

ABSTRAK

Perkembangan industri game digital di Indonesia mendorong pemanfaatan analisis data dalam game simulasi manajemen olahraga seperti Football Manager (FM). Salah satu tantangan utama dalam permainan ini adalah memprediksi Potential Ability (PA), yaitu atribut tersembunyi yang merepresentasikan batas maksimal kemampuan pemain di masa depan. Penelitian sebelumnya umumnya menggunakan pendekatan klasifikasi biner dan belum mengoptimalkan seleksi fitur, sehingga kurang merepresentasikan nilai PA yang bersifat numerik serta berpotensi menurunkan performa model. Oleh karena itu, penelitian ini menerapkan pendekatan regresi menggunakan algoritma Extreme Gradient Boosting (XGBoost) untuk memprediksi nilai PA secara kontinu.

Dataset diperoleh dari Kaggle dengan menggabungkan FM 2020 dan FM 2023 sebanyak 153.603 entri data. Tahapan preprocessing meliputi cleaning, harmonization, merging, serta refinement untuk mengonversi kode PA negatif menjadi nilai aktual, diikuti transformasi fitur kategorikal menggunakan Ordinal Encoder. Data dibagi dengan rasio 80:20 untuk pelatihan dan pengujian. Seleksi fitur dilakukan melalui XGBoost Feature Importance berbasis Gain dengan ambang 1,25 kali median, sementara optimasi hyperparameter menggunakan Optuna dengan metode TPE.

Hasil penelitian menunjukkan bahwa model XGBoost regresi menghasilkan performa terbaik setelah tuning dengan RMSE 9.86, MAE 7,43, dan R^2 sebesar 0,86. Proses seleksi fitur berhasil mereduksi atribut dari 60 menjadi 24 tanpa penurunan akurasi yang signifikan, dengan model akhir mempertahankan MAE 7,67, RMSE 10.21 dan R^2 0,85. Penelitian ini membuktikan bahwa seleksi fitur berbasis Gain mampu meningkatkan efisiensi komputasi sekaligus menghasilkan prediksi PA yang akurat dan objektif dalam konteks simulasi manajemen sepak bola digital.

Kata Kunci: *Football Manager, Potential Ability, XGBoost, Feature Importance, Optuna.*

ABSTRACT

The growth of the digital gaming industry in Indonesia has encouraged the use of data analytics in sports management simulation games such as Football Manager (FM). One of the main challenges in this game is predicting Potential Ability (PA), a hidden attribute that represents a player's maximum future performance. Previous studies generally applied binary classification approaches and did not optimize feature selection, which limited the representation of PA as a numerical value and potentially reduced model performance. Therefore, this study adopts a regression approach using the Extreme Gradient Boosting (XGBoost) algorithm to predict PA continuously.

The dataset was obtained from Kaggle by merging FM 2020 and FM 2023 data, resulting in a total of 153,603 records. The preprocessing stages included data cleaning, harmonization, merging, and refinement to convert negative PA codes into actual values, followed by categorical feature transformation using Ordinal Encoder. The data were split into training and testing sets with an 80:20 ratio. Feature selection was conducted using XGBoost Feature Importance based on the Gain metric with a threshold of 1.25 times the median, while hyperparameter optimization was performed using Optuna with the Tree-structured Parzen Estimator (TPE) method.

The results show that the tuned XGBoost regression model achieved the best performance with an RMSE of 9.86, an MAE of 7.43, and an R^2 of 0.86. Feature selection successfully reduced the number of attributes from 60 to 24 without a significant loss in accuracy, with the final model maintaining an MAE of 7.67, RMSE 10.21 and an R^2 of 0.85. This study demonstrates that Gain-based feature selection improves computational efficiency while maintaining high predictive accuracy, providing an objective predictive solution for player potential in digital football management simulations.

Keywords: *Football Manager, Potential Ability, XGBoost, Feature Importance, Optuna.*