



**PROJECT REPORT**

**COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR  
FOOTBALL MATCH OUTCOME PREDICTION USING LA LIGA MATCH  
STATISTICS**

**RONALDO JULIAN RAJAAR**  
**20.K1.0037**

**Faculty of Computer Science**  
**Soegijapranata Catholic University**  
**2026**

## ABSTRACT

Football match outcome prediction has become an important application of machine learning in sports analytics. This study compares the performance of Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost) for predicting LaLiga match outcomes using the FBref LaLiga Matches Dataset covering the 2019–2025 seasons. The dataset consists of 4,700 match records and was processed through data cleaning, median imputation, feature selection, chronological data splitting, and feature standardization. Nine numerical performance-based features were selected as predictor variables for model development. Model performance was evaluated using Accuracy, Precision, Recall, F1-score, a Classification Report, and a Confusion Matrix. The experimental results show that Logistic Regression achieved the highest overall performance among the evaluated models under the implemented experimental configuration. In addition, XGBoost was employed to analyze feature importance and identify the performance indicators that contributed most to match outcome prediction. These findings demonstrate that model performance is influenced by both dataset characteristics and model configuration, providing useful insights for selecting appropriate machine learning algorithms for football match outcome prediction.

*Keyword: Football match prediction; LaLiga; Logistic Regression; Machine learning; Random Forest; Sports analytics; XGBoost*