

ABSTRACT

Football match outcomes can be predicted using historical factors such as goals, shots, shots on target, and other team performance statistics that influence the final result. To improve the accuracy of these predictions, this study applies Gaussian Naïve Bayes algorithm to English Premier League data from the 2018-2024 seasons totaling 2.660 matches, through stages including data cleaning, encoding, normalization, the addition of derived features, train–test splitting, and parameter optimization of var_smoothing using Grid Search and Stratified K-Fold validation. Evaluation using accuracy, classification report, and confusion matrix shows a test accuracy of 66.02% and a cross-validation accuracy of 62.88%, with shots and shots on target being the most influential features, while yellow and red cards have minimal impact. This model provides probabilistic predictions for win, draw, or loss outcomes and offers deeper insights into the factors that influence match results based on historical data.

Keywords: English Premier League, Naïve Bayes, Match Prediction

ABSTRAK

Prediksi hasil pertandingan sepak bola dapat dilakukan menggunakan faktor-faktor historis seperti goals, *shots*, *shots on target*, serta statistik performa tim lainnya yang berpengaruh terhadap hasil akhir. Untuk meningkatkan akurasi prediksi tersebut, penelitian ini menerapkan algoritma Gaussian Naïve Bayes pada data Liga Inggris musim 2018-2014 berjumlah 2.660 pertandingan melalui tahapan pembersihan data, *encoding*, normalisasi, penambahan fitur turunan, pembagian data latih–uji, serta optimasi parameter *var_smoothing* menggunakan *Grid Search* dan *Stratified K-Fold Validation*. Evaluasi menggunakan akurasi, *classification report*, dan *confusion matrix* menunjukkan akurasi uji 66.02% dan akurasi *cross-validation* 62.88% dengan *shots* dan *shots on target* sebagai fitur yang paling berpengaruh, sedangkan kartu kuning dan kartu merah memiliki pengaruh rendah. Model ini memberikan prediksi probabilistik untuk kemungkinan menang, imbang, atau kalah, serta menawarkan pemahaman lebih mendalam mengenai faktor-faktor yang memengaruhi hasil pertandingan berdasarkan data historis,

Kata Kunci: Liga Inggris, Naïve Bayes, Match Prediction