this study is to develop and evaluate a predictive model for Rugby World Cup match outcomes using established machine learning classifiers. The specific objectives are to:

- Investigate and pre-process a long-run dataset of international rugby results to construct features suitable for predicting match outcomes from basic inputs such as competing teams and venue.
- Implement and compare the performance of four classification algorithms — Logistic Regression, Support Vector Machine, Random Forest, and AdaBoost — in predicting home win, away win, and draw outcomes.
- Identify the strengths, weaknesses, and sources of error in each model, with particular attention to class imbalance between draws and decisive results.

To the authors' knowledge, no prior published study has applied this specific combination of classifiers to Rugby World Cup outcome prediction using a dataset of this historical depth, making this comparison a novel contribution to the sports-analytics literature.

2. Literature Review

Predicting sports outcomes is inherently difficult because of the many interacting factors that determine a result, and this uncertainty is part of what makes sport engaging to follow. Machine learning applications in sport have expanded rapidly in recent years, spanning injury prediction (Van Eetvelde et al., 2021), training-load analytics (Rajšp & Fister, 2020), and team-selection optimisation (Ishi & Patil, 2021). This review focuses specifically on studies that use ML to predict match outcomes in team sports, with an emphasis on rugby.

2.1. Outcome Prediction in Football and Other Sports

Most sports-outcome prediction research has concentrated on football (Bunker & Susnjak, 2022; Horvat & Job, 2020; Stekler et al., 2010; Wunderlich & Memmert, 2021). A widely cited example built two supervised models — Random Forest and Gradient Boosting — for the FIFA 2022 World Cup using team rankings and performance metrics; Random Forest achieved roughly 78% precision, marginally ahead of Gradient Boosting at 77%, and was preferred for its lower risk of overfitting (Swart, 2017). Separately, probabilistic approaches using the Poisson distribution have also been applied to World Cup forecasting with reasonable success (Andrew, 2022), illustrating that both classification and probabilistic techniques have a role in this domain.

2.2. Machine Learning in Rugby Prediction

Work specific to rugby is comparatively sparse. Early studies applying neural networks and exponential smoothing to Greek Rugby Union matches reported accuracies between 55% and 64%. Later comparative work testing Naïve Bayes, K-means, K-Nearest Neighbour, decision trees, and artificial neural networks on professional rugby matches found neural networks generally strongest, while K-means performed best for narrowly defined tasks such as identifying large-margin wins. A Poisson-based estimator trained on observed scoring patterns has also been shown to outperform several supervised binary classifiers, including logistic regression, k-nearest neighbours, decision trees, SVM, neural networks, and random forest, for ranking purposes.

A neural network built with a Keras Sequential architecture correctly predicted the winners of 27 of 30 matches (roughly 90% accuracy) at the 2019 Rugby World Cup, and was within five points of the actual margin in about two-thirds of its first-round predictions; its performance declined later in the tournament because team dynamics were not carried forward between group-stage matches (Lamb, 2019). Support-vector-based models applied to Six Nations data achieved accuracies of 55–60% (Celentano et al.), while a comparison of decision trees, random forest, AdaBoost,

logistic regression, SVM, and ordinary least squares on 1,507 IRB matches over four seasons focused on points-difference prediction without reporting standard classification metrics (Kempe & Dohle).

More recent studies report accuracies above 80%: a model trained on three seasons of European Super League data achieved 85.5% test accuracy (Parmar et al., 2017), and a related study examined the 2016–17 English Premiership season to identify the most reliable predictive approach (Bennett et al., 2019). Koekemoer (2023) benchmarked several classifiers on Rugby World Cup data and found Random Forest most accurate at approximately 85%, ahead of Support Vector Classifier (82%), K-Nearest Neighbours (80%), AdaBoost (79%), Logistic Regression (78%), and Decision Tree (76%). Hvattum et al. (2010) combined an Elo-style rating system with gradient-boosted trees on a 1871–2023 rugby dataset, reaching 81% accuracy for match winners overall (82.5% at group stage, 71% in knockout rounds) and correctly predicting the score margin within seven points around 30% of the time.

2.3. Synthesis

Across this body of work, ensemble tree-based methods — particularly Random Forest and gradient boosting — consistently perform as well as or better than linear and margin-based classifiers such as logistic regression and SVM, though reported accuracies vary widely (55–90%) depending on dataset scope, feature richness, and competition specificity. Draws and away wins are consistently harder to predict than home wins across studies, reflecting the general rarity of draws in rugby and the historical strength of home advantage. However, no reviewed study applies this exact four-model comparison to Rugby World Cup fixtures using a dataset spanning the full 1871–2024 international record, which is the gap this study addresses.

3. Methodology

The prediction task treats each fixture as an independent event, on the premise that a team's historical performance and rating trajectory are the best available predictors of its performance in a future match. Predictions therefore combine historical form with an evolving ranking-points system.

3.1. Tools and Software

All analysis was implemented in Python within Jupyter Notebook, using NumPy and Pandas for data handling, Matplotlib and Seaborn for visualisation, and Scikit-learn for model development, hyperparameter search, and evaluation.

3.2. Dataset

The dataset comprises international men's rugby results compiled from publicly available match records (1871–2024) for ten leading nations: England, Wales, Ireland, Scotland, Italy, France, South Africa, New Zealand, Australia, and Argentina. Its structure was modelled on an established international-football results dataset to support methodological comparability. Table 1 summarises the original fields.

Table 1 Original dataset fields.

Field	Description
date	Date the match was played (standardised for time-based analysis)
home_team / away_team	Identity of the home and visiting teams
home_score / away_score	Final points scored by each side
competition	Tournament or series context of the match
stadium / city / country	Venue details, used to assess location effects
neutral	Boolean flag for neutral-venue matches
world_cup	Boolean flag identifying Rugby World Cup fixtures

3.3. Data Pre-processing

3.3.1. Data Cleaning

The raw dataset contained 23 missing values, confined entirely to the competition field. These were imputed using Scikit-learn's SimpleImputer with a most-frequent-value strategy, preserving the existing categorical distribution rather than introducing an artificial placeholder category.

3.3.2. Feature Engineering

Because feature quality typically contributes more to model performance than algorithm choice (Kuhn & Johnson, 2019), a set of derived variables was engineered to capture team strength, recent form, and match experience (Table 2).

Table 2 Engineered features added during pre-processing.

Feature	Description
winner / loser	Derived from the score margin; used for outcome labelling
ranking_points_home / away	Iteratively updated team-strength ratings
margin	home_score – away_score
result	Categorical target: home_win, away_win, or draw
home_form / away_form	Outcome record over the previous 5 matches
home_matches_played / away_matches_played	Cumulative match experience per team
total_matches	Combined experience of both competing teams

Categorical variables were converted to numeric form using Label Encoding, and all numeric features were standardised with a StandardScaler prior to model training.

3.3.3. Exploratory Data Analysis

Exploratory analysis showed that home teams tend to score more heavily than away teams, that ranking points for opposing sides are positively correlated (reflecting a tendency for similarly ranked teams to meet), and that match margin correlates strongly with score levels at both ends. Recent form was also informative: median home_form was approximately 3.5 for home wins, 2.5 for draws, and 2.0 for away wins, indicating that teams in good recent form at home are the most likely to win there. Total match experience showed comparatively little separation across outcome classes, suggesting that raw experience is a weaker discriminator than recent form or ranking.

3.3.4. Feature Selection

Both filter and wrapper approaches to feature selection were considered. The filter method — implemented via Scikit-learn's SelectKBest using univariate statistics — was adopted as the primary approach for its lower computational cost on this dataset, consistent with Lamb (2019).

3.4. Model Development and Training

The processed data was split 80/20 into training and test sets (Figure omitted for brevity), the larger training partition supporting model generalisation while the held-out test set provided an unbiased estimate of performance on unseen fixtures. Four classifiers were implemented:

- Logistic Regression — a probabilistic linear classifier using a sigmoid function to map inputs to class probabilities; the 'liblinear' solver was used with a One-vs-Rest scheme, suited to smaller, lower-resource datasets.
- Support Vector Machine (SVC) — implemented with a radial basis function (RBF) kernel to capture non-linear relationships between features, with gamma left at its default (scale) setting.
- Random Forest — an ensemble of decision trees combined by majority vote (feature bagging), chosen for its robustness to overfitting and capacity to handle high-dimensional, non-linear data.

- AdaBoost — an ensemble method that iteratively re-weights misclassified observations so that subsequent weak learners focus on harder cases, improving accuracy while remaining resistant to overfitting on smaller or noisier datasets.

3.5. Model Evaluation

Model performance was assessed using a confusion matrix (true/false positives and negatives per class) together with four standard metrics: accuracy — the proportion of correctly classified instances; precision — the proportion of predicted-positive cases that were correct; recall — the proportion of actual-positive cases correctly identified; and F1-score — the harmonic mean of precision and recall. The area under the ROC curve (AUC) was used to summarise each model's ability to discriminate between outcome classes across all decision thresholds.

3.6. Cross-Validation and Hyperparameter Tuning

Given the size of the dataset, cross-validation was used throughout model development, and hyperparameters for each algorithm were optimised using GridSearchCV and RandomizedSearchCV to maximise generalisation performance rather than in-sample fit.

4. Results and discussion

Table 3 reports precision, recall, F1-score, and AUC for each model across the three outcome classes (away win, draw, home win); Table 4 gives overall test-set accuracy; Figures 1–3 present the same results graphically.

Table 3 Classification performance by model and outcome class.

Model	Class	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Logistic Regression	Away win	60	62	61	78
	Draw	0	0	0	76
	Home win	76	79	77	81
SVM (RBF)	Away win	60	62	61	77
	Draw	0	0	0	66
	Home win	75	79	77	79
Random Forest	Away win	64	68	66	77
	Draw	0	0	0	69
	Home win	77	79	78	88
AdaBoost	Away win	61	62	62	83
	Draw	0	0	0	29
	Home win	75	80	78	11

Table 4 Overall test-set accuracy by model.

Model	Overall Test Accuracy
Logistic Regression	69%
SVM (RBF)	69%
Random Forest	72%
AdaBoost	70%

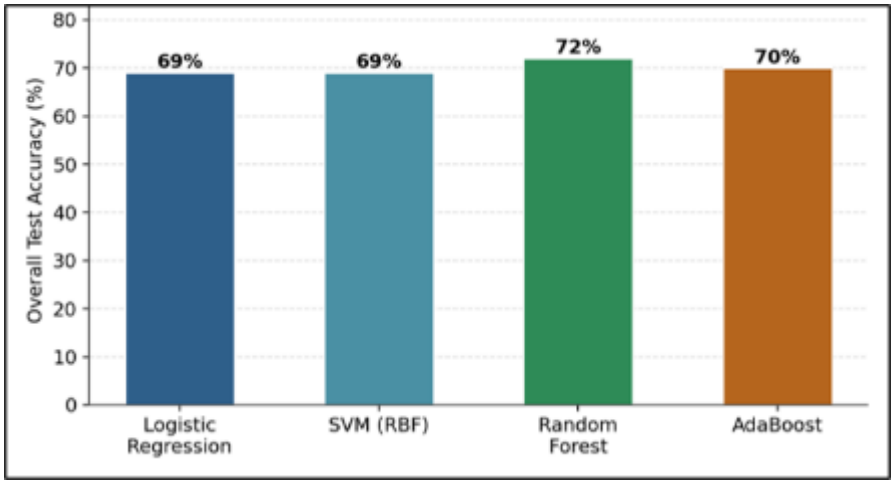

Figure 1 Overall predictive accuracy by model.

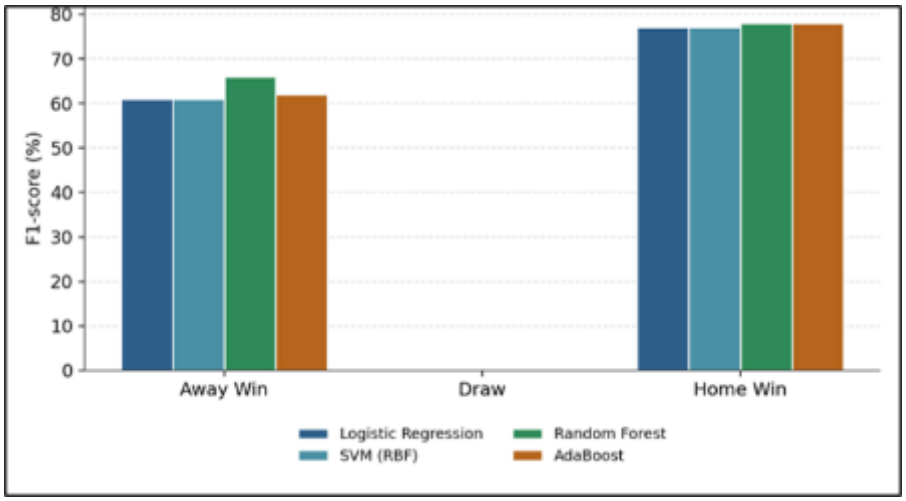

Figure 2 F1-score by match outcome class and model.

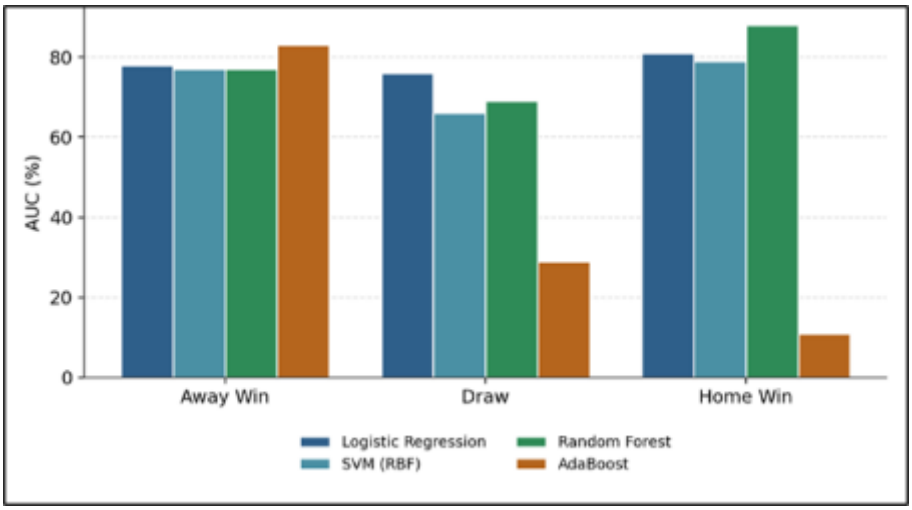

Figure 3 ROC-AUC by match outcome class and model.

4.1. Model-by-Model Findings

4.1.1. Support Vector Machine

The SVM classifier predicted home wins reliably (75% precision, 79% recall, AUC 0.79) and away wins moderately well (60% precision, 62% recall, AUC 0.77), but failed entirely on draws, with precision, recall, and F1-score all at 0% despite a moderate AUC of 0.66. The confusion matrix showed 38% of true away wins misclassified as home wins, while draws were never predicted, indicating the model defaults toward the two dominant, more frequent classes.

4.1.2. Logistic Regression

Logistic Regression produced a very similar pattern to SVM: strong home-win prediction (76% precision, 79% recall, AUC 0.81), fair away-win prediction (60% precision, 62% recall, AUC 0.78), and complete failure on draws in terms of precision, recall, and F1-score, despite a relatively high draw AUC of 0.76. This gap between AUC and threshold-based metrics for draws suggests that the model can partially rank draw-likelihood but the default 0.5 decision threshold is never crossed for this minority class.

4.1.3. Random Forest

Random Forest was the strongest overall performer, achieving 72% test accuracy — the best of the four models — with the highest home-win AUC recorded in the study (0.88) alongside 77% precision and 79% recall for that class. Away-win performance was moderate (64% precision, 68% recall, AUC 0.77), while draws remained entirely unpredicted (0% precision/recall/F1) despite a 69% AUC, again pointing to a thresholding problem driven by class imbalance rather than a total absence of predictive signal.

4.1.4. AdaBoost

AdaBoost achieved 70% accuracy, with strong home-win precision (75%) and recall (80%) and a very high away-win AUC of 0.83. However, its home-win AUC fell to just 0.11 despite strong precision/recall for that class — an inconsistency that most likely reflects sensitivity to threshold placement and probability calibration in AdaBoost's boosted-ensemble output rather than a genuine loss of discriminative ability, and highlights AUC's limitations as a standalone metric for boosted classifiers evaluated on imbalanced, multi-class problems. Draws were again unpredicted (0% precision/recall/F1) with the lowest draw AUC of the four models (0.29).

4.2. Comparative Analysis

Across all four models, home wins were predicted most reliably, away wins moderately well, and draws not at all in terms of threshold-based classification metrics. Random Forest's advantage over the other three models (72% vs 69–70%) is consistent with its ability to model non-linear interactions between ranking points, form, and experience without the strict decision-boundary assumptions of logistic regression or the single-kernel geometry of SVM. AdaBoost's strong away-win AUC (0.83) alongside its very low home-win AUC (0.11) illustrates that boosting techniques can trade class-level consistency for improvements on individual classes, and reinforces the value of examining per-class AUC rather than relying on a single aggregate score.

The complete inability of every model to predict draws is best explained by class imbalance rather than by any single algorithm's weakness. As Figure 6.5 in the underlying dataset analysis illustrates, non-World-Cup matches alone contain roughly 1,480 home wins and 1,020 away wins against only about 120 draws — an imbalance of more than ten to one. Standard classifiers trained to minimise overall error naturally gravitate toward the majority classes, since predicting 'no draw' is correct far more often than not. This mirrors the broader sports-analytics literature, where draws in low-scoring or high-scoring sports are consistently the hardest outcome to classify (Bunker & Susnjak, 2022).

4.3. Hyperparameter Tuning Outcomes

Table 5 reports the best hyperparameter configurations identified via grid and randomised search for the two ensemble models, which delivered the top two accuracy scores in this study.

Table 5 Best hyperparameters identified for the ensemble models.

Model	Best hyperparameters	Best CV accuracy
Random Forest	n_estimators = 100; min_samples_split = 2; min_samples_leaf = 6; max_features = 'sqrt'; max_depth = 12; bootstrap = True	0.72
AdaBoost	n_estimators = 100; learning_rate = 0.1	0.70

The relatively shallow maximum depth (12) and modest leaf size (6) selected for Random Forest suggest that a moderately regularised ensemble generalised better than a deeper, more complex forest — consistent with the model's comparatively strong and stable performance across classes.

5. Conclusion

This study compared four supervised machine learning classifiers — Logistic Regression, Support Vector Machine, Random Forest, and AdaBoost — for predicting Rugby World Cup match outcomes using a 153-year international results dataset enriched with ranking, form, and experience features. Random Forest emerged as the most effective model overall, achieving 72% test accuracy and the strongest discrimination for home-win outcomes (AUC = 0.88), followed by AdaBoost (70%) and a tie between Logistic Regression and SVM (69% each). All four models predicted home wins reliably and away wins moderately well, but none could correctly classify draws, a limitation attributable to severe class imbalance rather than to any specific algorithmic weakness.

These findings support the broader literature's conclusion that ensemble tree-based methods offer a modest but real advantage over linear and margin-based classifiers for team-sport outcome prediction, while underscoring that data balance and feature richness — not merely algorithm choice — are the binding constraints on further accuracy gains in rugby-specific prediction tasks.

Limitations

- Data availability: Rugby World Cup-specific outcome data is far less extensively documented and studied than equivalent data for other major sports, constraining the depth of historical signal available to the models.
- Dataset characteristics: the working dataset, while spanning 153 years, remains modest in size and homogeneity relative to what deep-learning approaches typically require, increasing the risk of overfitting to dominant classes.
- Feature diversity: the dataset does not include player-level statistics, tactical indicators, injury data, or weather conditions, any of which could materially improve predictive power, particularly for closely contested fixtures likely to end in a draw.

Recommendations for Future Research

- Apply resampling techniques such as SMOTE or targeted class-weighting to address the severe under-representation of draws before training, rather than relying on default decision thresholds.
- Incorporate domain-informed feature selection — for example, refining ranking-point calculations or introducing rugby-specific performance indicators — rather than relying solely on generic statistical feature selection methods.
- Validate tuned models against an external, held-out tournament dataset (rather than cross-validation alone) to confirm real-world generalisability before operational deployment.
- Expand the feature set to include player availability, recent injury history, travel distance, and weather conditions, which are plausible drivers of upset results and draws in rugby union.
- Explore probability calibration techniques (e.g., Platt scaling or isotonic regression) for boosted models such as AdaBoost, given the AUC inconsistencies observed for the home-win class in this study.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Andrew (2022) Predicting the World Cup 2022 with Python and Poisson distribution. Towards Data Science.

[2] Bennett, M., Bezodis, N., Shearer, D.A., Locke, D. and Kilduff, L.P. (2019) 'Descriptive conversion of performance indicators in rugby union', Journal of Science and Medicine in Sport, 22(3), pp.330–334.

[3] Bunker, R.P. and Thabtah, F. (2019) 'A machine learning framework for sport result prediction', Applied Computing and Informatics, 15(1), pp.27–33.

[4] Bunker, R. and Susnjak, T. (2022) 'The application of machine learning techniques for predicting match results in team sport: A review', Journal of Artificial Intelligence Research, 73, pp.1285–1322.

[5] Colomer, C.M., Pyne, D.B., Mooney, M., McKune, A. and Serpell, B.G. (2020) 'Performance analysis in rugby union: a critical systematic review', Sports Medicine-Open, 6, pp.1–15.

[6] Freund, Y. and Schapire, R.E. (1997) 'A decision-theoretic generalization of on-line learning and an application to boosting', Journal of Computer and System Sciences, 55(1), pp.119–139.

[7] Horvat, T. and Job, J. (2020) 'The use of machine learning in sport outcome prediction: A review', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(5), e1380.

[8] Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. International Journal of forecasting, 26(3), 460-470.

[9] Koekemoer, D. (2023) Predicting Rugby World Cup outcomes with supervised learning models. Kaggle.

[10] Kuhn, M. and Johnson, K. (2019) Feature Engineering and Selection: A Practical Approach for Predictive Models. Boca Raton: CRC Press.

[11] Lamb, R. (2019) Using machine learning to predict sport scores — a Rugby World Cup example. Towards Data Science.

[12] O'Donoghue, P. and Williams, J. (2004) 'An evaluation of human and computer-based predictions of the 2003 rugby union world cup', International Journal of Computer Science in Sport, 3(1), pp.5–22.

[13] Parmar, N., James, N., Hughes, M., Jones, H. and Hearne, G. (2017) 'Team performance indicators that predict match outcome and points difference in professional rugby league', International Journal of Performance Analysis in Sport, 17(6), pp.1044–1056.

[14] Richter, C., O'Reilly, M. and Delahunt, E. (2021) 'Machine learning in sports science: challenges and opportunities', Sports Biomechanics, pp.1–7.

[15] Stekler, H.O., Sendor, D. and Verlander, R. (2010) 'Issues in sports forecasting', International Journal of Forecasting, 26(3), pp.606–621.

[16] Swart, K. (2017). The Rugby World Cup as a global mega-event. In *The rugby world in the professional era* (pp. 108-117). Routledge.

[17] Van Eetvelde, H. et al. (2021) Machine learning methods in sport injury prediction and prevention: a systematic review. Journal of Experimental Orthopaedics.

[18] Wunderlich, F. and Memmert, D. (2021) 'Innovative approaches in sports science — lexicon-based sentiment analysis as a tool to analyze medial perception of sport', Sustainability, 13(1), p.61.