PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, **Number** 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

Championship Probability Prediction in the FIFA ASEAN Cup 2026 Challenge Division Using Logistic Regression

Mohamad Iqbal Ulumando
Informatics Study Program, Institut Teknologi Alberth Foenay Kupang, Indonesia
iqbal77ulumando@gmail.com

open access

***Corresponding Author**

Article History:
Submitted: 19-09-2026
Accepted: 23-09-2026
Published: 04-01-2027

Keywords:
FIFA ASEAN Cup; Championship Prediction; Logistic Regression; Football Performance; Machine Learning.

PERFECT: Journal of Smart Algorithms is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

This study aims to develop a Logistic Regression model for predicting the championship probability of teams participating in the FIFA ASEAN Cup 2026 Challenge Division. The study employs a quantitative approach using five performance indicators, namely FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference. Historical data from 2020–2025 were used to develop and evaluate the prediction model, while data from 2026 were used to estimate the championship probability of the participating teams. The historical dataset consisted of 36 observations and was divided into training and testing datasets using an 80:20 stratified split. Prior to modeling, the predictor variables were prepared using data preprocessing and standardization to ensure that the variables were suitable for Logistic Regression. Model performance was evaluated using Accuracy, Precision, Recall, and F1-score. The results showed that the Logistic Regression model achieved an Accuracy of 62.50%, Precision of 50.00%, Recall of 66.67%, and F1-score of 57.14% on the testing dataset. The prediction results indicated that Hong Kong obtained the highest championship probability of 77.53%, followed by Myanmar at 68.59% and Cambodia at 53.06%. Laos recorded 28.01%, while Timor-Leste and Brunei Darussalam obtained 0.72% and 0.71%, respectively. These results indicate that Logistic Regression provides an interpretable machine learning approach for estimating championship probability based on multiple football performance indicators and can support quantitative analysis of team championship potential in the FIFA ASEAN Cup 2026 Challenge Division.

INTRODUCTION

Football is one of the most widely followed sports in Southeast Asia and has become an important component of regional sporting competition. The development of football in the ASEAN region has encouraged national teams to improve their competitive performance through systematic preparation, performance evaluation, and data-based analysis. In this context, statistical and machine learning approaches can provide useful tools for analyzing historical team performance and estimating future competition outcomes (Fadilah, 2024). The FIFA ASEAN Cup 2026 Challenge Division provides an interesting context for predictive analysis because teams participating in this division have different levels of international experience and competitive performance. Differences in team quality can be reflected through measurable indicators such as FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference (Atmaja, 2026).

These indicators provide quantitative information that can be used to characterize the competitive strength of each team and estimate its potential to become the tournament champion. Previous studies have demonstrated the application of machine learning techniques to football prediction by utilizing historical match results and team-performance indicators (Rusdianto Hidayat et al., 2024). Classification-based methods are particularly useful when the prediction target consists of two possible outcomes, such as Champion and Non-Champion. However, the application of Logistic Regression to estimate championship probability in an ASEAN regional competition, particularly the Challenge Division, remains relatively limited. Logistic Regression is appropriate for this problem because it produces a probability value for a binary outcome and allows the contribution of multiple predictor variables to be considered simultaneously (Ulumando, 2026).

Therefore, this study aims to develop a Logistic Regression model to predict the championship probability of teams participating in the FIFA ASEAN Cup 2026 Challenge Division. The model uses five performance indicators, namely FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference, as predictor variables, while the championship status is used as the target variable (Ulumando, 2026). Historical tournament data are used to train and evaluate the model before it is applied to the participating teams in the 2026 competition. The results of this study are expected to provide a quantitative ranking of the championship potential of Challenge Division teams

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, Number 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

and demonstrate the usefulness of Logistic Regression as an interpretable machine learning approach for football championship prediction (Gifford & Bayrak, 2023). The prediction can also serve as an analytical reference for evaluating team competitiveness based on measurable historical performance indicators.

LITERATURE REVIEW

Football Prediction

Football prediction refers to the use of statistical methods, mathematical models, and machine learning techniques to estimate the potential outcomes of football competitions based on historical and current performance data. The increasing availability of football statistics has enabled researchers to develop predictive models using measurable indicators such as team ranking, winning performance, goals scored, goals conceded, and goal difference. Machine learning has become an important approach in football prediction because it can identify patterns within historical data and use those patterns to classify or estimate future outcomes. Previous research has demonstrated that predictive models can be applied to different football-related problems, including match-result prediction, team-performance classification, and tournament outcome estimation (Ulumando, 2026). The effectiveness of these approaches depends strongly on the quality of the dataset, the selection of relevant predictor variables, and the suitability of the classification algorithm. In the context of this study, football prediction is formulated as a binary classification problem in which teams are categorized as Champion or Non-Champion (Dewi & Latjompoh, 2021). The classification result is subsequently used to estimate the relative championship probability of teams participating in the FIFA ASEAN Cup 2026 Challenge Division.

Football Performance Indicators

Football team performance can be represented through several quantitative indicators that describe the competitive characteristics of a team. In this study, five indicators are used: FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference. FIFA Ranking provides an international measure of a national team's competitive position (Apryani et al., 2025). A relatively higher-ranked team generally indicates stronger historical and recent competitive performance. Win Rate represents the proportion of matches won by a team and provides an indication of its ability to achieve positive results. Average Goals Scored reflects attacking effectiveness, whereas Average Goals Conceded represents defensive performance. Goal Difference combines attacking and defensive outcomes and provides an overall indication of the difference between goals scored and goals conceded. The combination of these indicators provides a multidimensional representation of team performance. Rather than relying on a single statistic, the proposed model considers the joint characteristics of the five indicators to estimate the likelihood that a team belongs to the Champion class (Said, 2025).

Logistic Regression Algorithm

Logistic Regression is a supervised machine learning algorithm commonly used for binary classification problems. Unlike linear regression, which predicts a continuous numerical value, Logistic Regression estimates the probability that an observation belongs to a particular class (Pratama & Dahlan, 2025). This characteristic makes it suitable for predicting whether a football team is classified as Champion (1) or Non-Champion (0). The Logistic Regression model estimates the relationship between the predictor variables and the probability of belonging to the target class (Ulumando, 2025). In this study, FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference are used as independent variables, while championship status is used as the dependent variable (Randa & Lubis, 2025). The probability of a team belonging to the Champion class is estimated using the logistic or sigmoid function:

$$P(Y = 1|X) = \frac{1}{1+e^{-z}} \quad (1)$$

Where :

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (2)$$

where $P(Y = 1|X)$ represents the predicted probability of a team being classified as Champion, β_0 is the intercept, $\beta_1, \dots\dots, \beta_5$ are the model coefficients, and $X_1, \dots\dots, X_5$ represent FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference, respectively.

The output of the sigmoid function ranges from 0 to 1 and can therefore be interpreted as a probability. A probability closer to 1 indicates a higher likelihood of being classified as Champion, whereas a probability closer to 0 indicates a higher likelihood of being classified as Non-Champion. In this study, the probability value is converted into a percentage to facilitate comparison among the participating teams. One of the main advantages of Logistic Regression

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, **Number** 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

is its interpretability (Sulistya & Pakereng, 2026). The model can provide probability estimates that allow the championship potential of different teams to be compared directly. This is particularly relevant to the objective of this study, which is not only to classify teams but also to rank them according to their estimated championship probability. Logistic Regression has also been widely applied in sports analytics because of its relatively simple structure, ability to incorporate multiple predictor variables, and suitability for binary outcome prediction (Aziz & Sulistiani, 2026). Therefore, Logistic Regression provides an appropriate methodological foundation for estimating the championship probability of FIFA ASEAN Cup 2026 Challenge Division teams.

Previous Studies and Research Gap

Previous studies have shown that machine learning and statistical classification methods can effectively utilize football performance data to predict match outcomes and team competitiveness. Research on football prediction has employed various algorithms, including Logistic Regression, Naïve Bayes, decision trees, support vector machines, and neural networks. These studies generally demonstrate that historical performance indicators can provide meaningful information for predicting football outcomes (Wisswani et al., 2020). Berrar, Lopes, and Dubitzky developed a data- and knowledge-driven framework for machine learning models in soccer outcome prediction and emphasized the importance of appropriate feature selection, data preparation, and model evaluation in football prediction. Other studies have also demonstrated the applicability of probabilistic and classification-based approaches for estimating football match outcomes using historical performance characteristics (Ulumando, 2026).

However, most existing studies focus on individual match outcomes, such as win, draw, or loss, rather than directly estimating the championship probability of teams in a regional tournament division. Furthermore, research specifically examining the FIFA ASEAN Cup Challenge Division using Logistic Regression remains limited. This creates an opportunity to apply an interpretable binary classification model to estimate the relative championship potential of participating teams (Gifford & Bayrak, 2023). Therefore, this study addresses the research gap by developing a Logistic Regression-based championship prediction model for the FIFA ASEAN Cup 2026 Challenge Division. The study combines five measurable team-performance indicators and produces a probability-based ranking of participating teams. This approach provides a more direct assessment of championship potential and complements previous research that has primarily focused on individual match prediction (Ulumando, 2026).

METHOD

Research Design

This study employed a quantitative predictive research design using a supervised machine learning approach to estimate the championship probability of teams participating in the FIFA ASEAN Cup 2026 Challenge Division. The proposed prediction model uses the Logistic Regression algorithm, which is suitable for binary classification and provides probability estimates for the occurrence of a particular class. The research process consisted of several stages, including historical data collection, data preprocessing, dataset preparation and splitting, Logistic Regression model development, model evaluation, and prediction of championship probability for the FIFA ASEAN Cup 2026 Challenge Division. Historical team performance data were used to train and evaluate the model, while the resulting model was subsequently applied to the 2026 Challenge Division teams. The target variable was formulated into two classes: Champion (1) and Non-Champion (0). The resulting probability from the Logistic Regression model was used to estimate the likelihood of each team becoming champion and to produce a ranking based on the predicted championship probability. The complete research procedure is summarized in Figure 1, which illustrates the process from historical data collection and preprocessing through model development, evaluation, and prediction of the FIFA ASEAN Cup 2026 Challenge Division championship probability.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS
E-ISSN: E-ISSN: 3064-0377
Volume 4, Number 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

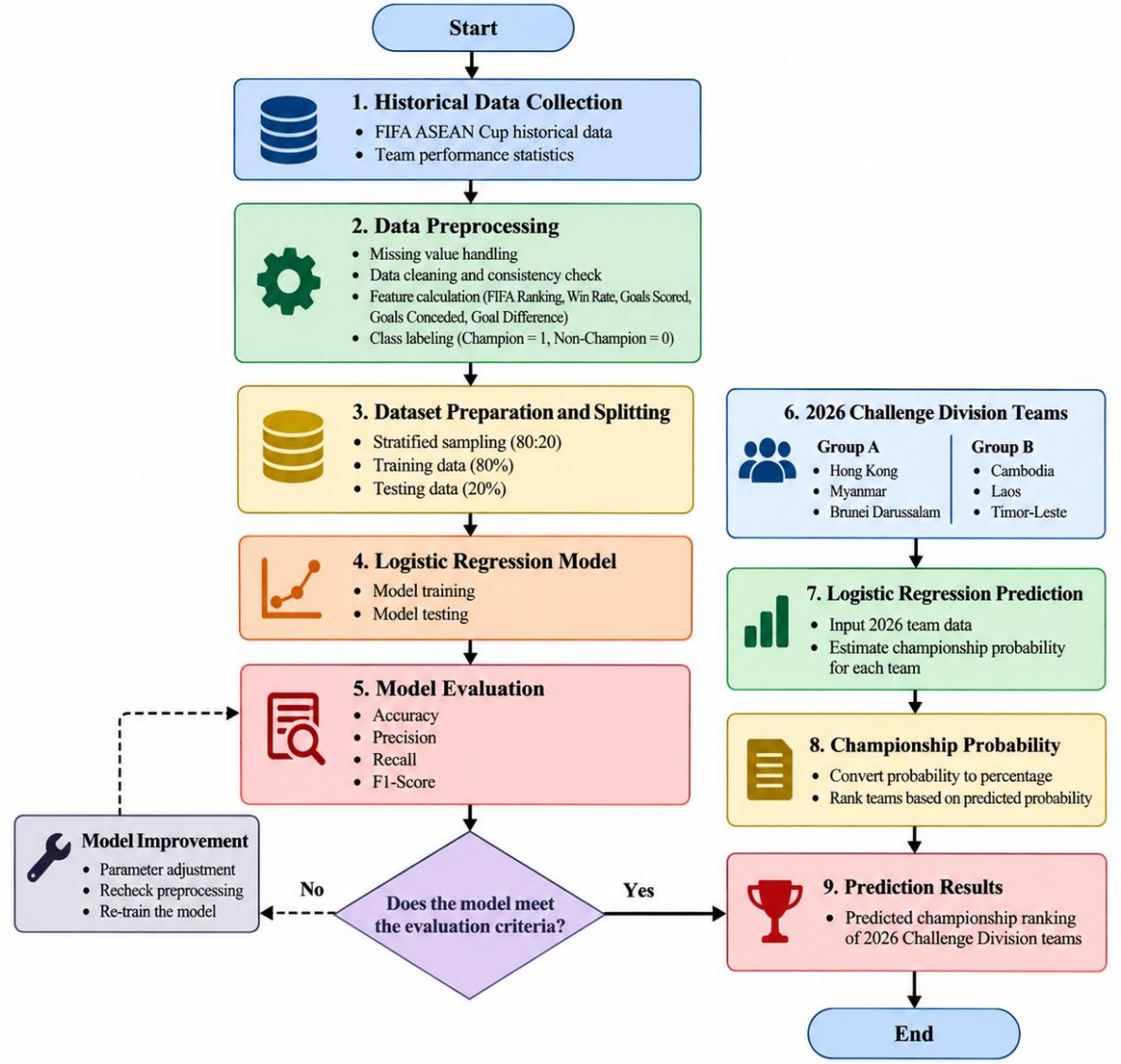


Figure 1. Research Flow of the Proposed Study

Data Collection

The data used in this study consist of historical football performance data and team characteristics relevant to the prediction of championship probability. The historical data were collected from relevant football statistical sources and FIFA-related information, including team ranking and match performance records. Five performance indicators were selected as predictor variables, namely FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference. These variables were selected because they represent different aspects of team performance, including international ranking, winning consistency, offensive capability, defensive performance, and overall goal performance. For the prediction stage, the study focuses on the six teams participating in the FIFA ASEAN Cup 2026 Challenge Division. The teams are divided into two groups:

Group A

1. Hong Kong
2. Myanmar
3. Brunei Darussalam

Group B

1. Cambodia
2. Laos
3. Timor-Leste

Historical observations were used as the training and testing dataset, while the 2026 team data were prepared

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, Number 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

separately as input for the final prediction.

Research Variables

The research variables consist of five independent variables and one dependent variable. The independent variables represent the performance characteristics of each national team, while the dependent variable represents the championship status.

Table 1. Research Variables

Variable	Description	Type
X1	FIFA Ranking	Independent Variable
X2	Win Rate	Independent Variable
X3	Average Goals Scored	Independent Variable
X4	Average Goals Conceded	Independent Variable
X5	Goal Difference	Independent Variable
Y	Championship Status	Dependent Variable

The target variable YY is represented using binary classification:

- 1 = Champion
- 0 = Non-Champion

The FIFA Ranking represents the relative international position of a national team. Win Rate represents the proportion of matches won by a team during the observation period. Average Goals Scored measures offensive productivity, while Average Goals Conceded represents defensive vulnerability. Goal Difference is calculated from the difference between average goals scored and average goals conceded. These five variables are subsequently used as input features for the Logistic Regression model.

Data Preprocessing

Data preprocessing was conducted to ensure that the dataset was suitable for model development and prediction. The preprocessing stage consisted of several activities, including data cleaning, missing-value checking, consistency checking, feature calculation, and class labeling. First, the collected data were examined to identify missing, duplicated, or inconsistent records. Records with incomplete information were handled according to the availability and consistency of the corresponding historical data. Second, the required performance indicators were calculated or organized from the available match records. Win Rate was obtained from the proportion of matches won, while Average Goals Scored and Average Goals Conceded were calculated based on the number of goals scored and conceded over the observation period. Goal Difference was obtained from the difference between goals scored and goals conceded. The target variable was then assigned based on the championship status of each historical observation. A value of 1 was assigned to teams classified as champions, whereas a value of 0 was assigned to non-champion teams. After preprocessing, the resulting dataset contained numerical predictor variables that could be processed by the Logistic Regression algorithm.

Dataset Splitting

The prepared historical dataset was divided into training data and testing data to evaluate the predictive performance of the Logistic Regression model. An 80:20 data splitting ratio was used, where approximately 80% of the observations were allocated as training data and 20% as testing data. To maintain the proportion of Champion and Non-Champion classes in both subsets, stratified sampling was applied. The training dataset was used to estimate the Logistic Regression model parameters, while the testing dataset was used to evaluate the model's ability to classify observations that were not included during training. The 2026 Challenge Division team data were not included in the training or testing process. Instead, they were retained as separate prediction data and were only introduced after the model had been evaluated.

Logistic Regression Model

Logistic Regression was employed as the main prediction algorithm because the target variable consists of two classes, namely Champion and Non-Champion. The algorithm estimates the probability that an observation belongs to the Champion class based on the values of the five predictor variables. The Logistic Regression model uses the following predictors:

- FIFA Ranking (X1X_1)
- Win Rate (X2X_2)
- Average Goals Scored (X3X_3)
- Average Goals Conceded (X4X_4)

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, **Number** 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

- Goal Difference (X5X_5)

The output of the model is a probability value ranging from 0 to 1. This probability represents the estimated likelihood that a team belongs to the Champion class. For classification purposes, the predicted class can be determined based on the probability threshold. However, in this study, the probability value itself is retained because the main objective is not only to classify teams but also to estimate and compare their championship probabilities. The trained Logistic Regression model was subsequently applied to the six FIFA ASEAN Cup 2026 Challenge Division teams. The predicted probabilities were converted into percentages and used to establish the championship probability ranking.

Model Evaluation

The performance of the Logistic Regression model was evaluated using the testing dataset. Four commonly used classification metrics were employed:

1. Accuracy
2. Precision
3. Recall
4. F1-Score

Accuracy measures the proportion of correctly classified observations among all testing observations. Precision measures the proportion of observations predicted as Champion that are actually Champion. Recall measures the ability of the model to correctly identify Champion observations. The F1-Score represents the harmonic mean of Precision and Recall.

These evaluation metrics were used to determine whether the trained Logistic Regression model provided an acceptable classification performance before being applied to the 2026 Challenge Division prediction. If the model did not meet the predetermined evaluation criteria, the preprocessing and model training stages could be reviewed and improved. If the model achieved satisfactory performance, the model was used for the championship probability prediction stage.

Prediction of Championship Probability

After the Logistic Regression model had been trained and evaluated, the model was applied to the performance data of the six teams participating in the FIFA ASEAN Cup 2026 Challenge Division. The prediction process consisted of several steps. First, the five predictor variables for each 2026 team were prepared using the same preprocessing procedure applied to the historical dataset. Second, the processed team data were entered into the trained Logistic Regression model. Third, the model generated a probability value representing the estimated likelihood of each team being classified as Champion. The resulting probability values were converted into percentages to facilitate interpretation. The six teams participating in the 2026 Challenge Division are distributed into two groups, as presented in Table 2.

Table 2. FIFA ASEAN Cup 2026 Challenge Division Teams by Group

Group	Teams
Group A	Hong Kong, Myanmar, Brunei Darussalam
Group B	Cambodia, Laos, Timor-Leste

Table 2 shows that the FIFA ASEAN Cup 2026 Challenge Division consists of six teams divided into two groups. Group A comprises Hong Kong, Myanmar, and Brunei Darussalam, while Group B consists of Cambodia, Laos, and Timor-Leste. Each team serves as an object of prediction in the Logistic Regression model based on the five performance indicators used in this study. The grouping provides the basis for identifying the teams included in the subsequent championship probability analysis. The resulting prediction probabilities were then used to compare the estimated championship prospects of the six participating teams. The teams were ranked according to their predicted championship probability, from the highest to the lowest value.

RESULT

Dataset Preparation

The dataset used in this study consists of historical performance data from six national teams participating in the FIFA ASEAN Cup 2026 Challenge Division, namely Hong Kong, Myanmar, Brunei Darussalam, Cambodia, Laos, and Timor-Leste. The dataset contains five predictor variables, namely FIFA Ranking (X1), Win Rate (X2), Average Goals Scored (X3), Average Goals Conceded (X4), and Goal Difference (X5). The initial dataset contains observations covering the period from 2020 to 2026, with six observations for each year, resulting in a total of 42 observations. Each observation represents the performance characteristics of one team in a particular year. The dataset was subsequently

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, Number 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

examined and organized to ensure that the team and year information were consistently represented. For the purposes of this study, the historical observations from 2021 to 2025 are used as the historical dataset for model development, while the 2026 observations are retained as prediction data for the FIFA ASEAN Cup 2026 Challenge Division. This separation is intended to prevent the 2026 observations from being included in the model training and testing process before the prediction stage. Accordingly, the historical dataset consists of 30 observations, representing six teams across five years (2021–2025). The 2026 prediction dataset consists of 6 observations, representing the six Challenge Division teams. The distribution of observations is presented in Table 3.

Table 3. Distribution of the Dataset by Year

Year	Number of Teams	Data Usage
2021	6	Historical Data
2022	6	Historical Data
2023	6	Historical Data
2024	6	Historical Data
2025	6	Historical Data
2021–2025	30	Model Training and Testing
2026	6	Prediction Data
Total	**36**	**Historical and Prediction Data**

The five predictor variables were maintained in numerical form to facilitate their use in the Logistic Regression model. FIFA Ranking represents the international ranking position of each team, while Win Rate represents the percentage of matches won. Average Goals Scored and Average Goals Conceded represent the average number of goals scored and conceded per match, respectively. Goal Difference represents the goal-performance differential recorded for each team. The resulting dataset therefore provides the input characteristics required for the subsequent descriptive analysis, dataset splitting, Logistic Regression model development, model evaluation, and prediction of championship probability for the FIFA ASEAN Cup 2026 Challenge Division.

Descriptive Statistics

Descriptive statistical analysis was conducted to provide an overview of the characteristics and distribution of the five predictor variables used in the Logistic Regression model. The analysis was performed on the 30 historical observations covering the 2021–2025 period, consisting of six teams observed across five years. The variables analyzed were FIFA Ranking (X1), Win Rate (X2), Average Goals Scored (X3), Average Goals Conceded (X4), and Goal Difference (X5). The descriptive statistics include the mean, standard deviation, minimum, and maximum values of each predictor variable. The results are presented in Table 4.

Table 4. Descriptive Statistics of Historical Data (2021–2025)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
FIFA Ranking (X1)	30	177.27	17.14	146.00	200.00
Win Rate (X2)	30	0.73%	0.64%	0.07%	2.67%
Average Goals Scored (X3)	30	1.51	0.81	0.11	3.20
Average Goals Conceded (X4)	30	7.62	3.54	3.21	15.95
Goal Difference (X5)	30	-4.49	3.97	-13.94	1.90

Based on Table 4, the FIFA Ranking (X1) has a mean value of 177.27, with values ranging from 146.00 to 200.00. The variation indicates differences in the international ranking positions of the six teams throughout the observation period. Lower FIFA ranking numbers indicate a relatively stronger international ranking position. The Win Rate (X2) has a mean value of 0.73%, with a minimum of 0.07% and a maximum of 2.67%. The relatively wide variation between the minimum and maximum values indicates differences in winning performance among the observed teams. The values in this table follow the percentage representation provided in the dataset. The Average Goals Scored (X3) has a mean of 1.51 goals per match, ranging from 0.11 to 3.20 goals per match. This indicates considerable variation in offensive productivity among the teams. Teams with higher values demonstrate a greater average ability to score goals during the observation period. The Average Goals Conceded (X4) has a mean of 7.62 goals per match, with values ranging from 3.21 to 15.95. This variable reflects differences in defensive performance. Higher values indicate that a team conceded more goals on average, representing relatively weaker defensive performance.

Meanwhile, the Goal Difference (X5) has a mean value of −4.49, with a minimum of −13.94 and a maximum of 1.90. The negative mean indicates that, overall, the historical observations in the dataset tend to have a negative goal differential. Only observations with positive values indicate that the average goals scored exceeded the corresponding goal-conceded measure. Overall, the descriptive statistics demonstrate that the six Challenge Division teams exhibit

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, Number 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

substantial variation across the five predictor variables. These variations provide the basis for distinguishing team performance characteristics within the Logistic Regression model. The historical data are subsequently divided into training and testing subsets before model development.

Dataset Splitting

The historical dataset consisted of 30 observations covering the 2021–2025 period. The dataset was divided into training and testing subsets using an 80:20 ratio. The training dataset was used to develop the Logistic Regression model, while the testing dataset was used to evaluate its predictive performance. Based on the specified proportion, 24 observations (80%) were allocated to the training dataset and 6 observations (20%) were allocated to the testing dataset. The observations from 2026 were not included in the training or testing datasets because they were specifically reserved for the prediction stage. The dataset distribution is presented in Table 5.

Table 5. Dataset Splitting for Logistic Regression

Dataset	Period	Number of Observations	Proportion	Data Usage
Training Dataset	2021–2025	24	80%	Model Training
Testing Dataset	2021–2025	6	20%	Model Evaluation
Prediction Dataset	2026	6	Not included in training/testing	2026 Championship Prediction
Total Dataset	**2021–2026**	**36**	**100% of available observations**	**Model Development and Prediction**

The 24 training observations were used to estimate the Logistic Regression model based on the five predictor variables, namely FIFA Ranking (X1), Win Rate (X2), Average Goals Scored (X3), Average Goals Conceded (X4), and Goal Difference (X5). The resulting model was then evaluated using the 6 testing observations. The 6 observations from 2026 were retained as an independent prediction dataset. These observations represent the six Challenge Division teams: Hong Kong, Myanmar, Brunei Darussalam, Cambodia, Laos, and Timor-Leste. Since the championship status of these teams is the subject of prediction, the 2026 observations are not assigned a known target value during the prediction stage. The separation of historical and prediction data ensures that the 2026 observations do not influence the estimation or evaluation of the Logistic Regression model. Therefore, the model developed from the 2021–2025 historical data can subsequently be applied to the 2026 team characteristics to estimate their respective championship probabilities.

Logistic Regression Model Results

The Logistic Regression model was developed using five predictor variables, namely FIFA Ranking (X1), Win Rate (X2), Average Goals Scored (X3), Average Goals Conceded (X4), and Goal Difference (X5). The target variable was Championship Status (Y), consisting of two classes: 0 for Non-Champion and 1 for Champion. The historical dataset from 2020–2025 was used for model development, while the 2026 data were reserved for the prediction stage. After preprocessing, the historical dataset consisted of 36 observations. The dataset was divided into 28 training observations (80%) and 8 testing observations (20%). Standardization was applied to the predictor variables before model training.

Table 6. Logistic Regression Model Configuration

Component	Description
Algorithm	Logistic Regression
Predictor Variables	X1, X2, X3, X4, X5
Target Variable	Championship Status (Y)
Class 0	Non-Champion
Class 1	Champion
Historical Data	2020–2025
Training Data	28 observations (80%)
Testing Data	8 observations (20%)
Scaling	Standardization
Prediction Data	2026 Challenge Division

The resulting Logistic Regression model produced the coefficient values presented in Figure 2. The model coefficients were −1.1815 for FIFA Ranking (X1), −0.4136 for Win Rate (X2), 0.1343 for Average Goals Scored (X3), −0.2402 for Average Goals Conceded (X4), and 0.8715 for Goal Difference (X5). The intercept value was −0.9942.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS
E-ISSN: E-ISSN: 3064-0377
Volume 4, **Number** 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

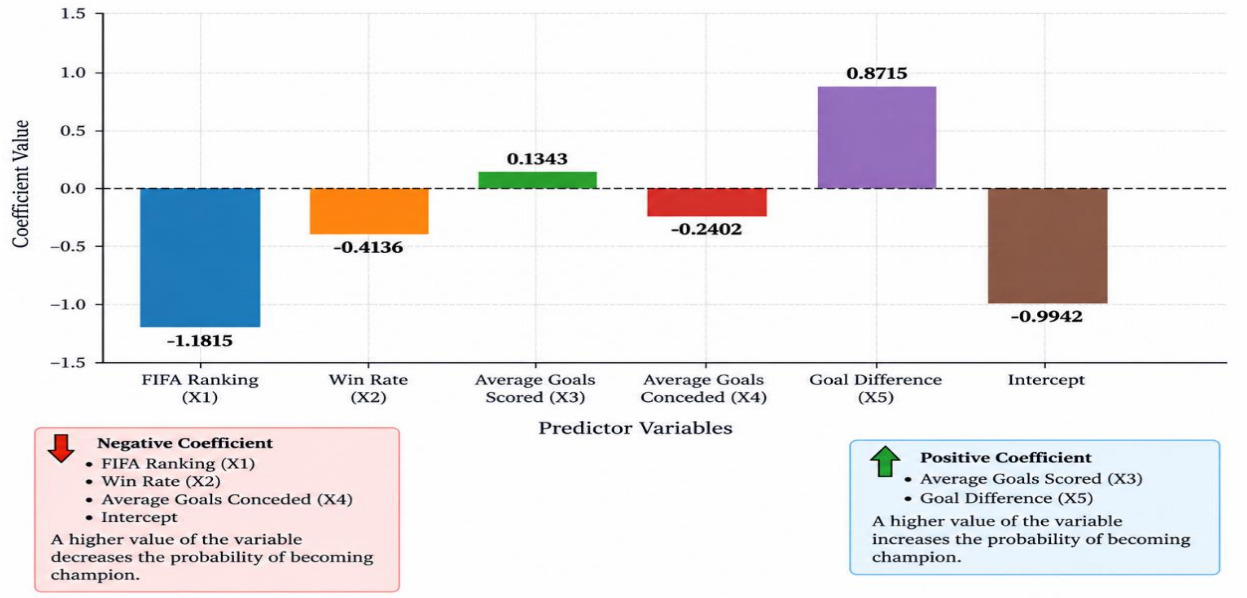


Figure 2. Logistic Regression Model Coefficients

The coefficient results show that Goal Difference (X5) had the strongest positive coefficient, with a value of 0.8715. This indicates a positive contribution of goal difference to the predicted probability of being classified as Champion. Average Goals Scored (X3) also had a positive coefficient of 0.1343. Meanwhile, FIFA Ranking (X1) had the largest negative coefficient at −1.1815, followed by Win Rate (X2) at −0.4136 and Average Goals Conceded (X4) at −0.2402. These coefficient signs indicate the direction of the relationship between each standardized predictor and the model's predicted Champion class. The trained Logistic Regression model was subsequently applied to the testing dataset to assess its classification performance. The results of the evaluation, including Accuracy, Precision, Recall, F1-score, and the confusion matrix, are presented in Section Evaluation of the Prediction Model.

Evaluation of the Prediction Model

The performance of the Logistic Regression model was evaluated using the testing dataset consisting of eight observations. Four standard classification metrics were used to measure the model performance: Accuracy, Precision, Recall, and F1-score. The evaluation results are presented in Figure 3.

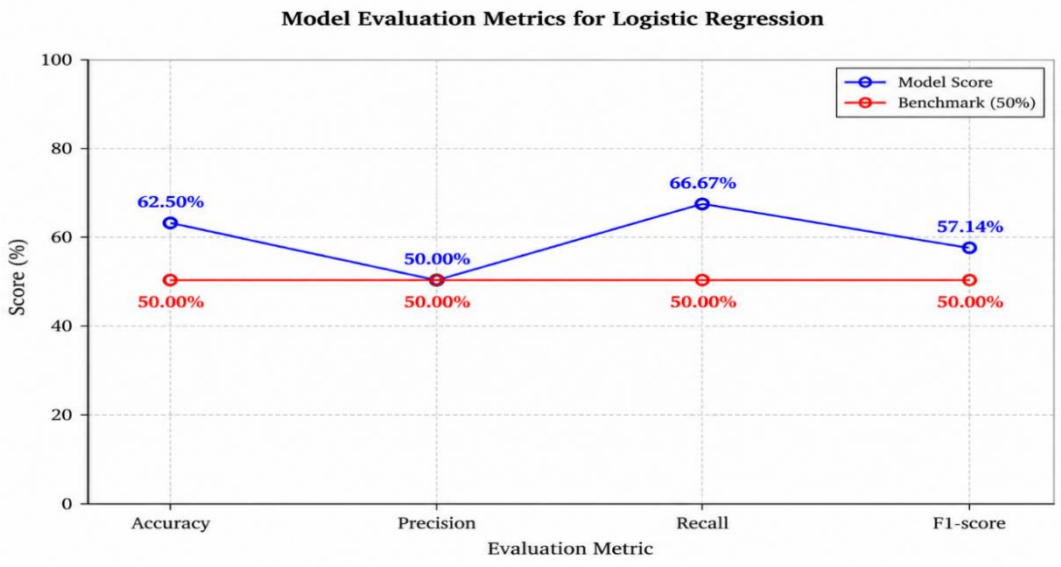


Figure 3. Evaluation Metrics of the Logistic Regression Model

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, Number 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

The model achieved an Accuracy of 62.50%, indicating that five out of eight testing observations were correctly classified. The Precision was 50.00%, showing that half of the observations predicted as Champion were correctly classified. Meanwhile, the model obtained a Recall of 66.67%, indicating that the model correctly identified two-thirds of the actual Champion observations. The resulting F1-score was 57.14%, reflecting the balance between Precision and Recall.

Table 7. Logistic Regression Model Evaluation Results

Evaluation Metric	Result
Accuracy	62.50%
Precision	50.00%
Recall	66.67%
F1-score	57.14%

The confusion matrix obtained from the testing process is presented in Table 8.

Table 8. Confusion Matrix of the Logistic Regression Model

Actual Class	Predicted Non-Champion	Predicted Champion
Non-Champion	3	2
Champion	1	2

The confusion matrix shows that the model correctly classified three Non-Champion observations and two Champion observations. However, two Non-Champion observations were incorrectly classified as Champion, while one Champion observation was incorrectly classified as Non-Champion. Overall, the Logistic Regression model achieved a 62.50% accuracy on the testing dataset. The Recall value of 66.67% indicates that the model was relatively better at identifying Champion observations than avoiding false Champion predictions, as reflected by the Precision value of 50.00%. The F1-score of 57.14% indicates a moderate balance between Precision and Recall. These results indicate that the Logistic Regression model can be used as a predictive approach for estimating championship probability. The trained model is therefore applied to the six teams participating in the FIFA ASEAN Cup 2026 Challenge Division. The resulting championship probabilities and team rankings are presented in the next section.

Championship Probability 2026

The trained Logistic Regression model was applied to the performance data of the six teams participating in the FIFA ASEAN Cup 2026 Challenge Division. The prediction used five performance indicators, namely FIFA Ranking (X1), Win Rate (X2), Average Goals Scored (X3), Average Goals Conceded (X4), and Goal Difference (X5). The model generated a championship probability for each team based on its 2026 performance indicators. The resulting probabilities were then used to establish a ranking of the teams according to their estimated championship potential.

Table 9. Championship Probability of FIFA ASEAN Cup 2026 Challenge Division Teams

Rank	Team	Championship Probability
1	Hong Kong	77.53%
2	Myanmar	68.59%
3	Cambodia	53.06%
4	Laos	28.01%
5	Timor-Leste	0.72%
6	Brunei Darussalam	0.71%

The results show that Hong Kong obtained the highest estimated championship probability of 77.53%, placing the team in first position. Myanmar followed in second place with a probability of 68.59%. Cambodia ranked third with 53.06%, while Laos obtained a substantially lower probability of 28.01%. Timor-Leste and Brunei Darussalam recorded the lowest estimated championship probabilities, with 0.72% and 0.71%, respectively. These results indicate that, according to the Logistic Regression model, Hong Kong and Myanmar have the strongest estimated potential to become champions of the 2026 Challenge Division based on the five predictor variables.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, **Number** 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

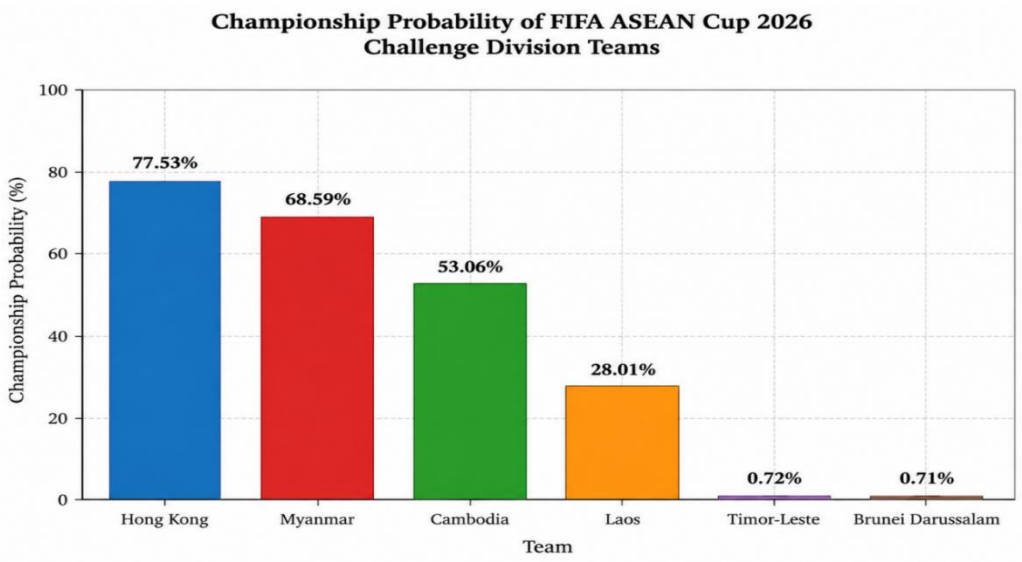


Figure 4. Championship Probability of FIFA ASEAN Cup 2026 Challenge Division Teams

The championship probability results provide a quantitative ranking of the six participating teams. The relatively high probabilities obtained by Hong Kong, Myanmar, and Cambodia indicate stronger model-based championship potential, whereas the considerably lower probabilities of Laos, Timor-Leste, and Brunei Darussalam indicate weaker predicted championship potential. It should be noted that the probability values represent the estimated probability of the Champion class for each team. Therefore, the values are interpreted individually and do not necessarily have to sum to 100%.

DISCUSSION

The results of this study demonstrate that Logistic Regression can be used to estimate the championship probability of teams in the FIFA ASEAN Cup 2026 Challenge Division based on multiple performance indicators. The coefficient analysis shows that Goal Difference (X5) had the strongest positive coefficient at 0.8715, followed by Average Goals Scored (X3) with a coefficient of 0.1343. These positive coefficients indicate that, within the fitted model, higher values of these variables are associated with an increased probability of being classified as Champion. Goal Difference therefore represents an important predictor in distinguishing teams with stronger overall goal performance. This finding is consistent with the role of goal-related indicators in representing the balance between offensive and defensive performance.

FIFA Ranking (X1) produced the largest negative coefficient, with a value of −1.1815, while Win Rate (X2) and Average Goals Conceded (X4) also produced negative coefficients of −0.4136 and −0.2402, respectively. The negative coefficient for FIFA Ranking can be interpreted in relation to the ranking scale, where a lower numerical ranking represents a stronger international position. Thus, the negative coefficient is consistent with the expectation that a better FIFA Ranking position is associated with greater championship potential. In contrast, the negative coefficients for Win Rate and Average Goals Conceded should be interpreted cautiously because the model was developed from a relatively small historical dataset. Relationships among predictor variables and the limited number of observations may influence the direction and magnitude of individual coefficients. Therefore, these coefficients describe the relationships captured by the fitted model rather than establishing independent causal effects.

The model evaluation results indicate moderate classification performance, with an Accuracy of 62.50%, Precision of 50.00%, Recall of 66.67%, and F1-score of 57.14%. The confusion matrix shows that five of eight testing observations were correctly classified, consisting of three Non-Champion and two Champion observations. However, two Non-Champion observations were classified as Champion and one Champion observation was classified as Non-Champion. The higher Recall than Precision indicates that the model was relatively more successful in identifying Champion observations than in ensuring that all observations predicted as Champion were actually Champion. This result is important when interpreting the championship probabilities because the model retains some uncertainty in distinguishing the two classes.

The 2026 prediction results provide a quantitative comparison of the estimated championship potential of the six Challenge Division teams. Hong Kong obtained the highest reported probability at 77.53%, followed by Myanmar at 68.59% and Cambodia at 53.06%, while Laos, Timor-Leste, and Brunei Darussalam obtained substantially lower values. These results indicate that the selected performance indicators can be used to differentiate the estimated

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, **Number** 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

championship potential of the participating teams. However, the probability values should be interpreted as model-based estimates rather than definitive predictions of the tournament winner. In addition, because the probability generated by Logistic Regression represents the probability of belonging to the Champion class for each observation, the values are interpreted individually and do not necessarily sum to 100%.

The findings also highlight the importance of data characteristics in applying Logistic Regression to football championship prediction. The model was developed from a relatively limited historical dataset, and the evaluation was performed on only eight testing observations in the reported model configuration. Consequently, the obtained performance metrics and championship probabilities should be interpreted within the scope of the available data. A larger historical dataset covering more tournament editions, together with additional indicators such as recent match results, possession, shots on target, tournament-stage performance, and opponent strength, could provide a more comprehensive representation of team competitiveness. Such additional information may improve the robustness of the model and provide a broader basis for estimating championship probability in future research.

CONCLUSION

This study developed a Logistic Regression model to predict the championship probability of teams participating in the FIFA ASEAN Cup 2026 Challenge Division using five performance indicators: FIFA Ranking, Win Rate, Average Goals Scored, Average Goals Conceded, and Goal Difference. The historical dataset from 2020–2025 was used for model development and evaluation, while the 2026 data were used as the prediction dataset. The evaluation results showed that the Logistic Regression model achieved an Accuracy of 62.50%, Precision of 50.00%, Recall of 66.67%, and F1-score of 57.14%. These results indicate that the model was able to classify Champion and Non-Champion observations, although its predictive performance remains moderate. Based on the 2026 prediction results, Hong Kong obtained the highest estimated championship probability of 77.53%, followed by Myanmar with 68.59% and Cambodia with 53.06%. Laos obtained 28.01%, while Timor-Leste and Brunei Darussalam recorded 0.72% and 0.71%, respectively. The results demonstrate that Logistic Regression can provide a quantitative ranking of championship potential based on the selected performance indicators. Overall, the study demonstrates the potential of Logistic Regression as an interpretable machine learning approach for football championship prediction. The resulting probabilities can provide an analytical reference for comparing team performance and estimating championship potential in the FIFA ASEAN Cup 2026 Challenge Division. Future research should consider a larger historical dataset, additional performance indicators, and tournament-specific match variables to improve the robustness and predictive performance of the model.

REFERENCES

Apryani, I., Fauzi, A., Kusumaningrum, D. S., & Handayani, H. H. (2025). Analisis Sentimen Performa Timnas Sepak Bola Indonesia pada Kolom Komentar Aplikasi TikTok Menggunakan Algoritma Machine Learning. *Jurnal Accounting Information System (AIMS)*, *8*(1), 76–89. https://doi.org/10.32627/aims.v8i1.1160

Atmaja, V. (2026). *Mengulas Pesaing Timnas Indonesia Jelang Piala AFF 2026: Thailand sampai Vietnam Makin Kuat, Saatnya Skuad Garuda Beri Bukti*. Bola.Com. https://www.bola.com/indonesia/read/6655058/mengulas-pesaing-timnas-indonesia-jelang-piala-aff-2026-thailand-sampai-vietnam-makin-kuat-saatnya-skuad-garuda-beri-bukti

Aziz, A., & Sulistiani, H. (2026). Klasifikasi Risiko HIV Berbasis Data Agregat Populasi Menggunakan Random Forest, SVM, dan Logistic Regression. *Jurnal Algoritma*, *23*(1), 1732–1743. https://doi.org/10.33364/algoritma/v.23-1.3492

Dewi, K. R., & Latjompoh, M. (2021). Penerapan Metode Permainan Sepak Bola Verbal Terhadap Application Of The Verbal Football Game Method To Student' s Cognitive Learning Outcomes In Classification. *Jambura Edu Biosfer Journal*, *3*(1), 37–48.

Fadilah, R. (2024). *Sejarah Piala AFF: Ajang sepak bola paling bergengsi di Asia Tenggara*. ANTARA News. https://www.antaranews.com/berita/4356891/sejarah-piala-aff-ajang-sepak-bola-paling-bergengsi-di-asia-tenggara

Gifford, M., & Bayrak, T. (2023). A predictive analytics model for forecasting outcomes in the National Football League games using decision tree and logistic regression. *Decision Analytics Journal*, *8*(August), 1–10. https://doi.org/10.1016/j.dajour.2023.100296

Pratama, F. R., & Dahlan. (2025). ANALISIS SENTIMEN OPINI NETIZEN DI INSTAGRAM TERHADAP BERITA SEPAKBOLA LIGA INDONESIA MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION. *JITET: Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(3), 1338–1345.

Randa, T. M., & Lubis, L. A. (2025). PERBANDINGAN KINERJA MODEL REGRESI LOGISTIK, SUPPORT VECTOR MACHINE, DAN NAÏVE BAYES DALAM MEMPREDIKSI KEPUASAN MAHASISWA TERHADAP PROSES PEMBELAJARAN. *STATMAT (Jurnal Statistika Dan Matematika)*, *7*(3), 426–438.

Rusdianto Hidayat, R., Yersin Putri Larung, E., Sinaga, E., Sari Kardi, I., Fariz Prima Putra, M., Bari Samal, A., &

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

PERFECT
JOURNAL OF SMART ALGORITHMS

E-ISSN: E-ISSN: 3064-0377
Volume 4, **Number** 1, January 2027
https://doi.org/10.62671/perfect.v4i1.336

Babingga, H. (2024). Analitik prediktif sepakbola: model machine learning bri liga 1 indonesia. *MULTILATERAL: Jurnal Pendidikan Jasmani Dan Olahraga*, *23*(4), 386–399.

Said, H. (2025). Prediksi Posisi Optimal Pemain Sepakbola berdasarkan Aspek Kepribadian Tim PON Sepak Bola Putra Jawa Timur. *Jurnal Ilmiah Multidisiplin*, *4*(1), 670–678.

Sulistya, R. A. F., & Pakereng, M. A. I. (2026). ANALISIS SENTIMEN MAHASISWA TERHADAP SIASAT UKSW BERDASARKAN KUESIONER MENGGUNAKAN METODE LOGISTIC REGRESSION. *JURNAL PENDIDIKAN TEKNOLOGI INFORMASI (JUKANTI)*, *9*(1), 20–31.

Ulumando, M. I. (2025). Analisis Prediksi Penerimaan Mahasiswa Baru Menggunakan Regresi Linear : Perbandingan Perhitungan Manual Dan Pemodelan Pada ITAF Kupang. *Journal of Computer and Information Systems Ampera*, *6*(3), 1–20.

Ulumando, M. I. (2026). KLASIFIKASI KANDIDAT MOST VALUABLE PLAYER (MVP) PADA TURNAMEN SEPAK BOLA ASEAN CHAMPIONSHIP 2026 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) BERDASARKAN STATISTIK PERFORMA PEMAIN. *SULIWA: Jurnal Multidisiplin Teknik, Sains, Pendidikan Dan Teknologi*, *3*(3), 268–285.

Ulumando, M. I. (2026). PENGEMBANGAN MODEL PREDIKSI PELUANG JUARA TURNAMEN SEPAK BOLA ASEAN CHAMPIONSHIP 2026 (AFF CUP) MENGGUNAKAN EXTREME GRADIENT BOOSTING (XGBOOST) BERDASARKAN STATISTIK PERFORMA TIM NASIONAL. *SULIWA: Jurnal Multidisiplin Teknik, Sains, Pendidikan Dan Teknologi*, *3*(2), 218–235.

Ulumando, M. I. (2026). PREDIKSI PELUANG JUARA PIALA DUNIA FIFA 2026 MENGGUNAKAN ALGORITMA RANDOM FOREST DAN SIMULASI MONTE CARLO BERDASARKAN DATA STATISTIK PERFORMA TIM NASIONAL. *SULIWA: Jurnal Multidisiplin Teknik, Sains, Pendidikan Dan Teknologi*, *3*(2), 147–163.

Wisswani, N. W., Tulili, T. R., Andrijasa, M. F., Satria, M. R., Wahyudi, E., & Jumaini, S. (2020). Klasifikasi Tipe Gelandang Sepak Bola Berdasarkan Data Kemampuan Menggunakan Metode Naive Bayes. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, *12*(2), 46. https://doi.org/10.46964/justti.v12i2.370

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.