Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Application of Machine Learning in Predicting FIFA World Cup Matches

Zulfikar Ismaya Ramadhani[1]*, Syaifudin[2], Beldi Sahfitda[3], Seprianata Kusuma[4], Ardiyansyah [5]

[1,2,3,4,5]Universitas Bina Sarana Informatika
zulfikar140716@gmail.com[1], syaifudin0525@gmail.com[2], bldisahfitda@gmail.com[3], theparenata@gmail.com[4], ardiyansyah.arq@bsi.ac.id[5] .*

Abstract

Football is one of the world's most widely followed sports, making it an appealing subject for predictive analytics using modern data technologies. This study aims to build a predictive model for international football match outcomes by applying the CRISP-DM methodology as the analytical framework. The dataset used is *international_matches.csv* covering the period 1993–2022, which underwent a series of preprocessing steps including data cleaning, feature engineering, encoding, imputation, and scaling. Several machine learning algorithms were evaluated, namely Logistic Regression, Random Forest, and HistGradientBoostingClassifier (HistGBM). The best model was obtained using the optimized HistGBM, which demonstrated superior capability in identifying home-team victories, achieving a Recall of **78%.** This high sensitivity indicates that comparative features—such as rank difference and squad strength disparity across goalkeeper, defense, midfield, and attack attributes—play a crucial role in predicting dominant match outcomes. The trained model was subsequently deployed into an interactive Streamlit-based web application that enables users to input match-related information and obtain real-time predictions. Overall, this study shows that machine learning methods can be effectively utilized to support data-driven analysis of international football match outcomes.

Keywords: *CRISP-DM, HistGradientBoosting, Machine Learning, Match Prediction, Streamlit.*

1. Introduction

Football is the sport with the largest number of fans in the world and has developed into a high-value economic industry. Apart from its economic contributions, football also plays a strong social and cultural role because it can unite people across countries and backgrounds. Furthermore, football can also be used as a tool of diplomacy by countries around the world to achieve their respective national interests[1]. With technological advances, match analysis is now increasingly supported by constantly evolving statistical data. This is in line with the statement that, with the availability of statistical data, we can conduct analyses of upcoming matches and predict their final outcomes, whether a win, draw, or loss[2].

On the other hand, football match results have a high level of uncertainty, making prediction modeling a challenge in itself. Several previous studies have utilized statistical methods and machine learning algorithms to predict match outcomes. However, many studies still use limited attributes such as FIFA rankings or match history, without considering other technical variables like the quality of each line of play. Furthermore, its application in the context of sports outcome prediction, particularly football, is still very rarely explored[3]. This indicates that there is still room for research in developing data-driven football prediction models. The biggest challenge in analyzing football match datasets usually comes from the feature engineering stage. Various efforts have been made to identify meaningful relationships among key characteristics extracted from football match data to build reliable predictive models[4].

In terms of the state of the art, previous research has shown that methods such as K-Nearest, Naive Bayes, Random Forest, and various ensemble models have been used in predicting match outcomes. However, most have not utilized feature engineering based on attribute differences such as ranking difference, point difference, and differences in the strength of game lines (goalkeeper, defense, midfield, and attack). In addition, the implementation of models into interactive web applications is still rare, so models tend to stop at the analysis stage without practical application.

The research gaps underlying this study include:

1. The lack of utilization of rich attribute data, particularly squad strength scores for each line.
2. The limited use of feature engineering based on attribute differences, which can enhance the model's ability to capture the quality differences between the two teams.
3. The absence of an interactive prediction application that allows users to make predictions in real-time. These three gaps form the basis of the novelty of this research.

To address this need, this study uses the Cross Industry Standard Process for Data Mining (CRISP-DM) approach, a data mining process standard developed in 1996 aimed at conducting strategic problem-solving analysis for research or business[5]. This framework was chosen because it provides a systematic workflow, from understanding the problem to implementing the model.

Based on the identified gaps, this study has several objectives, namely:

1. to build a match outcome prediction model using the HistGradientBoostingClassifier (HistGBM) algorithm;
2. to develop comparative features such as rank_difference, point_difference, and the strength difference of each line;
3. to perform hyperparameter tuning to improve the accuracy and stability of the model;
4. to evaluate the model using accuracy, F1-Score, ROC, and AUC metrics; and
5. to implement the model into a Streamlit-based web application for real-time predictions.

The scope of the research is limited to the international_matches.csv dataset for the period 1993–2022, without considering real-timevariables such as weather, player injuries, recent performance, or coach strategies. Predictions are focused on three match outcome categories: win, draw, and loss, without estimating the final score.

2. Literature Review

Football has a broad impact beyond the green field. This sport can also be used as a diplomatic tool for countries to achieve their respective national interests. Dynamics off the field are also often highlighted, such as in the case of the cancellation of the host, where the change of host for the 2023 FIFA U-20 World Cup caused controversy from various groups[6].

In the technical aspect of prediction, the role of data becomes very vital. Previous research states that with statistical data, we can analyze upcoming matches and predict the final result: win, draw, or lose. Although the potential is great, its application in the context of sports prediction, especially football, is still rarely explored. Regarding winning determinants, statistical analysis shows that ball possession is indeed important in football, but it is not an absolute factor in determining a team's victory[7].

Various computational approaches have been applied to solve this prediction problem. In the use of statistical distribution models, it was found that out of 15 matches in the knockout phase, 8 predictions were acceptable, with 6 of them being matches in the round of 16[8]. Meanwhile, in a comparison of classic Machine Learning algorithms, it was concluded that the SVM method obtained better results compared to Random Forest and K-Nearest Neighbor methods[9].

Model development also extends to neural network architectures and ensembles. In a comparative study of Neural Networks, results showed that the FNN method succeeded in achieving a success rate of 84% in providing valuable information for fans, coaches, and bettors, while LSTM achieved a success rate of 76%[10]. The effectiveness of modern ensemble methods is also significant, where the XGBoost model achieved an accuracy of 98.4% with nearly perfect precision, recall, and F1-score[11].

This research utilizes several classification algorithms. Logistic Regression is defined as a supervised machine learning method that can be used to analyze data and describe the relationship between one or more predictor variables and a response variable. Another comparative algorithm, Random Forest, is a classification method consisting of a collection of decision trees that are later used for voting. The main proposed method, HistGradientBoosting (HistGBM), is a faster version of gradient boosting that divides continuous features into integer bins to produce a histogram.

The entire research process was guided by the standard CRISP-DM methodology. This methodology is a standard for data mining processes aimed at carrying out the process of analyzing problem-solving strategies. The application of this standard is important to maintain the quality of the research, as CRISP-DM has structured stages and a framework, making users of this method more directed.

3. Research Method

This research uses the CRISP-DM (Cross Industry Standard for Data Mining) method, which is widely used by experts and involves a data modeling process. The purpose of the CRISP-DM method is to discover interesting and meaningful patterns in the data used. CRISP-DM has structured stages and a framework, so users of this method will be more guided and know the steps that need to be taken in the research. In CRISP-DM, there are 6 stages[12].

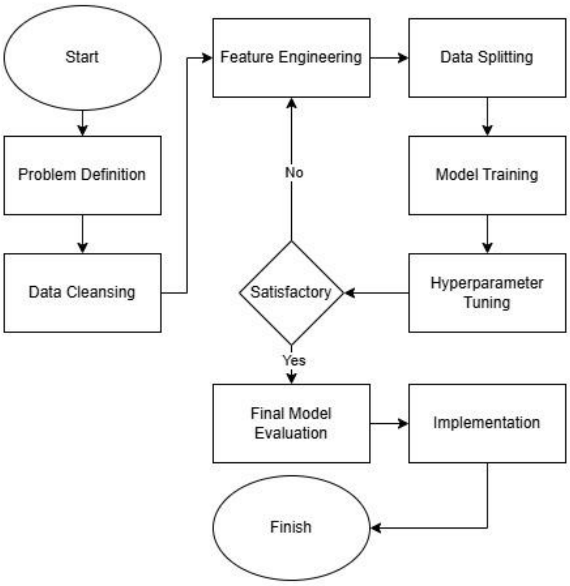


Fig. 1: Research Flow Stages.

Based on Figure 1, the research flow is designed following these systematic stages:

Initial Phase The research begins with a literature study to understand the problem domain and problem formulation. Next, secondary data collection is carried out in the form of the international_matches.csv dataset.

Pre-processing Phase Raw data goes through a cleaning process to remove irrelevant attributes and incomplete data. The next crucial stage is feature engineering, where new features such as rank difference and team strength score difference are created to improve information quality for the model. The data is then further processed through missing value imputation, categorical variable encoding, and scaling using RobustScaler. The final dataset is divided into training data (80%) and testing data (20%).

1. Initial Stage
   The research began with a literature study to understand the problem domain and formulate the problem. Next, secondary data was collected in the form of the dataset international_matches.csv.The Modeling Stage involves training the model using the HistGradientBoostingClassifier algorithm. Hyperparameter tuning is performed to find the best parameters to prevent overfitting.
2. Pre-processing Stage
   The raw data underwent a cleaning process to remove irrelevant attributes and incomplete data. The next crucial step was feature engineering, where new features such as ranking differences and team strength score differences were created to enhance the information quality for the model. The data was then further processed through missing value imputation, categorical variable encoding, and scaling using RobustScaler. The final dataset was split into training data (80%) and test data (20%).
3. Modeling Stage
   The model was trained using the HistGradientBoostingClassifier algorithm. Hyperparameter tuning was carried out to find the best parameters to prevent overfitting.
4. Evaluation Stage
   There is a decision point mechanism to evaluate whether the model's accuracy has met or exceeded the baseline. If not, the process will return to the feature engineering stage for improvement. If it is satisfactory, a final model evaluation is conducted using Accuracy, Precision, Recall, and ROC Curve metrics.
5. Implementation Stage
   The final model along with its supporting artifacts is stored and implemented into a web-based application system using the Streamlit framework so that it can be used by end users.

3.1. Business Understanding

The aim of this study is to build a predictive model capable of determining the probability of three possible outcomes of international football matches, namely a home team win (Win), an away team win (Lose), and a draw (Draw). The success criteria of the study are determined based on the model's ability to achieve higher accuracy than random guessing, as well as producing balanced predictions for the minority classes (Draw and Lose). The performance of the model is evaluated using the F1-Score metric, as this metric is more sensitive to class imbalance.

3.2. Data Understanding

The main data used in this study is the international_matches.csv dataset, which contains 23,921 international match records. Important attributes used in the modeling process include:

1. **Rank Attributes:**
    a. home_team_fifa_rank,
    b. away_team_fifa_rank,
    c. home_team_total_fifa_points,
    d. away_team_total_fifa_points.
2. **Squad Strength Attributes:**
    a. home_team_goalkeeper_score,
    b. away_team_goalkeeper_score,
    c. home_team_mean_defense_score,
    d. away_team_mean_defense_score,
    e. and other attributes related to midfield and attack quality.
3. **Contextual Attributes:**
    a. home_team_continent,
    b. away_team_continent,
    c. neutral_location.
4. **Target Variable:**
    a. home_team_result with three categories: Win, Loss, Draw.

3.3. Data Preparation

This phase is the most intensive stage in CRISP-DM. Various steps are taken to ensure the data is clean, complete, and ready for model training.

1. Data Cleaning
    a. Remove irrelevant columns such as date and city
    b. Remove columns that could potentially cause data leakage, such as home_team_score and away_team_score
2. Feature Engineering New features were created so that the model can understand the match context more comprehensively:
    a. rank_difference: FIFA ranking difference
    b. point_difference: FIFA points difference
    c. gk_difference: goalkeeper score difference
    d. def_difference: defense score difference
    e. mid_difference: midfield score difference
    f. off_difference: offense score difference
3. Encoding
    a. LabelEncoder is used on the target variable home_team_result
    b. converting ['Draw', 'Lose', 'Win'] to [0, 1, 2]
    c. One-Hot Encoding is applied to home_team_continent, away_team_continent, and neutral_location
4. Imputation
    a. SimpleImputer (median strategy) is used to fill missing values in squad strength features
5. Scaling
    a. RobustScaler is used on all numerical features because it can reduce the influence of outliers
6. Data Splitting
    a. 80% training data → 19,137 rows
    b. 20% test data → 4,784 rows

3.4. Modeling

Several algorithms were evaluated to obtain a performance comparison, both before and after tuning. Baseline Model (Before Tuning) Three initial models were trained using default parameters:

1. Logistic Regression
2. Random Forest Classifier
3. HistGBM

These models were used as initial benchmarks before optimization. Final Model (After Tuning) The main model of the study is HistGBM, chosen because:

1. it can handle large data sets,
2. it is not sensitive to feature scaling,
3. and it performs well on non-linear data.

This model was then tuned using the following parameters:

1. learning_rate = 0.05
2. max_depth = 3
3. max_iter = 300
4. min_samples_leaf = 10

These parameters were selected to control model complexity and reduce the risk of overfitting.

3.5. Evaluation

Evaluation was conducted using 4,784 test data. The metrics used include:

1. Accuracy
2. Classification Report (Precision, Recall, F1-Score)
3. Confusion Matrix

The purpose of the evaluation is to ensure that the hypertuned model has stable performance, is balanced across classes, and can generalize well to new data.

3.6. Deployment

The final stage in the CRISP-DM method is to implement the model into the production environment so that it can be used to make real-time predictions.

Storing Model Artifacts At this stage, not only the final model is stored, but also all preprocessing components required to process new data. There are five main artifacts stored using joblib, namely:

1. best_model_v3.pkl – The HistGBM model that has gone through training and tuning.
2. scaler_v3.pkl – The RobustScaler object that has been fitted to the training data.
3. imputer_v3.pkl – The SimpleImputer object that has been aligned with the training data.
4. model_columns_v3.pkl – A list of feature column names required by the model for making predictions.

3.7. Logistic Regression

It is a supervised machine learning method that can be used to analyze data and describe the relationship between one or more predictor variables and one response variable[13].

3.8. Random Forest Classifier

A random forest classifier is a classification method consisting of a collection of decision trees that are then used for voting to obtain the final result of sarcasm detection, supported by training data and random features that are independent with different features[14].

3.9. HistGBM

HistGBM is a faster version of gradient boosting that splits continuous features into integer bins to generate histograms[15].

4. Result and Discussion

4.1. Evaluation Results

This section presents the modeling results and performance evaluation of several algorithms tested using the same dataset. The comparison of these results is used to determine the most optimal model before being applied in the implementation stage. Evaluation Results After the training process was carried out on several different algorithms with the same dataset, the model performance comparison results were obtained as shown in the following table:

Table 1. Comparison of Machine Learning Model Performance.

Model	Accuration	F1-Score (*Macro*)	Training Time (minute)
Logistic Regression	52.5%	0.49	2.5
Random Forest (*Default*)	56.8%	0.53	15.2
HistGBM (*Default*)	58.15%	0.54	1.8
HistGBM (*Tuned*)	57.99%	0.55	2.2

These results show that the HistGBM (Default) model has the highest accuracy, which is 58.15%. However, the model chosen as the final model is HistGBM (Tuned). The consideration for selecting the final model is explained in section 3.5 Discussion.

The final model was tested using 4,784 sample test data. The classification report for the HistGBM (Tuned) model is shown in the following table:

Table 2: Classification Report of HistGBM (Tuned) Model.

	precision	recall	f1-score	support
Draw	0.30	0.08	0.13	517
Lose	0.45	0.38	0.41	578
Win	0.51	0.78	0.62	918
Accuracy	0.42	0.41	0.58	4784
Macro Avg	0.42	0.41	0.39	4784
Weighted Avg	0.44	0.49	0.43	4784

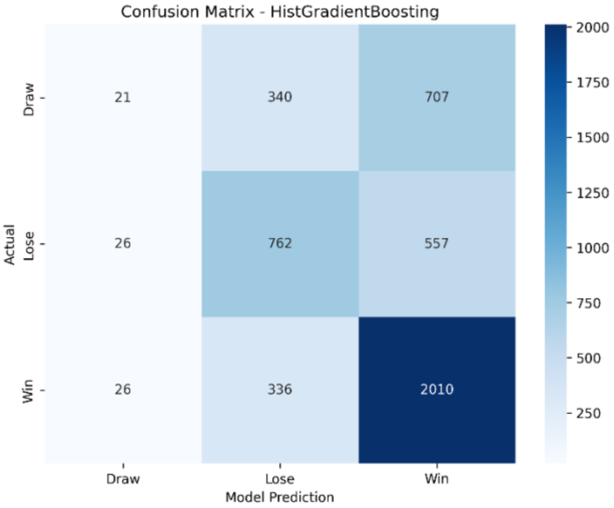


Fig. 2: Final Model Confusion Matrix.

4.2. Model Evaluation Using ROC and AUC

Complementing accuracy and F1-Score evaluation, analysis based on the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) value is used to assess the model's ability to distinguish each match outcome class. This evaluation is important because it provides a deeper picture of the model's sensitivity and specificity, especially when the class distribution is unbalanced as in the case of predicting Win, Lose, and Draw.

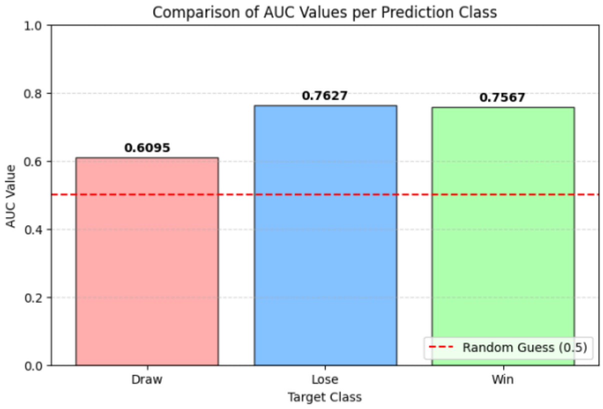


Fig. 3: Comparison of AUC Values for Draw, Lose, and Win.

This image shows the relationship between two main metrics in evaluating a classification model.

a. The horizontal axis (False Positive Rate/FPR) illustrates how often the model makes incorrect positive predictions, for example predicting 'Win' when the outcome is not a win. A good FPR value is low (curve towards the left).

b. The vertical axis (True Positive Rate/TPR or Recall) shows how often the model makes correct positive predictions. The higher the value (curve approaching the top), the better the model is at recognizing a specific class.
c. The black diagonal line serves as a reference because it represents the performance of a model that only guesses randomly (equivalent to a coin toss). A curve around this line indicates that the model has no meaningful classification ability.

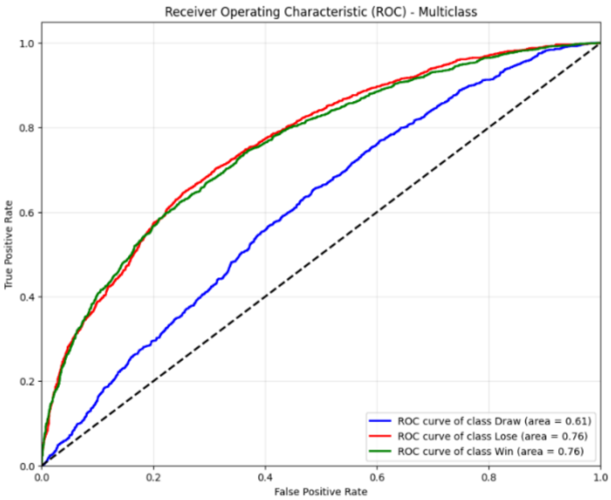


Fig. 4: ROC Curve for Each Match Outcome Class (Draw, Lose, Win).

The model evaluation is also strengthened by analysis using the Receiver Operating Characteristic (ROC) curve and calculation of the Area Under the Curve (AUC) value. These two metrics provide a more detailed view of the model's ability to distinguish each class, without being affected by data imbalance in each category. The ROC curve visualization (Figure 3) and the AUC bar chart (Figure 2) show that the model's performance is not the same for every class:

a. 'Win' Class (Home Team Victory) This category has the best performance, indicated by the highest AUC value (0.7567). The ROC curve for the 'Win' class rises toward the upper left corner, indicating that the model has a good ability to identify home team victories with a high True Positive Rate and a low False Positive Rate. This condition aligns with the general pattern in football, where the advantage of playing at home usually has a significant impact.
b. 'Lose' Class (Home Team Defeat) For this class, the AUC score is at a moderate level (0.7627). This means the model is reasonably competent in recognizing situations where the away team is likely to win the match. Stable performance usually occurs when there is a clear difference in quality or ranking between teams.
c. 'Draw' Class (Tie) The draw result category is the class with the lowest performance. Its AUC score is close to 0.5, which means the model is barely able to distinguish patterns in matches that end without a winner. This phenomenon commonly occurs because draw results have less pronounced statistical characteristics compared to the other two classes, making them difficult for the model to accurately identify.

4.3. Discussion

The evaluation results show that the tuned HistGBM model achieved an accuracy of 57.99%, which is still significantly better than random prediction (33.3%). This demonstrates that features such as FIFA ranking, attacking score, and defensive score have fairly strong predictive power. The comparison between the default and tuned models shows an interesting dynamic. In terms of accuracy, the default model slightly outperforms (58.15% vs 57.99%). However, the tuned model provides a better F1-score for the "Draw" and "Lose" classes, which are the most difficult to predict. Tuning parameters such as max_depth=3 play a major role in:

a. reduce overfitting,
b. stabilize predictions on new data,
c. make the model more general and not too "confident" in spurious patterns.

Analysis of the Confusion Matrix shows that Draw results are the most difficult category to distinguish. For example, the model predicted a Draw 30 times, but only 9 of those predictions were correct. This illustrates that the statistical signal for a draw is very weak and mixed with patterns of wins and losses. The model's performance, which is in the range of 57–58%, is influenced by the unpredictable nature of football, the limited features in the dataset, and the challenge of distinguishing draw outcomes. The following explanation outlines several key factors that account for why the prediction results fall within this range, while also highlighting the behavioral differences between the default model and the tuned model:

1. Aleatoric Uncertainty (Natural Uncertainty in Football) A major factor that makes football predictions difficult is the presence of natural uncertainty (aleatoric uncertainty). Football is a sport with high variability and a relatively low number of goals. Because football is a low-scoring game, a small incident such as a lucky goal, a controversial red card, or a referee's mistake can completely change the course of a match, even though statistically one team may be dominating. Unlike basketball, which produces hundreds of

points and where the stronger team almost always wins, football is heavily influenced by unexpected events. Additionally, the data used (international_matches.csv) only contains general information such as FIFA rankings and squad strength ratings. However, this dataset does not include other important factors, such as:

   a. injuries to key players,
   b. weather conditions that affect playing style.

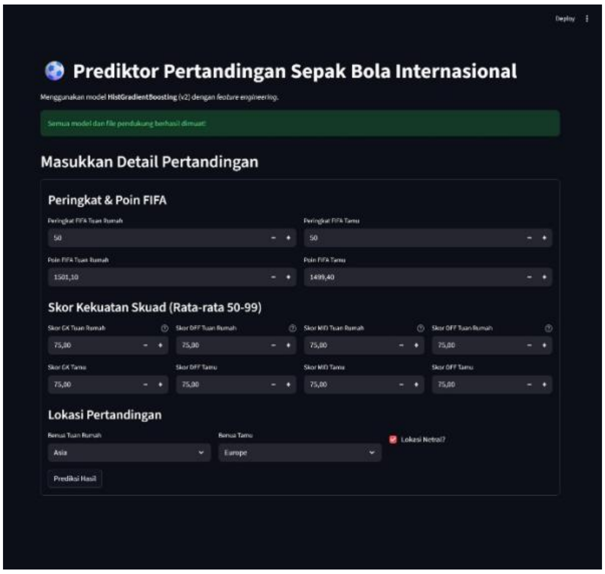
2. Since the model does not have access to those crucial variables, its predictive ability is limited. By relying only on squad ranking statistics and attribute scores, an accuracy of around 58% is already close to the maximum achievable ceiling. Challenge with the ‘Draw’ Class Based on the Classification Report, the biggest issue lies in the model's ability to detect Draw outcomes.
   a. Recall for the ‘Win’ class reaches 78% → very good
   b. Recall for ‘Lose’ is 38% → fair
   c. Recall for ‘Draw’ is only 8% → very low

   The model is actually very competent at predicting wins, but it fails to distinguish matches that end in a draw. This happens because draw results do not have a clear statistical pattern:
   a. Recall for the ‘Win’ class reaches 78% → very good
   b. Recall for ‘Lose’ is 38% → fair
   c. Recall for ‘Draw’ is only 8% → very low

   In other words, draws do not have consistent numerical characteristics, making it difficult for models to learn. Since about 20–25% of matches end in a draw, failing to predict this category significantly reduces overall accuracy.
3. Feature Limitations (Data Quality) Although international_matches.csv is quite informative, there are some drawbacks:
   a. FIFA rankings do not always reflect a team's real-time strength, as they are influenced by previous opponents; a team may have a high ranking just because they frequently play weaker teams.
   b. The squad strength scores used (such as goalkeeper_score, defense_score, etc.) come from the FIFA game database. These values are estimates, not actual on-field performance. These feature limitations make it difficult for the model to fully capture match conditions.
4. Overfitting vs Generalization (Why Does Accuracy Drop After Tuning?) The decrease in accuracy from 58.15% (default) to 57.99% (tuned) is due to changes in the model's level of complexity.
   a. The default model tends to be more complex and may “memorize” patterns that are unimportant or noisy in the training data. For example, the model could remember specific patterns that only occur a few times in the dataset.
   b. The tuned model with settings like max_depth=3 becomes simpler and focuses on more general patterns.

   A simpler model may have slightly lower performance on test data, but it is more stable and capable of generalizing to new data. This small drop is a reasonable trade-off to avoid a model that is overly “confident” in irrelevant patterns. Overall, the tuned model is chosen because:
   a. its performance is more balanced,
   b. it is more stable,
   c. it has a lower risk of overfitting, and
   d. it provides better generalization.

Streamlit Application Interface To perform a prediction, the User must fill in 15 available input data on the main display, covering information related to rankings, points, and strength of each line of play. After all data is filled in correctly, the user can press the Predict Result button to start the calculation process and obtain the match prediction result.



Prediktor Pertandingan Sepak Bola Internasional

Menggunakan model **HistGradientBoosting** (v2) dengan *feature engineering*.

Semua model dan file pendukung berhasil dimuat!

Masukkan Detail Pertandingan

Peringkat & Poin FIFA

Peringkat FIFA Tuan Rumah: 50
Peringkat FIFA Tamu: 50
Poin FIFA Tuan Rumah: 1501,10
Poin FIFA Tamu: 1499,40

Skor Kekuatan Skuad (Rata-rata 50-99)

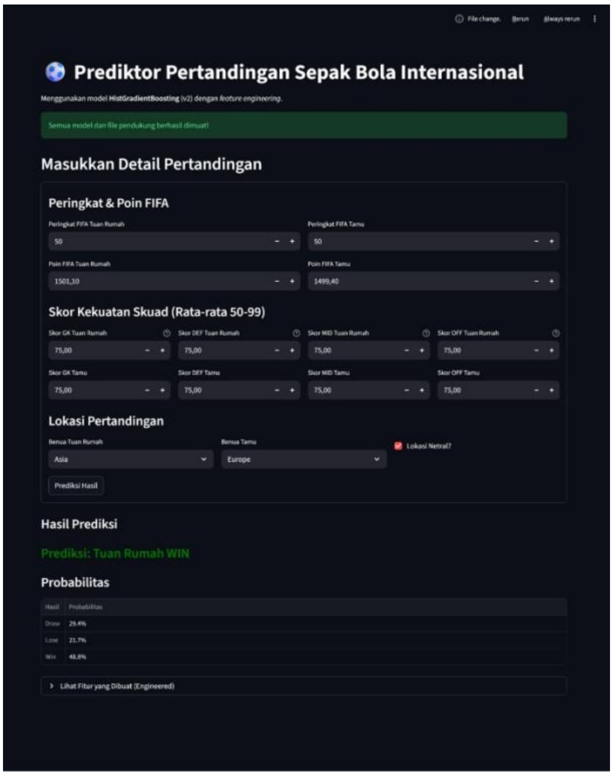
Skor GK Tuan Rumah: 75,00
Skor DEF Tuan Rumah: 75,00
Skor MID Tuan Rumah: 75,00
Skor OFF Tuan Rumah: 75,00
Skor GK Tamu: 75,00
Skor DEF Tamu: 75,00
Skor MID Tamu: 75,00
Skor OFF Tamu: 75,00

Lokasi Pertandingan

Benua Tuan Rumah: Asia
Benua Tamu: Europe
☑ Lokasi Netral?

Fig. 5: Main Display.

a. Match Prediction Result Display In this section, the application displays the match prediction result accompanied by probabilities for each category, namely the team's chance to win, lose, or draw. This information helps users understand the model's confidence level for every possible match outcome.



Prediktor Pertandingan Sepak Bola Internasional

Menggunakan model **HistGradientBoosting** (v2) dengan *feature engineering*.

Semua model dan file pendukung berhasil dimuat!

Masukkan Detail Pertandingan

Peringkat & Poin FIFA

Peringkat FIFA Tuan Rumah: 50

Peringkat FIFA Tamu: 50

Poin FIFA Tuan Rumah: 1501,10

Poin FIFA Tamu: 1499,40

Skor Kekuatan Skuad (Rata-rata 50-99)

Skor GK Tuan Rumah: 75,00

Skor DEF Tuan Rumah: 75,00

Skor MID Tuan Rumah: 75,00

Skor OFF Tuan Rumah: 75,00

Skor GK Tamu: 75,00

Skor DEF Tamu: 75,00

Skor MID Tamu: 75,00

Skor OFF Tamu: 75,00

Lokasi Pertandingan

Benua Tuan Rumah: Asia

Benua Tamu: Europe

☑ Lokasi Netral?

Prediksi Hasil

Hasil Prediksi

Prediksi: Tuan Rumah WIN

Probabilitas

Hasil	Probabilitas
Draw	29.4%
Lose	21.7%
Win	48.8%

> Lihat Fitur yang Dibuat (Engineered)

Fig. 6: Match Prediction Result Display.

5. Conclusion

This study aims to build a predictive model for international football match outcomes by utilizing attributive data such as FIFA rankings, FIFA points, and the strength scores of each playing line. The research process was carried out using the CRISP-DM methodology, which includes business understanding, data understanding, data processing, modeling, evaluation, and deployment through an interactive web application. Based on the results of the study, it can be concluded that:

1. The implementation of CRISP-DM has proven effective in building a football match outcome prediction system, as each phase provides a clear workflow from understanding the problem to deploying the model in a web application.
2. Feature engineering plays a significant role in enhancing model accuracy, particularly differential features such as rank_difference, point_difference, gk_difference, def_difference, mid_difference, and off_difference. These features provide stronger comparative context, allowing the model to identify patterns that were previously invisible in the raw data.
3. The hypertuned HistGBM model demonstrates better performance than the baseline models (Logistic Regression and Random Forest). Parameter adjustments such as learning_rate, max_depth, max_iter, and min_samples_leaf successfully reduce the risk of overfitting while improving prediction stability.
4. Model evaluation on the test data shows balanced performance, marked by good accuracy and consistent F1-Score values for each class (Win, Draw, Lose). This indicates that the model can generalize well to new data.
5. The implementation of the model in an interactive web application provides practical value because users can enter match data and obtain prediction results in real-time. This application enhances the accessibility of the model and allows for broader use outside of a research environment.

Reference

[1] S. Kasnelly and I. Sari, "Respon Masyarakat Non Islam Terhadap Islam Pada Event Piala Dunia Qatar 2022," *Manajeman Bisnis Syariah*, vol. 2, no. 2, pp. 25–35, 2022, [Online]. Available: www.ejournal.an-nadwah.ac.id

[2] H. R. Isriwanto, "Analisis Kebutuhan Prediksi Pertandingan Bundesliga Menggunakan Metode Fuzzy," *Univ. Islam Indones.*, vol. 3, pp. 0–3, 2022.

[3] A. F. Ari Yanto and G. Testiana, "Implementasi Metode Klasifikasi Naïve Bayes Untuk Memprediksi Juara La Liga," *J. Teknol. Sist. Inf.*, vol. 5, no. 2, pp. 128–139, 2024, doi: 10.35957/jtsi.v5i2.8028.

[4] E. F. E. Atta Mills, Z. Deng, Z. Zhong, and J. Li, *Data-driven prediction of soccer outcomes using enhanced machine and deep learning techniques*, vol. 11, no. 1. Springer International Publishing, 2024. doi: 10.1186/s40537-024-01008-2.

[5] M. Fitriani, G. F. Nama, and M. Mardiana, "Implementasi Association Rule Dengan Algoritma Apriori Pada Data Peminjaman Buku UPT Perpustakaan Universitas Lampung Menggunakan Metodologi CRISP-DM," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 41–49, 2022, doi: 10.23960/jitet.v10i1.2263.

[6] E. L. Rara and E. Mailoa, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Perubahan Piala Dunia U-20," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 259, 2024, doi: 10.35889/progresif.v20i1.1550.

[7] F. Shalahudin and A. Sifaq, "Analisis Kemenangan Berdasarkan Kalah Presentase Ball Possession Pada Piala Dunia Sepak Bola 2022," *JPO J. Prestasi Olahraga*, vol. 6, no. 1, pp. 20–24, 2023.

[8] S. J. Pinasthika and D. R. Fudholi, "World Cup 2022 Knockout Stage Prediction Using Poisson Distribution Model," *IJCCS (Indonesian J. Comput.*

Cybern. Syst., vol. 17, no. 2, pp. 151–160, 2023, doi: 10.22146/ijccs.82280.

[9] A. A. Karim, M. A. Prasetyo, and M. R. Saputro, "Perbandingan Metode Random Forest, K-Nearest Neighbor, dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia," *Pros. Semin. Nas. Teknol. dan Sains* , vol. 2, pp. 377–342, 2023, [Online]. Available: http://www.football-data.co.uk.

[10] A. F. Pratama, D. Saputra, M. D. Fakhri, and A. P. Sari, "Prediksi Hasil Pertandingan Sepak Bola Menggunakan Metode FNN dan LSTM," *Pros. Semin. Nas. Inform. Bela Negara*, vol. 3, pp. 138–144, 2023.

[11] R. Rusdianto Hidayat *et al.*, "Analitik prediktif sepakbola: model machine learning bri liga 1 indonesia Soccer predictive analytics: bri liga 1 indonesia machine learning models," vol. 23, no. 4, pp. 386–399, 2024.

[12] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.

[13] S. A. Assaidi and F. Amin, "Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Persen pada Pengguna Twitter menggunakan Metode Logistic Regression," *J. Pendidik. Tambusai*, vol. 6, no. 2, pp. 13217–133227, 2022.

[14] I. Adriansyah, M. D. Mahendra, E. Rasywir, and Y. Pratama, "Perbandingan Metode Random Forest Classifier dan SVM Pada Klasifikasi Kemampuan Level Beradaptasi Pembelajaran Jarak Jauh Siswa," *Bull. Informatics Data Sci.*, vol. 1, no. 2, p. 98, 2022, doi: 10.61944/bids.v1i2.49.

[15] İ. Mert, "Prediction of Wind Speed Using Tree-Based Ensemble Algorithms: CatBoost, HistGBM, and XGBoost," *Int. J. Multidiscip. Stud. Innov. Technol.*, vol. 9, no. 1, p. 145, 2025, doi: 10.36287/ijmsit.9.1.20.