

ABSTRACT

ADDITION OF EXPECTED GOALS (XG) IN FOOTBALL MATCH OUTCOME PREDICTION BASED ON MATCH STATISTICS USING RANDOM FOREST AND XGBOOST

By

Fadgham Albar Ronggoputra

21/479723/PA/20802

The popularity of football around the world has encouraged football to become more modernized, one of which is the use of data-driven approach in decision making in football. One evidence of its development is the presence of a modern statistical metric, namely expected goals (xG), which is believed to be able to assess the course of a match by separating the quality of team performance and luck factors after the full time whistle. This research proposes a method of adding xG features in football match outcome prediction into the match statistics dataset. This research uses the English Premier League match statistics dataset from 2017/2018 to 2023/2024. The addition of xG features provided by Opta and acquired through football reference (FBREF), this method aims to improve the performance of the model. The method was evaluated using confusion matrix, accuracy, precision, recall, and F1-score metrics.

This research implements a machine learning model for football match prediction. Web scraping and pre-processing methods are applied to improve the prediction accuracy value. The results showed that the addition of xG features resulted in an increase in accuracy, reaching an increase of 0.5% to 1.5%. The highest performance was achieved by the Random Forest model with the addition of xG resulting in an accuracy value of 66.54%.

Kata Kunci: Football Match Outcome Prediction, *Expected Goals*, XG, *Machine Learning*, XGBoost, *Random Forest*

INTISARI

PENAMBAHAN FITUR *EXPECTED GOALS* (xG) TERHADAP PREDIKSI HASIL PERTANDINGAN SEPAK BOLA BERDASARKAN STATISTIK PERTANDINGAN MENGGUNAKAN *RANDOM FOREST* DAN *XGBOOST*

Oleh

Fadgham Albar Ronggoputra

21/479723/PA/20802

Popularitas olahraga sepak bola yang tinggi di seluruh dunia mendorong perkembangan sepak bola bertransisi menjadi lebih modern, salah satunya banyaknya penggunaan data dalam pengambilan keputusan di sepak bola. Salah satu bukti perkembangannya adalah kehadiran metrik statistik modern yaitu *expected goals* (xG) yang diyakini dapat menilai jalannya suatu pertandingan dengan memisahkan kualitas performa tim dan faktor keberuntungan usai pertandingan berakhir. Penelitian ini mengusulkan metode penambahan fitur xG dalam prediksi hasil pertandingan sepak bola ke dataset statistik pertandingan sepak bola. Penelitian ini Menggunakan dataset statistik pertandingan *English Premier League* musim 2017/2018 sampai 2023/2024. Penambahan fitur xG yang disediakan oleh Opta dan diakuisisi melalui *football reference* (FBREF), metode ini bertujuan untuk meningkatkan performa model. Metode dievaluasi menggunakan metrik *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*.

Penelitian ini mengimplementasikan model *machine learning* untuk prediksi pertandingan sepak bola. Metode *web scraping* dan pra-pemrosesan diterapkan untuk meningkatkan nilai *accuracy* prediksi. Hasil menunjukkan bahwa penambahan fitur xG menghasilkan peningkatan yang dalam nilai *accuracy*, mencapai kenaikan sebesar 0.5% sampai 1.5%. Performa tertinggi diraih oleh model *Random Forest* dengan penambahan xG menghasilkan nilai *accuracy* sebesar 66.54%.

Kata Kunci: Prediksi Hasil Pertandingan Sepak Bola, *Expected Goals*, xG, *Machine Learning*, XGBoost, *Random Forest*