

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sepak bola adalah olahraga yang dimainkan oleh dua tim yang masing-masing terdiri dari 11 pemain. Pemenang pertandingan ditentukan dari tim yang mampu mencetak gol lebih banyak dari lawannya. Sepak bola menjadi olahraga paling populer di dunia dengan jumlah penggemar mencapai hampir 4 miliar orang. Di Indonesia sendiri, dari total populasi sekitar 275 juta jiwa, sebanyak 77% penduduk memiliki minat terhadap sepak bola (Saputro, dkk. 2024). Hasil dari sebuah pertandingan sepak bola sulit untuk diperkirakan, namun dapat diperkirakan dari faktor-faktor pertandingan sebelumnya untuk memprediksi hasil suatu pertandingan (Zein dan Gunawan 2022). Memprediksi dapat memberikan gambaran realistis dalam bentuk persentase serta membantu menghitung peluang kemenangan secara lebih objektif.

English Premier League (EPL) disebut juga liga inggris adalah salah satu kompetisi sepak bola paling bergengsi di dunia. Kompetisi ini diikuti oleh klub-klub terbaik di Inggris. EPL juga menjadi liga olahraga dengan jumlah penonton terbesar secara global, dengan siarannya disiarkan ke 634 juta rumah di 212 negara serta berpotensi menjangkau hingga 4,7 miliar pemirsa (Alim dan Murni 2023). Faktor utama dari daya tarik liga inggris dikarenakan adanya pemain-pemain bintang yang berasal dari berbagai negara, rivalitas yang sengit antar klub, serta pelatih berpengalaman yang membawa strategi permainan menjadi lebih menarik.

Dalam sepak bola banyak penggemarnya melakukan peralaman atau prediksi. Prediksi adalah proses untuk mengestimasi nilai di masa mendatang. Diperlukan data historis dalam melakukan prediksi, yang mencakup berbagai faktor dari pertandingan-pertandingan sebelumnya. Terdapat faktor-faktor yang berpengaruh terhadap suatu prediksi, seperti gol, jumlah kemenangan, dan faktor lainnya.(Zein dan Gunawan 2022). Faktor-faktor tersebut akan digunakan sebagai variabel dalam berbagai metode prediksi, termasuk dalam penerapan algoritma *machine learning*. *Machine learning* adalah salah satu keilmuan dari kecerdasan buatan (AI) yang membahas pada algoritma serta teknik yang memungkinkan komputer untuk belajar dan beradaptasi secara otomatis yang dapat membantu dalam berbagai permasalahan (Putra dan Fatimah 2023) . Prediksi bukan jawaban pasti akan kejadian yang terjadi, tetapi usaha untuk mendekati kemungkinan yang akan terjadi (R. A. Pratama, Apriandari, dan Indrayana 2023). Berdasarkan analisis yg dilakukan terdapat beberapa faktor krusial di dalam prediksi sebuah pertandingan yaitu *shot*, *shot on target* ,*corner*, dan *red card*. Faktor tersebut dikatakan krusial karena banyak mempengaruhi suatu pertandingan sepak bola.

Naïve Bayes adalah metode klasifikasi probabilitas yang sederhana, yang menentukan probabilitas suatu kelas dengan menghitung frekuensi serta kombinasi nilai dari dataset yang tersedia, dimana algoritma ini berdasarkan Teorema Bayes dan mengasumsikan bahwa setiap atribut tidak memiliki keterkaitan satu sama lain terhadap nilai yang terdapat dalam variabel kelas (Martantoh dan Yanih 2022). Hasil pengujian menunjukkan bahwa metode *Naïve Bayes* berhasil mencapai tingkat akurasi sebesar 63% dengan menggunakan data 2280 pertandingan

(Prabowo 2020). Pendekatan ini menjadikan *Naïve Bayes* sebagai salah satu algoritma yang efektif dalam kasus klasifikasi, termasuk prediksi pertandingan sepak bola. Pertimbangan berbagai faktor memungkinkan model ini memberikan estimasi probabilistik terhadap tiga kemungkinan: menang, kalah, atau imbang.

Supervised learning adalah metode dalam *machine learning* yang melatih model menggunakan data berlabel untuk memprediksi atau mengklasifikasikan data baru. Model ini belajar dari pola hubungan antara *input* dan *output* pada data historis, kemudian digunakan untuk memprediksi hasil berdasarkan variabel tertentu (Liu dan Wu 2012). Dalam prediksi pertandingan sepak bola, *supervised learning* digunakan untuk menganalisis data historis yang sudah dikategorikan, seperti kemenangan, kekalahan, atau hasil imbang. Algoritma *Naïve Bayes* sering diterapkan karena kemampuannya menangani dataset kecil serta kemudahan implementasinya. Model ini dievaluasi menggunakan metrik seperti akurasi dan presisi guna meningkatkan keakuratan prediksi (Tiwari 2022).

Metode *Machine learning* digunakan untuk prediksi pertandingan sepak bola salah satunya menggunakan algoritma *Naïve Bayes*. Berdasarkan penelitian terdahulu, Darmasatyo mengatakan bahwa menggunakan metode *Naïve Bayes* dapat meningkatkan akurasi sebanyak 63% (Prabowo 2020). Fabio melakukan penelitian terhadap rekomendasi posisi pemain dengan membandingkan posisi asli pemain dan posisi hasil rekomendasi dengan hasil akurasi sebesar 65% (F. F. Pratama dan Nurhasanah 2020). Penelitian lain yang dilakukan oleh Sherlyn yaitu memprediksi potensi pendaftaran siswa di SMK dengan menggunakan model *Naïve Bayes* dimana hasilnya didapatkan tingkat akurasi sebesar 86% (Putri 2023). Selain

itu juga , Riski melakukan penelitian terhadap nilai akademis mahasiswa dengan menggunakan algoritma *Naïve Bayes* dimana memanfaatkan perhitungan probabilitas dan statistik data sebelumnya untuk memprediksi data di masa depan berdasarkan pada data sebelumnya (Annisa dan Sasongko 2020).

Berdasarkan pernyataan yang dilakukan oleh (Prabowo 2020), masih terdapat celah dalam pengembangan model prediksi hasil pertandingan sepak bola, terutama dalam hal eksplorasi parameter yang berpengaruh serta keterbatasan metode klasifikasi yang diterapkan. Pengambilan keputusan berbasis risiko dimana pada probabilitas ini memberikan fleksibilitas untuk mempertimbangkan skenario yang mungkin terjadi, meningkatkan nilai tim untuk para investor dengan melihat seberapa besar kemungkinan kemenangan suatu tim, strategi dalam tim jika peluang kemenangan kecil maka pelatih bisa menyesuaikan pola permainannya, untuk keperluan riset lanjutan, memberikan analisis pertandingan berbasis angka yang kredibel dan transparan. Selain itu, akurasi prediksi masih dapat ditingkatkan dengan pendekatan yang lebih optimal dalam pengolahan data historis pertandingan. Untuk mengatasi keterbatasan tersebut, penelitian ini menerapkan algoritma *Gaussian Naïve Bayes* yang dirancang untuk memproses fitur numerik melalui asumsi distribusi *Gaussian*. Penelitian ini juga melibatkan beberapa langkah teknis seperti *preprocessing data (cleaning, encoding, dan normalisasi menggunakan StandardScaler)*, penambahan fitur turunan melalui *feature engineering*, pembagian data menggunakan *train–test split*, serta optimasi parameter *var_smoothing* melalui *Grid Search* dengan validasi *Stratified K-Fold*. Evaluasi model dilakukan menggunakan akurasi, *error rate*, *classification report*,

confusion matrix, serta analisis *feature importance* untuk mengetahui fitur yang paling berpengaruh dalam proses prediksi. Oleh karena itu, penelitian ini bertujuan mengoptimalkan model prediksi dengan menerapkan algoritma *Naïve Bayes*, yang dikenal sebagai metode probabilistik sederhana namun cukup efektif dalam klasifikasi berbasis data historis untuk meningkatkan akurasi prediksi.

Hasil penelitian ini nantinya diharapkan dapat memberikan kontribusi signifikan dalam pengembangan model prediksi hasil pertandingan sepak bola liga Inggris berdasarkan data historis yang menggunakan metode *Naïve Bayes*. Dengan demikian, penelitian ini mengusulkan implementasi algoritma *Naïve Bayes*, yang mengandalkan probabilitas bersyarat untuk menganalisis berbagai faktor, seperti jumlah gol, statistik tim, riwayat pertemuan, serta faktor lainnya. Model ini akan menghasilkan prediksi dalam bentuk persentase probabilitas untuk kemenangan, kekalahan, atau hasil imbang. Diharapkan, model ini dapat meningkatkan akurasi prediksi serta menjadi referensi dalam analisis pertandingan sepak bola.

1.2 Rumusan Masalah

- a. Bagaimana mengoptimalkan hasil prediksi pertandingan sepak bola menggunakan algoritma *Naïve Bayes* ?
- b. Apa saja faktor-faktor yang berpengaruh dalam menentukan hasil pertandingan sepak bola berdasarkan model *Naïve Bayes*, serta bagaimana tingkat akurasi model prediksi tersebut?

1.3 Tujuan Penelitian

- a. Menerapkan algoritma *Naïve Bayes* dalam memprediksi hasil pertandingan sepak bola berdasarkan data-data sebelumnya dalam bentuk persentase.
- b. Menganalisis faktor-faktor yang berpengaruh dalam menentukan hasil pertandingan sepak bola serta mengukur tingkat akurasi prediksi menggunakan algoritma *Naïve Bayes*.

1.4 Manfaat Penelitian

- a. Manfaat akademik

Penelitian ini menambah wawasan dalam bidang machine learning dalam penerapan *Naïve Bayes* untuk prediksi hasil pertandingan sepak bola serta memberikan kontribusi terhadap pengembangan model prediksi berbasis statistik dan probabilitas dalam dunia olahraga khususnya sepak bola.

- b. Manfaat praktis

Penelitian ini membantu dalam memahami faktor-faktor utama yang memengaruhi hasil pertandingan dan meningkatkan strategi analisis pertandingan bagi pelatih dan tim model ini dapat dijadikan alat bantu untuk menyusun strategi permainan dengan mempertimbangkan performa tim dan lawan. Bagi penggemar sepak bola penelitian ini memberikan wawasan mengenai cara kerja prediksi berbasis data yang dapat digunakan untuk referensi dalam berdiskusi. Penelitian ini juga dapat menjadi referensi untuk model prediksi olahraga lain dengan menggunakan *machine learning*, sehingga dapat membuka peluang penelitian lebih lanjut.

1.5 Batasan Masalah

Berikut merupakan batasan masalah yang diterapkan dalam penelitian ini :

- a. Dataset yang digunakan dalam penelitian ini berasal dari situs *football-data* yang dapat di akses melalui tautan : <https://www.football-data.co.uk/>
- b. Penelitian ini berfokus pada penerapan dan analisis algoritma *Naïve Bayes* dalam memprediksi hasil pertandingan sepak bola tanpa mengembangkan aplikasi atau sistem berbasis machine learning yang terintegrasi.
- c. Model prediksi yang digunakan berbentuk presentase tanpa memperhitungkan faktor non-teknis seperti psikologi pemain, cuaca, atau keputusan wasit.
- d. Evaluasi model dilakukan dengan mengukur tingkat akurasi prediksi menggunakan algoritma *Naïve Bayes* .
- e. Penelitian ini dilakukan dalam lingkungan pemrograman berbasis *Python*, menggunakan pustaka machine learning seperti *scikit-learn* untuk membangun dan menguji model prediksi.

1.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini dirancang untuk memberikan gambaran yang runtut mengenai alur penelitian. Adapun susunan sistematikanya meliputi:

Bab I Pendahuluan: memuat latar belakang penelitian, perumusan masalah, batasan penelitian, tujuan, manfaat, serta sistematika penulisan skripsi.

Bab II Tinjauan Pustaka: berisi teori-teori yang relevan, landasan konseptual yang digunakan sebagai dasar penelitian, serta hasil penelitian sebelumnya.

Bab III Metodologi Penelitian: menjelaskan metode yang diterapkan, mencakup roadmap penelitian, uraian metodologi, tahapan penelitian, serta penyajian diagram alir dan *fishbone* diagram untuk memperjelas alur penelitian.

Bab IV Hasil dan Pembahasan: menyajikan hasil penelitian, analisis data, evaluasi model, serta pembahasan yang mengaitkan temuan dengan teori maupun penelitian terdahulu.

Bab V Penutup: berisi kesimpulan dari penelitian yang dilakukan dan memberikan saran untuk penelitian berikutnya.