Sports Match Prediction Analysis Applying Machine Learning Model with Rule-Based Approach

Jay Abaleta, Ma. Christina Navarro, Jerome Alvez, Vinard James Damasco, Yvan Sinues, John Miguel Jusayan and Harvey Delos Santos

July 3, 2025

2024 IEEE 9th International Conference on Engineering Technologies and Applied Sciences (ICETAS)

Sports Match Prediction Analysis Applying Machine Learning Model with Rule-Based Approach

1st Jay Abaleta, MSIT
Computer Science Department
Adamson University
Metro Manila, Philippines
jay.abaleta@adamson.edu.ph

2nd Ma. Christina Navarro, MSCS
Computer Science Department
Adamson University
Metro Manila, Philippines
ma.christina.navarro@adamson.edu. ph

3rd Jerome Alvez, MSCS
Computer Science Department
Adamson University
Metro Manila, Philippines
jerome.alvez@adamson.edu.ph

4th Yvan Sinues
Computer Science Department
Adamson UniversityMetro
Cavite, Philippines
yvan.sinues@adamson.edu.ph

5th John Miguel Jusayan
Computer Science Department
Adamson University
Cavite, Philippines
john.miguel.jusayan@adamson.edu.ph

6th Vinard James Damasco
Computer Science Department
Adamson University
Makati City, Philippines
vinard.james.damasco@adamson.edu.ph

7th Harvey Delos Santos
Computer Science Department
Adamson University
Makati City, Philippines
harvey.delos.santos@adamson.edu.ph

Abstract - This research introduces an innovative approach to match prediction, combining machine learning with a rule-based approach, focusing on PBA (Philippine Basketball Association), and PVL (Premier Volleyball League). Objectives include creating a hybrid model, optimizing datasets, and deploying the model in a web application. The research shows accuracy for basketball (82.86%) and volleyball (90.20%), with random forest model and feature importance analysis highlighting key predictors. The conclusion emphasizes the model's value in sports prediction and recommends further exploration of the proposed approach.

I. Introduction

Research introduces sports match prediction, crucial for sports enthusiasts, gamblers seeking informed insights into sporting events. In the Philippines, where basketball reigns supreme and volleyball gains traction among the youth, study underscores the cultural significance of these sports. Notably, PBA, or Philippine Basketball Association, and PVL, or Premier Volleyball League serve as vital sources of data on players and team performance.

Machine learning models with a rule-based approach, the research contends that traditional prediction methods are becoming unreliable in the face of rapid technological advancements. With over 40 million Filipinos immersed in basketball and the PBA's enduring influence, the wealth of generated data offers an opportune arena for predictive analysis.[1]

Research advocates for the superiority of machine learning models as opposed to conventional techniques, emphasizing their efficacy in navigating the dynamic sports landscape. By using a method based on rules, the study aims to enhance the interpretability of machine-learning models, enabling a better evaluation of their suitability for specific sports. In essence, study positions machine learning as a valuable asset for sports industry professionals, fans, and analysts, offering an avenue to elevate the sports experience through more accurate predictions of future game outcomes.

1.1 Background of Study

This research integrates machine learning and a rule-based approach to predict basketball and volleyball match outcomes in the Philippines. The analysis delves into various factors such as team performance, historical data, and game-specific information. The machine learning model is trained with extensive data, including team per game performance, past results, and related information.

Focusing on basketball and volleyball in the Philippines, the study considers historical team performances, encompassing metrics like goals, free throws, rebounds, etc. The goal is to provide accurate match predictions through model training and meticulous analysis. These predictions are intended for use by sports enthusiasts, fans, and even sports betting enthusiasts, empowering them to make well-informed decisions based on the insights' generated by the integrated machine learning and rule-based approach.

1.2 Objectives of the Study

This study's goal is to use machine learning to forecast the next sporting event. which is (Random Forest Algorithm) ensemble of multiple decision trees with Rule-Based Approach. Training and tested on two sports domains: Basketball and Volleyball. The results will be the predicted game outcomes including the odds of winning and the model's prediction accuracy that the beneficiaries will be bettors and enthusiasts of the said sports. The study also aims:

- To create hybrid model from random forest model with rule-based approach
- To optimize the collected datasets by applying feature selection and pre-processing
- To deploy the model into a website application
- Proving the efficiency of the algorithm used in predicting outcomes
- To be able to predict sports match games for basketball and volleyball outcomes and their odds of winning.

1.3 Significance of the Study

This section of research elucidates the significance and advantages of the study. The findings of this study will prove

979-8-3503-6314-2 /23/$31.00 ©2024 IEEE

valuable to researchers, enthusiasts, teams, agencies, organizations, and other related professionals.

- In the sports Context, emphasizing the increasing accessibility of sports data. The need for ongoing research efforts in match prediction analysis is highlighted, particularly with the rising prevalence. The study suggests that these models offer promising solutions surpassing traditional forecasting methods.
- In a societal Context, potential to enhance the viewing experience for sports fans by providing accurate game outcome predictions. This information can add excitement, anticipation, and assist viewers in making informed choices related to betting, sports preferences, and overall enjoyment.
- In an organizational context, The model's performance metrics serve as benchmarks for continuous improvement. Organizations can strive to enhance the model by refining features, incorporating more data, or adopting advanced modeling techniques.

1.4 Scope and Delimitation

The study exclusively centers on historical team stat sheet data from past matches in the PBA and PVL, deliberately excluding real-time data. Its scope encompasses the utilization of machine learning models with a rule-based approach to predict outcomes and winning odds in head-to-head matchups. By training these models using past data, Research aims to enhance predictive accuracy. Enthusiasts and bettors will be provided with anticipated outcomes and winning probabilities, enabling decision-making in bets. The absence of real-time data highlights the study's emphasis on leveraging historical statistics for comprehensive predictive analysis. Additionally, the research will investigate the impact of rule-based approaches on predictions and underscore the significance of machine learning methodologies in assessing extensive historical datasets related to team performance.

The scope of the study is focused on specific teams and sports within the PBA and PVL, specifically the 2022-2023 PBA Cup and Governor's Cup for basketball, and a corresponding season for volleyball. Analysis utilizes existing datasets, mainly from past seasons, with an additional emphasis on ongoing sports match season data for testing predictions. Notably, player stats, motivational factors, injuries, team dynamics, and overtime are excluded from the study. The primary analytical tool is the Random Forest Model with a rule-based approach, producing output related to head-to-head results, winning odds, and a decision tree plot highlighting key features influencing outcomes.

1.5 Conceptual Framework

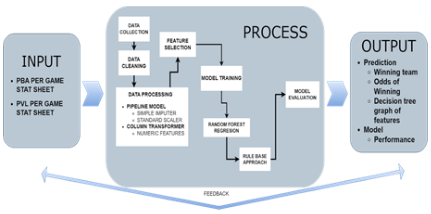


Fig. 1.5.1 Conceptual Framework of Study

Fig. 1.5.1 displays the conceptual framework of the study. It covers the three phases; input, process, output, and feedback.

The input phase involves gathering PBA team historical data and PVL stats.

The process phase comprises several steps, comprising feature extraction, data preprocessing, data cleansing, data gathering, and data classification, and evaluation and selection of the best model. In the data collection step, Data from historical teams vs team sports match statistics from the PBA and PVL league will be used. Pipeline Model which comprises simple imputer and standard scalar, and Column Transformer which numeric features will be employed for data preprocessing.

Feature selection that will involve using Team stats of Home and Away of the sports match, statistics of teams based from different matches of the processed data. The Model classification step utilizes a pre-trained Random Forest with rule-base approach to predict match outcome.

Model Evaluation involves assessing the model's performance using Confusion Matrix, ROC curve, Feature Importance, and other matrices used for classification graphs. Finally, in the output phase, the results of predicting the sports match game of teams give the odds of winning as well as a decision tree graph of the features in deciding the winner.

II. Literature Review

2.1 Sports Analytics

A study by H., Bai, Z., & Bai, X. (2021) This research delves into the significance of sports big data, exploring its background, management and applications. Emphasizing evaluation and prediction in sports data, the study scrutinizes representative research to address key issues. As sports and information technology rapidly intersect, the findings offer valuable insights for future researchers, guiding comprehensive awareness of sports big data. [6]

McCabe and Trevathan's (2008) sports prediction study leverages artificial intelligence, employing a multi-layer perceptron model to forecast outcomes. The model, incorporating team quality features, outperforms human analysts across diverse scenarios. Emphasizing neural network utility, especially multi-layer perceptrons, the research highlights accurate predictions with minimal data. Addressing challenges like noisy environments, it extends to predicting probabilistic events using machine learning. The multi-layer perceptron impressively performs in predicting sports match results despite limited information and external influences. [7]

The study by Haghighat and Rastegari (2014) leverages predictive analytics and data mining techniques on a decade of basketball health data. It identifies factors associated with player injuries, develops injury prediction models, evaluates lineup performance, and explores contributors to team success. This research highlights data analytics' potential to enhance basketball performance. As analytics advances, it promises innovative applications for game improvement. [15]

2.2 Productive Analytics

The study by Constantinou, Fenton, and Neil (2013) focuses on creating a comprehensive machine-learning framework to predict sporting event outcomes. The model considers indicators like past game results, player performance metrics, and opposition data. The first objective aims to develop an inclusive machine learning design for refining rule-based systems in sports. The second goal focuses on enhancing prediction accuracy by integrating more precise rule-based systems into existing machine learning and deep learning technologies. Both research areas aim to mutually complement each other for improved sports predictions. [16]

Garnica-Caparros et al.'s (2022) study underscores exploiting market inefficiencies in sports betting, asserting that bookmakers often rely on outdated data, providing an opportunity for gamblers to employ machine learning. The authors develop a model predicting NBA game outcomes, incorporating previous betting data and additional variables. Demonstrating consistent outperformance of bookmakers' odds, their algorithm proves advantageous over time, highlighting the potential of machine learning in creating more accurate sports outcome prediction models. [18]

The review explores the significance of linear regression in statistical and machine learning, particularly simple and multiple regression models. Maulud and Abdulazeez (2020) analyze 23 recent papers, emphasizing the importance of accurate model variable selection, dataset quality, and performance evaluation. The study underscores the critical role of regression modeling in obtaining valid statistical results, emphasizing the need for precision in predicting and comparing values for optimal efficiency. [5]

Jadhav, Das, Khatode, & Degaonkar. (2016) explored SVM for NBA Playoff predictions, analyzing 2012-2015 game data. Achieving 88% accuracy, their study showcased SVM's effectiveness in forecasting winning teams. The research demonstrated drawing distinct hyperplanes on features, illustrating the model's simplicity. This accurate prediction tool offers potential for NBA teams to devise strategies and aids sports bettors in making informed decisions. [19]

2.3 Machine-Learning-Based Approach in Predictive Analysis

Lotfi and Rebbouj (2021) investigated the combination of sports analytics and machine learning, emphasizing the growing synergy between data science and sports. The study reviews existing literature to reveal the complex mechanisms guiding data modeling and analysis in sports. It highlights the pivotal role of technological advancements in predicting sporting outcomes and aims to unveil the intricate variables affecting athletic performance. The research introduces a robust data mining analysis tool driven by machine learning, envisioning a transformative era of data-driven decision-making for precise performance predictions and strategic interventions in sports.[13]

Breiman's (2001) study underscores the significance of Random Forest Machine Learning in prediction. It highlights their effectiveness in mitigating overfitting, attributing it to the injection of randomness. The approach enhances accuracy in regression and classification tasks by evaluating predictors' strength and correlations. Out-of-bag estimates offer practical insights, challenging the belief that random forests lag behind arcing-type algorithms in accuracy. The conclusion raises intriguing questions about the analogy with boosting and arcing algorithms, renowned for bias and variance reduction. [8]

Biau's (2010) study explores on prediction using decision tree sub-groups. Despite widespread application, limited research delves into statistical properties and mathematical algorithms. This review investigates Random Forest Regression, revealing that prediction convergence depends solely on strong features, not noisy variables. The findings highlight its consistency and sparsity adaptability in predictive modeling. [10]

In Zhao et al. 's (2023) study, a team representation graph using an undirected graph and homogeneous network, coupled with a graph convolutional network, achieved a 66.90% average game prediction success rate. Integration with random forest-based feature extraction increased accuracy to 71.54%, outperforming previous models.

The approach considers spatial organization and team interactions, proving superior in predicting basketball game outcomes. Comparative analysis with baseline models and prior research validates the proposed method's efficacy, providing valuable insights into basketball play dynamics.[9]

Haghighat et al. (2013) emphasized in their study that decision trees are a potent and widely utilized tool for predicting sports outcomes, distinguishing them from artificial neural networks (ANN). Unlike ANN, decision trees offer transparent rules, categorizing data through attribute-based inquiries, forming a hierarchical tree structure. This clarity in the decision-making process sets decision trees apart, providing a distinct advantage in result prediction compared to the concealed paths of ANN. [11]

Lam's (2018) study emphasizes the underexplored intersection of real-world applications particularly in sports analytics. Addressing the scarcity of machine-learning approaches in this domain, the study introduces TLGProb, the Sparse Spectrum Gaussian Process Regression algorithm. Notably, TLGProb excels in predicting NBA 2014/2015 season outcomes, achieving 85.28% accuracy and outperforming existing models, highlighting its significance in advancing sports analytics.[12]

Alonso and Babac (2022) explored machine learning for predicting NBA game outcomes, aiming to enhance accuracy amid sports uncertainty. Their study revealed Naive Bayes as the most effective classifier, achieving 70.5% accuracy and maintaining a balance between accuracy and recall. KNN yielded 70% accuracy, but lower recall for Lost categorization. Conversely, Decision Trees struggled with 60% accuracy and imbalanced metrics. Considerations for dataset size and feature correlation are crucial for future predictions, especially with new probable matches emerging. [20].

III. Research Methodology

The approach, and procedures that will be used in the current investigation are provided and explored in detail in this chapter. This chapter clarifies to achieve the research objectives, this chapter emphasizes the creation of the research technique and research process.

3.1 Data Collection

To gather the data, the research used sports match (team vs team) stats of basketball and volleyball from the PBA and PVL 2022-2023 data. Throughout the data collection process, the data gathered has been 178 team matchup statistical data from PBA Leagues and 324 team matchup statistics from PVL has been obtained.

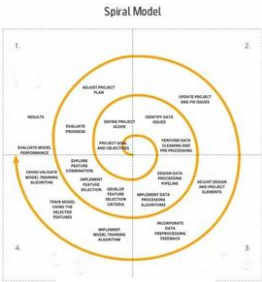
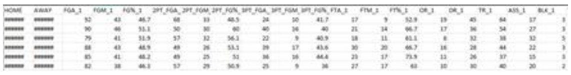
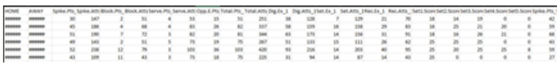


Fig. 3.2.1



HOME	AWAY	FGA_1	FGM_1	FG%_1	2PT_FGA_	2PT_FGM	2PT_FG%	3PT_FGA_	3PT_FGM	3PT_FG%	FTA_1	FTM_1	FT%_1	OR_1	DR_1	TR_1	ASS_1	BLK_1
######	######	92	43	46.7	68	33	48.5	24	10	41.7	17	9	52.9	19	45	64	17	3
######	######	90	46	51.1	50	30	60	40	16	40	21	14	66.7	17	36	54	27	3
######	######	79	41	51.9	57	32	56.1	22	9	40.9	18	11	61.1	6	32	38	32	5
######	######	88	43	48.9	49	26	53.1	39	17	43.6	30	20	66.7	16	28	44	22	3
######	######	85	41	48.2	49	25	51	36	16	44.4	23	17	73.9	11	26	37	15	3
######	######	82	38	46.3	57	29	50.9	25	9	36	27	17	63	10	30	40	20	2

Fig 3.2.1.1 Sample PBA Match Dataset



HOME	AWAY	Spike.Pts_	Spike.Atts	Block.Pts_	Block.Atts	Serve.Pts_	Serve.Atts	Opp.E.Pts	Total.Pts_	Total.Atts	Dig.Ex_1	Dig.Atts_1	Set.Ex_1	Set.Atts_1	Rec.Ex_1	Rec.Atts_	Set1.Score	Set2.Score	Set3.Score	Set4.Score	Set5.Score	Spike.Pts_
######	######	30	147	2	51	4	53	15	51	251	38	128	7	129	21	70	18	14	19	0	0	42
######	######	45	186	4	68	4	83	26	82	337	58	159	16	158	29	83	16	25	21	20	0	59
######	######	51	190	7	72	3	82	20	81	344	63	173	14	156	31	91	18	16	26	21	0	68
######	######	49	143	2	51	5	75	19	75	267	51	133	15	111	26	62	25	25	25	0	0	43
######	######	52	238	12	79	3	103	36	103	420	93	216	14	203	40	95	25	20	25	25	8	59
######	######	43	109	11	43	3	75	18	75	225	31	94	14	87	14	43	25	0	0	0	0	27

Fig 3.2.1.2 Sample PVL Match Dataset

3.2 Data Preprocessing

It Is crucial for enhancing the consistency and accuracy of machine learning models. Techniques like converting data to float and creating a numerical transformer involving normalization and imputation are employed. The 'numeric_transformer' pipeline plays a key role, ensuring mean 0 and variance 1 for features and handling missing values. The 'preprocessor' utilizes this transformer for numerical feature preprocessing, as there are no categorical features in the code. Furthermore, similar preprocessing steps are applied to season per game data, maintaining alignment with historical data processing. The combined effect of these techniques is to better prepare the data for efficient analysis of machine learning and model comprehension.

3.3 Feature Selection

Feature selection in machine learning involves defining relevant features, using them in model training and evaluation, and ensuring consistency throughout the entire pipeline. The aim is to identify the most informative subset of features for predicting team performance in terms of wins and losses. This process includes creating two feature lists, team1_features and team2_features, containing relevant data columns. Team1_features focus on the team's performance statistics, while team2_features include opponent performance statistics. When predicting the next season's performance, the same feature selection is applied to maintain consistency. Selected features are those present in the feature lists, ensuring a robust and comparable approach across seasons.

3.4 Model Training

The model training process begins by loading necessary libraries and past season data for either basketball or volleyball. The data undergoes preprocessing, including type conversion and the creation of interaction features. Relevant features are selected, and the target variable 'Outcome' is defined based on team scores. A rule-based prediction function is established, and a preprocessing pipeline is created with imputation and scaling. A Random Forest Classifier is trained. Utilizing a pipeline while the model evaluation metrics are missing in the provided code, the trained model is serialized using joblib, enabling future predictions for determining match outcomes between user-inputted teams in the chosen sport.

3.4.1 Random Forest Model

The machine learning model Random Forest Classifier from scikit-learn, specializing in predicting outcomes (Win/Lose) in team-vs-team sports using historical game data. Trained on a dataset with team statistics, it undergoes preprocessing with features standardized and missing values handled. Using decision trees as base models, it captures complex data relationships. Engineered features and a rule-based function enhance predictive capabilities. The model, with defined parameters, is persistently stored and works through a model_pipeline for data preprocessing. During predictions, users input two team names, and the model forecasts the winner with associated win probability. The ensemble of decision trees is the basis of the prediction formula. The random Forest model formula:

$$\hat{Y}_N = \frac{1}{N}\sum_{i=1}^{N} f_i(X)$$

where:

- $\hat{Y}$ is the predicted value for the number of wins or losses in the basketball season,
- N is the total number of Random Forest ensemble decision trees,
- $f_i(X)$ represents the prediction from the individual decision tree for the input features

3.4.2 Rule-based Approach

Predetermined criteria, including comparing field goal percentages or blocks, are part of the rule-based approach to match prediction. The `rule_based_prediction` function evaluates these conditions, returning predictions like Team1 winning if their `'FG%'` is higher. A default tie prediction (`'Tie'`) is made when no specific rule applies. Effectiveness depends on chosen rules' relevance to the sport and dataset, requiring adjustments based on domain knowledge and data characteristics for optimal results.

3.5 Model Evaluation

An essential first step in evaluating the effectiveness of machine learning models is model evaluation. for predicting the outcomes of sports match outcomes. The primary objectives are defined first, specifying the goals of the investigation. A baseline model is established as a benchmark, and Test, validation, and training sets of data are separated apart. For stability and optimization, cross-validation and hyper parameter adjustment are optional.

Accuracy is one of the criteria used to evaluate the model. following training and testing on the validation set:

- Accuracy: (TP + TN) / (TP + TN + FP + FN)

- Precision: TP / (TP + FP)
- Recall: TP / (TP + FN)
- F1 Score: 2 * (Precision * Recall) / (Precision + Recall).

A confusion matrix is developed, which offers a comprehensive evaluation of genuine positives. Interpretability, error analysis, external validation, comparison analysis, sensitivity analysis, and thorough documentation all support an all-encompassing research methodology. These components aid in continuous improvement and iteration as well.

3.6 Hardware Specification

This study uses hardware with the following specifications in Table 3.9.1.

Hardware	***Specification***
Operating System	Windows 10&11
System Type	64 bit
Processor	AMD Quad Core R3 - 2200G Vega 2 graphics @3.6 Hz (4 CPUs)
RAM	16 GB
Solid State Drive	120 GB
Hard Drive	1TB + 500GB

Table 3.9.1 Hardware specifications

IV. Results and Discussion

The findings of our research, with a particular emphasis on the use of machine learning techniques with a rule-based approach to predicting the outcome of a sporting event in terms of attributes and winning odds. The discussion interprets these results, providing insights into the predictive capabilities of Random Forest Algorithm with a rule-based approach.

4.1 Model Performance

Following training and testing for use in a classification model, we assessed the Random Forest Model's accuracy, precision, recall, and F1 Score, as indicated in Table 4.1.

Sports	Accuracy	Precision	Recall	F1 Score
Basketball	82.86%	80.95%	89.47%	85.00%
Volleyball	90.20%	88.46%	92.00%	90.20%

Table 4.1 Model Performance

4.2 Random Forest Model

A 100 decision tree was used to forecast the winning probabilities for the team, providing a reliable technique. The Confusion Matrix assessment showed that it was successful in identifying complex associations in the data. The feature significance scores (derived from the decrease in impurity across decision trees) presented in Figures 4.2.1 and 4.2.2 indicate the relative contribution of each feature to the overall prediction performance. These scores highlight the significant roles features have in predictions, without showing the direction of the association. Although feature importance should be taken into account when choosing features, care must be taken because of the relativity of the scores and the possibility of problems with highly linked qualities.

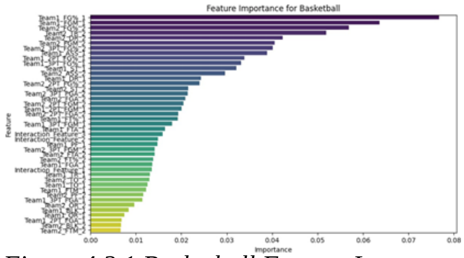


Figure 4.2.1 Basketball Feature Importance

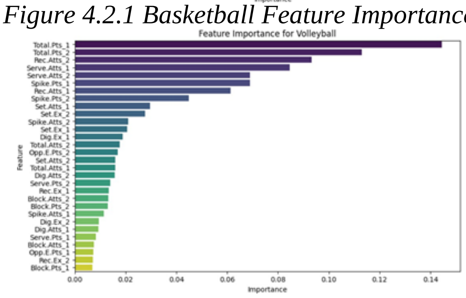


Figure 4.2.2 Volleyball Feature Importance

4.3 Evaluation of Predictions

We evaluated the usefulness of our model by training the model with 80% of data and 20% for testing. The accuracy and other metrics were assessed using confusion matrix results and ROC curve as shown in Figures 4.3.1, 4.3.2, 4.3.3, 4.3.4 which interprets that the overall accuracy of the Confusion Matrix for basketball is 82.86% while in volleyball is 90.20%.

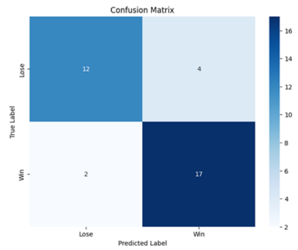


Figure 4.3.1. Basketball Confusion Matrix

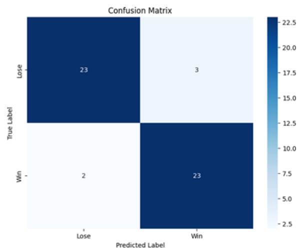


Figure 4.3.2. Volleyball Confusion Matrix

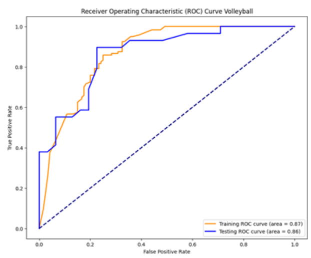


Figure 4.3.3. Volleyball ROC Curve

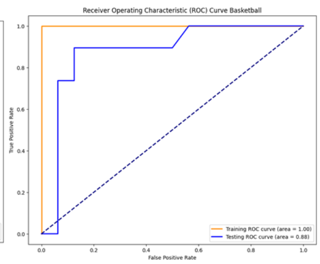


Figure 4.3.4. Basketball ROC Curve

4.4 Discussion

Our research's discussion part focuses on analyzing the findings and their implications

- Because of the model's predicted accuracy, sports fans may watch games with more enthusiasm and anticipation, which helps them make well-informed judgments about their betting and sports preferences.
- By combining a rule-based methodology that improves interpretability for sports experts and bridges the gap

between machine learning complexity and usable insights, our study adds to the expanding area of sports predictive analytics.

- Even with the model's successes, it is crucial to be aware of its drawbacks, such as relativity of feature importance scores and difficulties with highly correlated attributes. Further research ought to investigate ways to overcome these constraints and improve the model's performance
- The study, which focuses on basketball and volleyball in the Philippines, raises the possibility of wider applicability to different sports and offers a more thorough comprehension of its generalizability.

V. Conclusion and Recommendation

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

5.1 Conclusion

Accurately forecasting match outcomes has become a valuable talent in the ever-changing world of sports. Sports prediction is an area that has drawn the interest of many people, whether it is for the excitement of a fan's insight or the tactical decisions of a sports enthusiast. This study conducted an experiment on predicting the match outcomes of PBA, or Philippine Basketball Association, and PVL, or Premier Volleyball League by integrating a machine learning model with a rule-based approach. The researchers assessed and analyzed the datasets from the PBA and PVL teams' stat sheets, and then used random forest regression to train the machine learning model.

The Accuracy, Precision, Recall, F1 Score, Confusion Matrix, and the ROC were evaluated to train and test its performance in the classification model. As a results in the training and evaluation of random forest regression in predicting the sports match outcome, the PBA has an accuracy of 82.86%, precision has 80.95%, recall has 89.47%, and the F1 score has 85.00%, while in the PVL it has an accuracy of 90.20%, precision has 88.46%, recall has 92.00%, and the F1 score has 90.20%. based on Table 4.1.

5.2 Recommendation

Integrating a machine learning model with a rule-based method gives future research prospects in the realm of sports match prediction, where precision and accuracy are critical. Our research highlights the potential benefits of this hybrid model, which integrates machine learning predictive capability with rule-based system interpretability. This study only used a machine learning model using random forest regression to predict match results for a specific group of PBA and PVL teams and sports, such as basketball and volleyball. We recommend that researchers and sport analysts continue to investigate this approach, focusing on model optimization and adaptation, applying models to real-world scenarios, studying other sports and machine learning models.

References

[1] The Sporting Blog. (2023, May 16). The most popular sports in the Philippines. April 12, 2023 https://thesporting.blog/blog/most-popular-sports-in-the-Philippines

[2] Gu, Z., & He, C. (2021). Application of Fuzzy Decision Tree Algorithm Based on Mobile Computing in Sports Fitness Member Management. Wireless Communications and Mobile Computing. Retrieved October 29, 2023. https://doi.org/10.1155/2021/4632722

[3] Batra, S. (2017, April 16). The Role of Data Analytics in Modern Day Sports. Researchgate. Retrieved October 29, 2023, from https://www.researchgate.net/publication/3305 78956_The_Role_of_Data_Analytics_in_Mod ern_Day_Sports

[4] Sankaranarayanan, V. V., Sattar, J., & Lakshmanan, L. V. S. (2014). Auto-play: A Data Mining Approach to ODI Cricket Simulationand Prediction. https://doi.org/10.1137/1.9781611973440.121

[5] Maulud, D. H., & Abdulazeez, A. M. (2020). A review on Linear Regression comprehensive in Machine Learning. Journal of Applied Science and Technology Trends, 1(4), 140–147. Retrieved October 29, 2023 https://doi.org/10.38094/jastt1457

[6] H., Bai, Z., & Bai, X. (2021). Sports Big Data: Management, Analysis, Applications, and Challenges. Sports Big Data: Management, Analysis, Applications, and Challenges. Retrieved October 29, 2023https://doi.org/10.1155/2021/6676297

[7] McCabe, A., & Trevathan, J. (n.d.). Artificial Intelligence in Sports Prediction. (PDF) Artificial Intelligence in Sports Prediction | Jarrod Trevathan - Academia.edu. Retrieved September 12, 2023 https://www.academia.edu/3165997/Artificial_Intelligence_in_Sports_Prediction

[8] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). April 12, 2023 https://doi.org/10.1023/A:1010933404324

[9] Zhao, K., Chun-Jie, D., & Tan, G. (2023).

[10] Enhancing Basketball Game Outcome Prediction through Fused Graph Convolutional Networks and Random Forest Algorithm. Entropy, 25(5), 765. September 12, 2023 https://doi.org/10.3390/e25050765

[11] Biau, G. (2010). Analysis of a random forests model. HAL (Le Centre Pour La Communication Scientifique Directe). September 12,2023 https://hal.archives-ouvertes.fr/hal-00476545

[12] Haghighat, M., Rastegari, H., & Nourafza, N. (2013). A Review of Data Mining Techniques for Result Prediction in Sports. April 12,2023 https://www.researchgate.net/publication/2590 82292_A_Review_of_Data_Mining_Techniques_for_Result_Prediction_in_Sports

[13] Lam, M. W. (2018, February). One-Match-Ahead Forecasting in Two-Team Sports with Stacked Bayesian Regressions. Sciendo. Retrieved October 29, 2023, from https://sciendo.com/article/10.1515/jaiscr-201 8-0011

[14] Lotfi, S., & Rebbouj, M. (2021, July). Machine Learning for sport results prediction using algorithms. Researchgate. Retrieved October 29, 2023, from https://www.researchgate.net/publication/3536 65925_Machine_Learning_for_sport_results_prediction_using_algorithms

[15] Passi, K., & Pandey, N. (2018). Increased Prediction Accuracy in the Game of Cricket Using Machine Learning. International Journal of Data Mining & Knowledge Management Process, 8(2), 19–36. April 12,2023 https://doi.org/10.5121/ijdkp.2018.8203