

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang didapat dari penelitian ini adalah:

- a. Untuk mengolah dan menganalisis data pertandingan *English Premier League* guna memprediksi hasil klasemen secara akurat, penelitian ini dilakukan melalui beberapa tahapan yang sistematis. Pertama, data pertandingan dikumpulkan dari sumber terpercaya yaitu situs *FBref.com*, yang mencakup berbagai statistik seperti jumlah kemenangan, hasil imbang, kekalahan, jumlah gol, tembakan, penguasaan bola, dan atribut lainnya yang relevan dengan performa tim. Selanjutnya, data yang diperoleh melalui proses *web scraping* ini dilakukan tahap pra-pemrosesan, yang meliputi pembersihan data (*data cleaning*), penanganan nilai yang hilang (*missing values*), *encoding* variabel kategorikal, serta seleksi fitur untuk menentukan atribut yang paling berpengaruh terhadap hasil pertandingan. Setelah data siap, dilakukan pemisahan antara data pelatihan dan data pengujian dengan rasio 80:20, kemudian dibangun model prediksi menggunakan dua algoritma *machine learning*, yaitu *Naïve Bayes* dan *Random Forest*.
- b. Algoritma *Naïve Bayes* menghasilkan akurasi 69,83% dalam memprediksi klasemen akhir *Premier League*. Namun, nilai *precision*, *recall*, dan *f1-score* masih rendah, terutama dalam memprediksi hasil imbang atau kalah. Hal ini menunjukkan bahwa model belum mampu membedakan beberapa kelas dengan baik. Salah satu penyebabnya adalah asumsi *Naïve Bayes* yang menganggap semua fitur saling bebas, padahal dalam sepak bola banyak fitur yang saling terkait. Sementara itu, algoritma *Random Forest* mendapatkan akurasi 99,57%, hasil sangat baik. Nilai *precision*, *recall*, dan *f1-score* juga tinggi dan seimbang, menandakan model mampu mengenali pola data dengan baik.

- c. Visualisasi klasemen prediksi menunjukkan algoritma *Random Forest* lebih akurat dalam mengurutkan tim berdasarkan performa, menghasilkan hasil prediksi yang lebih mendekati kenyataan dibandingkan dengan *Naive Bayes*.

5.2. Saran

Saran yang dapat dijadikan acuan untuk pengembangan penelitian selanjutnya:

1. Penambahan Fitur

Penelitian selanjutnya disarankan untuk menambahkan fitur-fitur yang lebih kompleks seperti performa pemain individu, kondisi cuaca, histori pertemuan antar tim, atau faktor non-teknis seperti cedera pemain dan jadwal pertandingan.

2. Penanganan Ketidakseimbangan Data

Distribusi kelas hasil pertandingan (menang, seri, kalah) pada data dapat mempengaruhi performa model. Oleh karena itu, disarankan untuk melakukan *balancing dataset* melalui teknik seperti SMOTE (*Synthetic Minority Over-sampling Technique*) atau *undersampling* untuk meningkatkan akurasi model.

3. Eksplorasi Model Lain

Di samping penggunaan *Naive Bayes* dan *Random Forest*, penelitian mendatang dapat diarahkan untuk mengevaluasi algoritma lain seperti *Support Vector Machine (SVM)*, *Gradient Boosting*, maupun metode *Deep Learning* seperti LSTM yang memiliki kemampuan dalam mengolah data berurutan.

Dengan memperhatikan beberapa saran tersebut, diharapkan hasil prediksi klasemen dapat lebih akurat dan aplikatif untuk keperluan analisis sepak bola atau sistem pendukung keputusan dalam konteks olahraga.