

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sepak bola menjadi cabang olahraga yang paling banyak diminati oleh masyarakat di berbagai belahan dunia, dan *English Premier League* (EPL) terkenal sebagai salah satu kompetisi sepak bola paling kompetitif secara global. Setiap tahunnya, sebanyak 20 tim papan atas dari Inggris saling bersaing untuk mendapatkan gelar juara EPL. Persaingan yang ketat serta pertandingan yang penuh tensi tinggi menjadikan prediksi pemenang liga sangat menarik namun juga menantang. Hingga saat ini, klub yang selalu berpartisipasi dalam setiap musim EPL, yaitu *Arsenal*, *Chelsea*, *Everton*, *Liverpool*, *Manchester United*, dan *Tottenham Hotspur*.

Pada sebuah data dalam pertandingan, statistik adalah komponen yang utama (Fauzan & Bawono, 2021). Statistik dalam dunia olahraga, terutama sepak bola, mencakup berbagai elemen yang dapat dimanfaatkan untuk menilai kinerja sebuah tim. Untuk meningkatkan ketepatan dalam memprediksi hasil pertandingan, peran analisis statistik menjadi sangat krusial. Tingkat akurasi prediksi sangat dipengaruhi oleh relevansi data statistik pertandingan serta keandalan analisis yang dilakukan sebelum pertandingan berlangsung (Wheatcroft, 2021). Hasil pertandingan sepak bola bersifat tidak pasti, namun melalui analisis data dari berbagai pertandingan sebelumnya, dapat ditemukan pola-pola tertentu yang dapat dimanfaatkan untuk memperkirakan hasil pertandingan yang akan datang (Kristen Satya Wacana Salatiga, 2019).

Prediksi hasil pertandingan kerap menjadi topik yang menarik bagi para penggemar sepak bola. Seiring dengan kemajuan teknologi, berbagai data penting terkait pertandingan kini dapat diakses dengan lebih mudah. Informasi tersebut dapat dianalisis dan dimanfaatkan untuk memperkirakan hasil pertandingan di masa depan. Salah satu cabang dalam ilmu komputer yang sering kali digunakan untuk membuat prediksi berbasis data adalah *Machine Learning* (Winata et al., 2020).

Algoritma *Naive Bayes* dan *Random Forest* merupakan dua algoritma yang cukup populer dalam bidang prediksi. Algoritma *Naive Bayes* mengandalkan

konsep probabilitas dalam membuat prediksi, sementara *Random Forest* bekerja dengan menggabungkan sejumlah pohon keputusan guna menghasilkan prediksi akurat. Kedua algoritma ini telah diterapkan dalam berbagai penelitian terkait prediksi hasil pertandingan sepak bola, dan hasilnya menunjukkan bahwa keduanya mampu memberikan tingkat akurasi prediksi yang tinggi.

Penelitian tentang prediksi pada pertandingan olahraga menggunakan algoritma *Naïve Bayes* sudah banyak dilakukan. Salah satu penelitian yang berjudul “Prediksi Hasil Pertandingan Liga *Serie A* Menggunakan Metode *Naïve Bayes*” oleh Ridwan Adi Pratama, Winda Apriandari, Didik Indrayana. Dalam penelitian ini, menggunakan sejumlah atribut atau variabel dalam dataset, antara lain: HT, AT, HTHG, HTAG, HS, AS, HST, AST, HF, AF, HC, AC, HY, AY, HR, AR, dan FTR. Penelitian memprediksi hasil pertandingan ini berdasarkan klasemen akhir Liga *Serie A*. Dengan menerapkan algoritma *Naïve Bayes*, diperoleh tingkat akurasi sebesar 75,79%.(Pratama et al., 2023). Penelitian berjudul “Perbandingan Metode *Random Forest*, *K-Nearest Neighbor*, dan *SVM* Dalam Prediksi Akurasi Pertandingan Liga Italia” yang dilakukan oleh Ahmad Assril Karim, Muhammad Ary Prasetyo, dan Muhammad Rohid Saputro membagi dataset sebanyak 380 data menjadi dua data, yaitu 80% sebagai data latih dan 20% sebagai data uji. Dari hasil prediksi algoritma *Random Forest*, diperoleh akurasi sebesar 62%. Persentase akurasi tersebut diperoleh berdasarkan perhitungan dari data pelatihan yang digunakan (Karim et al., 2023).

Keakuratan prediksi Tingkat akurasi dalam prediksi sangat dipengaruhi oleh kualitas data yang digunakan. Jika data bersifat tidak lengkap, tidak terstruktur dengan baik, atau kurang relevan, hal tersebut dapat menurunkan kinerja model prediktif yang dibangun. Selain itu, pemahaman terhadap hasil prediksi juga perlu diperhatikan, karena dalam penerapan *Machine Learning*, tidak hanya akurasi yang penting, tetapi juga bagaimana cara model membuat keputusan. Berdasarkan hal tersebut, penelitian ini akan memusatkan perhatian pada analisis serta perbandingan antara dua algoritma untuk menilai akurasi, efektivitas, dan tingkat interpretabilitas dalam memprediksi peringkat akhir di English Premier League. Diharapkan hasil dari studi ini dapat bermanfaat dalam mengembangkan model prediksi yang lebih andal dan efisien di bidang sepak bola.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang, beberapa masalah telah diidentifikasi. Adapun rumusan masalah dalam penelitian ini dapat dirinci ke dalam beberapa poin utama berikut:

1. Bagaimana cara mengolah dan menganalisis data pertandingan *English Premier League* untuk memprediksi hasil klasemen secara akurat?
2. Sejauh mana tingkat akurasi algoritma *Naïve Bayes* dan *Random Forest* dalam prediksi terhadap klasemen akhir *English Premier League*?
3. Manakah algoritma yang menunjukkan kinerja lebih unggul dalam memprediksi hasil klasemen antara *Naive Bayes* dan *Random Forest*?

## 1.3. Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut :

1. Identifikasi faktor statistik yang paling berpengaruh dalam prediksi hasil klasemen *English Premier League*.
2. Menganalisis dan membandingkan akurasi model yang dihasilkan dari kedua algoritma dalam memprediksi hasil klasemen EPL.
3. Melakukan evaluasi kinerja antara algoritma *Naive Bayes* dan *Random Forest* untuk menentukan metode yang memiliki tingkat akurasi tertinggi dalam memprediksi klasemen.

## 1.4. Manfaat Penelitian

Adapun manfaat dari penelitian ini sebagai berikut :

1. Menyediakan wawasan baru dalam analisis data olahraga, khususnya dalam konteks prediksi hasil pertandingan sepak bola.
2. Memberikan informasi yang berguna bagi penggemar sepak bola dan pihak-pihak terkait dalam industri olahraga mengenai tim sepak bola EPL.
3. Memberikan acuan bagi penelitian berikutnya analisis data serta prediksi di bidang olahraga.