

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sepak bola merupakan salah satu olahraga yang paling populer dan digemari di seluruh dunia, dengan jumlah penggemar yang fantastis mencapai 3,5 miliar orang menurut survei FIFA Global Football Development (GFD) pada tahun 2023. Popularitasnya yang luar biasa bahkan sering kali dianggap tidak memerlukan pembuktian ilmiah. Dalam permainan ini, setiap pemain memiliki peran strategis sesuai dengan posisinya, yaitu pemain bertahan yang terdiri dari kiper (goalkeeper) dan bek (defender), pemain tengah (midfielder), serta penyerang (forward). Ketiga posisi ini memiliki kontribusi penting dalam menentukan keberhasilan tim di lapangan [1]. Penempatan posisi pemain sangat berpengaruh terhadap kerangka tim dalam sebuah pertandingan, di mana kurang optimalnya penempatan posisi dapat menjadi faktor kekalahan dalam tim. Sebagai contoh, perubahan posisi Septian D. Maulana dari gelandang menjadi penyerang dapat memengaruhi kontribusinya terhadap tim. Contoh lain adalah Firman Utina, yang menunjukkan peningkatan performa setelah posisinya diubah dari penyerang menjadi gelandang [2]. Sementara itu, di dunia sepak bola profesional, La Liga dikenal sebagai liga sepak bola tingkat teratas dalam sistem liga Spanyol. Liga ini selalu menjadi pusat perhatian karena pertandingan-pertandingan sengit dan duel menarik antara klub-klub raksasa yang terus memikat hati penggemar sepak bola dari berbagai penjuru dunia. Namun, memprediksi klub mana yang akan menjuarai La Liga bukanlah hal yang mudah. Klub sering kali menggunakan metode klasifikasi berdasarkan track record atau catatan kinerja mereka untuk memperkirakan peluang menjadi juara. Analisis ini melibatkan penilaian terhadap peringkat akhir klasemen liga beberapa tahun sebelumnya, yang menunjukkan bahwa klub harus mampu mencapai nilai tertentu secara konsisten untuk meraih gelar juara. Dengan demikian, performa yang solid dan stabilitas tim menjadi faktor penentu utama dalam meraih kesuksesan di liga yang kompetitif ini [3].

Klasifikasi adalah salah satu metode penting dalam data mining dan machine learning yang bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang telah dipelajari dari data sebelumnya [4]. Proses ini memungkinkan sistem untuk mempelajari hubungan antara fitur-fitur dalam data dan menghasilkan model prediktif yang dapat digunakan untuk berbagai aplikasi [5]. Metode klasifikasi digunakan dalam berbagai bidang, termasuk analisis sentimen untuk memahami opini publik, prediksi penyakit untuk mendukung diagnosis medis, dan pengelompokan data olahraga untuk menganalisis performa atlet atau tim [6]. Salah satu algoritma yang sering digunakan dalam klasifikasi adalah Logistic Regression, yang bekerja dengan memprediksi probabilitas suatu kejadian berdasarkan fungsi logistik atau sigmoid [4]. Logistic Regression dikenal karena kesederhanaannya dalam implementasi, kemudahan interpretasi hasil, dan kemampuannya menangani data dengan hubungan linear, menjadikannya pilihan yang populer dalam berbagai penelitian [6]. Selain itu, algoritma ini mampu menangani klasifikasi baik untuk data biner maupun multi-kelas, memberikan fleksibilitas dalam berbagai konteks aplikasi. Dalam klasifikasi multiclass, suatu data dapat dikategorikan ke dalam lebih dari dua kelas. Salah satu metode yang digunakan untuk menangani klasifikasi multiclass adalah Multinomial Logistic Regression (MLR), yang merupakan perluasan dari Logistic Regression [5]. Dengan keunggulan-keunggulan tersebut, Logistic Regression tetap menjadi salah satu metode yang banyak digunakan dalam pengembangan model klasifikasi modern.

Berdasarkan penelitian sebelumnya, proses klasifikasi dilakukan dengan membandingkan tingkat akurasi antara algoritma Logistic Regression dan Multinomial Naïve Bayes dalam kasus analisis sentimen ulasan film IMDB. Hasilnya menunjukkan bahwa algoritma Logistic Regression mencapai akurasi yang tinggi sebesar 89,32%, sedangkan metode Multinomial Naïve Bayes memperoleh akurasi sebesar 85,28% [7]. Penelitian selanjutnya tentang klasifikasi dilakukan dengan membandingkan algoritma Logistic Regression, SVM, dan Artificial Neural Network pada kasus prediksi penyakit jantung. Hasilnya menunjukkan bahwa Logistic Regression memperoleh akurasi tertinggi sebesar 86% pada skenario pembagian data 80:20 dibandingkan dengan algoritma lainnya

[8]. Kemudian, penelitian lain menggunakan Logistic Regression dan SVM pada kasus prediksi kegagalan mesin dengan teknik SMOTE untuk mengatasi ketidakseimbangan data. Logistic Regression mencapai akurasi sebesar 84%, lebih rendah dibandingkan SVM yang mencapai akurasi 93% [9]. Selanjutnya, pada penelitian klasifikasi email phishing, Logistic Regression dibandingkan dengan Support Vector Classification, di mana Logistic Regression memperoleh akurasi sebesar 96,5%, sedikit lebih rendah dibandingkan Support Vector Classification yang mencapai akurasi 97,4% [10].

Berdasarkan uraian diatas, penelitian ini akan melibatkan serangkaian tahapan, Algoritma Logistic Regression akan diterapkan untuk mengklasifikasikan posisi pemain sepak bola , disertai dengan beberapa upaya optimasi melalui *tuning hyperparameter* guna menilai tingkat akurasi dari Logistic Regression tersebut.

1.2 Rumusan Masalah

Berdasarkan hasil penjelasan dari latar belakang yang telah disampaikan, dapat dirumuskan masalah sebagai berikut:

1.2.1 Permasalahan

Penempatan posisi pemain yang tidak sesuai dapat berdampak negatif pada performa tim sepak bola. Oleh karena itu, diperlukan metode klasifikasi yang efektif untuk mengelompokkan pemain berdasarkan posisi mereka, guna mendukung pengambilan keputusan strategis yang lebih baik dalam tim.

1.2.2 Solusi Permasalahan

Algoritma *Logistic Regression* digunakan untuk melakukan klasifikasi posisi pemain sepak bola berdasarkan atribut-atribut pemain. Selain itu, dilakukan tuning hyperparameter pada algoritma *Logistic Regression* guna meningkatkan akurasi hasil klasifikasi yang diperoleh.

1.2.3 Pertanyaan Penelitian

1. Seberapa tinggi tingkat akurasi yang mampu dicapai oleh algoritma *Logistic Regression* dalam mengklasifikasikan posisi pemain sepak bola?

2. Apakah proses optimasi tuning hyperparameter pada algoritma *Logistic Regression* memberikan hasil yang tinggi pada akurasi klasifikasi posisi pemain sepak bola?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Penelitian ini bertujuan untuk:

1. Menggunakan *Logistic Regression* dalam mengklasifikasi posisi pemain
2. Mengevaluasi performa model berdasarkan akurasi, precision, recall, dan confusion matrix
3. Menentukan kombinasi parameter terbaik untuk *Logistic Regression* menggunakan hyperparameter tuning

1.3.2 Manfaat

Penelitian ini memberikan kontribusi dalam bidang data science, khususnya dalam penerapan *Logistic Regression* untuk klasifikasi multiclass, yang dapat menjadi referensi bagi penelitian lain terkait analisis data olahraga. Selain itu, penelitian ini memberikan wawasan kepada pelatih dan analis olahraga dalam menentukan posisi pemain secara objektif berdasarkan data statistik, serta mendukung pengembangan sistem otomatis untuk membantu proses penentuan posisi pemain dalam tim.

1.4 Batasan Masalah

Penelitian memiliki beberapa Batasan antara lain:

1. Dataset yang digunakan adalah dataset dari website kaggle berdasarkan game EA FC SPORTS 25 yang berisikan 529 pemain dari *La Liga Hypermotion*.
2. Penelitian hanya menggunakan atribut statistik dan tidak mempertimbangkan faktor non teknis seperti preferensi pelatihan atau gaya bermain.
3. Model klasifikasi hanya menggunakan algoritma *Logistic Regression*.

1.5 Sistematika Laporan Skripsi

Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi pendahuluan dari penelitian. Di dalamnya akan dijelaskan latar belakang masalah yang muncul, solusi yang diajukan untuk mengatasi masalah tersebut, tujuan dan manfaat penelitian, batasan-batasan dalam penelitian, serta sistematika penulisan proposal.

BAB II TINJAUAN PUSTAKA

Bab ini memuat kajian pustaka yang menguraikan materi dan metode yang menjadi landasan teori dalam penelitian. Materi yang dibahas meliputi Sepakbola, Algoritma *Logistic Regression*, Klasifikasi, Preprocessing, serta penelitian-penelitian terdahulu yang relevan.

BAB III METODE PENELITIAN

Bagian ini membahas Metode Penelitian yang mencakup penjelasan tentang dataset yang digunakan, arsitektur sistem, tahapan pengujian, serta skenario uji coba yang akan dilakukan.

BAB IV ANALISIS DAN HASIL PENELITIAN

Bagian ini membahas hasil dan pembahasan penelitian, termasuk hasil yang diperoleh dari penerapan algoritma *Logistic Regression* dalam memecahkan masalah. Selanjutnya, dilakukan pengujian untuk mengukur dan menganalisis kinerja serta efektivitas dari hasil yang dicapai.

BAB V PENUTUP

Bagian ini berisi kesimpulan dari implementasi yang telah dilakukan, serta saran-saran yang ditujukan untuk pengembangan dan peningkatan penelitian di masa mendatang.