

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sepak bola adalah olahraga yang sangat penting dalam kehidupan masyarakat di seluruh dunia. Selain menjadi sarana hiburan, sepak bola juga berkembang menjadi sebuah industri yang melibatkan berbagai aspek strategis. Dalam hal manajemen, klub sepak bola harus mampu mengelola sumber daya dengan baik, mulai dari pemain, staf pelatih, hingga perencanaan pertandingan jangka pendek dan jangka panjang. Di sisi ekonomi, sepak bola memiliki peran besar dalam pendapatan klub, hak siar, sponsor, penjualan tiket, serta memberikan kontribusi terhadap perekonomian daerah dan negara. Dari segi ilmu pengetahuan, sepak bola terus berkembang dengan penerapan pendekatan ilmiah dalam menganalisis performa pemain, meningkatkan kebugaran, dan merancang strategi permainan. Dengan meningkatnya persaingan antar klub, keputusan dalam sepak bola modern semakin membutuhkan dasar yang logis dan didukung data. Oleh karena itu, keberhasilan sebuah tim tidak hanya tergantung pada kemampuan individu pemain, tetapi juga pada keakuratan dalam memilih formasi yang tepat serta mengelola permainan secara menyeluruh dan terencana. Pemilihan formasi ini seringkali menjadi kunci untuk merespons kekuatan dan kelemahan lawan yang dihadapi (Wahyudi, 2023).

Seiring dengan kemajuan teknologi informasi, sepak bola kini dipahami sebagai permainan yang dapat dianalisis secara ilmiah menggunakan data, bukan hanya dinilai berdasarkan hasil akhir atau pengamatan visual semata. Data pertandingan yang sebelumnya hanya digunakan sebagai catatan statistik sederhana kini berkembang menjadi sumber informasi strategis yang bernilai tinggi. Berbagai aspek pertandingan seperti lawan, lokasi pertandingan, posisi lawan di klasemen, rata-rata gol yang dicetak lawan, rata-rata gol kebobolan lawan, dan penguasaan bola serta efektivitas formasi dapat dianalisis secara sistematis menggunakan pendekatan berbasis data. (Fujii, 2024) menjelaskan bahwa pemanfaatan machine learning dalam bidang olahraga, memungkinkan analisis permainan secara lebih mendalam, berdasarkan data historis dan karakteristik lawan. Pendekatan tersebut

relevan untuk diterapkan dalam sepak bola karena permainan ini melibatkan kerja sama antar pemain, penerapan formasi yang beragam, serta kondisi pertandingan yang dapat berubah selama berlangsungnya permainan.

Pada tingkat sepak bola profesional, pemanfaatan analisis data telah menjadi bagian penting dalam proses perencanaan formasi dan evaluasi kinerja tim. Data pertandingan digunakan untuk menilai efektivitas formasi yang diterapkan, sehingga pelatih dapat menyesuaikan pola permainan tim melalui pemilihan formasi yang sesuai dengan karakteristik lawan. Informasi yang dihasilkan dari analisis tersebut memberikan dasar yang lebih terukur dalam pengambilan keputusan teknis. Oleh karena itu, penggunaan *data analytics* berperan dalam meningkatkan efektivitas formasi dalam pertandingan serta membantu menjaga kestabilan performa tim sepanjang musim kompetisi.

Di Indonesia, sepak bola merupakan olahraga dengan tingkat popularitas yang sangat tinggi. Antusiasme masyarakat terhadap kompetisi domestik, khususnya Liga 1, menunjukkan besarnya perhatian publik terhadap perkembangan sepak bola nasional. Namun, tingginya minat tersebut belum sepenuhnya diiringi dengan pemanfaatan pendekatan ilmiah berbasis data dalam kajian akademik maupun praktik kepelatihan. (Pradana et al., 2024) menunjukkan bahwa kajian ilmiah mengenai sepak bola di Indonesia masih terbatas, sehingga pengembangan pengetahuan berbasis riset data belum optimal untuk mendukung peningkatan kualitas sepak bola nasional.

Keterbatasan kajian akademik berbasis data menyebabkan masih terbatasnya referensi ilmiah yang dapat dimanfaatkan sebagai dasar dalam menentukan formasi permainan pada sepak bola Indonesia. Akibatnya, banyak keputusan teknis dan taktis masih bergantung pada pengalaman dan intuisi semata, tanpa dukungan analisis kuantitatif yang memadai. Padahal, pendekatan berbasis data memiliki potensi besar untuk membantu klub dalam mengevaluasi performa tim secara lebih objektif dan berkelanjutan.

Salah satu klub yang menarik untuk dikaji dalam konteks ini adalah Persebaya Surabaya. Sebagai klub dengan sejarah panjang dan basis pendukung yang besar, tim yang berjulukan Bajul Ijo ini menunjukkan performa yang kompetitif dalam kompetisi BRI Liga 1 musim 2024/2025. Hingga pekan ke-27, Persebaya berada di

papan atas klasemen dengan perolehan 48 poin dari 27 pertandingan, yang mencerminkan kemampuan tim dalam bersaing di level tertinggi kompetisi nasional (Saleh, 2025). Capaian tersebut menunjukkan bahwa Persebaya memiliki potensi besar untuk terus meningkatkan kualitas performa tim.

Meskipun Persebaya Surabaya berada di papan atas klasemen, hal itu tidak selalu mencerminkan konsistensi dalam menjalankan formasi permainan yang adaptif terhadap lawan. Sepanjang musim, tim ini mengalami beberapa perubahan, seperti pergantian pelatih dan variasi dalam penerapan formasi. Beberapa kajian dan laporan analisis pertandingan menunjukkan bahwa formasi yang digunakan terkadang belum sepenuhnya sesuai dengan karakteristik lawan yang dihadapi (Putra, 2024). Hal ini menunjukkan adanya tantangan dalam menemukan formasi yang bisa diterapkan secara konsisten.

Perubahan formasi yang terjadi secara berulang menunjukkan bahwa proses penentuan keputusan taktis dalam sepak bola bersifat kompleks dan melibatkan berbagai variabel. Pelatih perlu mempertimbangkan karakteristik lawan, serta tuntutan kompetisi dalam waktu yang terbatas. Tanpa dukungan analisis berbasis data, keputusan tersebut berpotensi menjadi kurang optimal dan sulit dievaluasi secara objektif. Oleh karena itu, diperlukan suatu pendekatan yang dapat membantu pelatih dalam mengklasifikasikan formasi berdasarkan pola data pertandingan yang telah terjadi sebelumnya, khususnya yang berkaitan dengan karakteristik lawan. Dalam konteks tersebut, penerapan machine learning menjadi salah satu alternatif pendekatan yang relevan.

Machine learning bisa memproses banyak data sekaligus dan menemukan hubungan antar variabel yang sulit dilihat saat dianalisis secara manual. Dalam olahraga sepak bola, teknologi ini digunakan untuk menganalisis kinerja pemain dan tim, mengevaluasi efektivitas formasi yang digunakan, serta membantu mengklasifikasikan formasi yang paling sesuai berdasarkan karakteristik lawan dari pertandingan sebelumnya. (Kim, 2023) menunjukkan bahwa formasi sepak bola dapat diidentifikasi dan dianalisis secara otomatis menggunakan pendekatan *machine learning* berbasis data pertandingan. Hasil penelitian tersebut menunjukkan bahwa pola formasi memiliki karakteristik yang dapat dipelajari oleh

model sehingga dapat digunakan untuk mendukung analisis taktis dalam sepak bola.

Random Forest adalah salah satu algoritma pengajaran mesin yang paling banyak digunakan dalam masalah klasifikasi karena sangat baik untuk menangani data yang kompleks dan *non-linear*. Selain itu, karena menggunakan pendekatan pengajaran kelompok, yang menggabungkan banyak pohon keputusan dalam proses pengambilan keputusan, algoritma ini mampu menghasilkan hasil klasifikasi yang stabil. (Midoul, 2026) Untuk tugas klasifikasi, *Random Forest* adalah salah satu algoritma yang paling efektif karena menggunakan mekanisme voting mayoritas, di mana hasil akhir klasifikasi didasarkan pada pilihan yang paling banyak diberikan oleh semua pohon keputusan yang telah dibuat. Karena proses pembentukan setiap pohon keputusan melibatkan pemilihan sampel dan fitur secara acak, algoritma ini dikenal tahan terhadap overfitting.

Random Forest memiliki keunggulan dalam melakukan klasifikasi dan menemukan tingkat kepentingan setiap fitur *feature importance* yang digunakan dalam model. Melalui kemampuan tersebut, peneliti dapat mengetahui atribut mana yang memiliki pengaruh terbesar terhadap hasil klasifikasi. (Teixeira et al., 2025) menekankan bahwa pemetaan perilaku taktis menggunakan kecerdasan buatan memungkinkan klasifikasi peran individu dan dinamika kolektif tim secara lebih objektif. Dalam konteks penelitian ini, fitur-fitur yang digunakan difokuskan pada karakteristik lawan untuk menentukan penggunaan formasi oleh Persebaya Surabaya. Informasi tersebut dimanfaatkan untuk memahami pola hubungan antara profil lawan dengan formasi yang digunakan, sehingga mendukung proses pengambilan keputusan taktis yang lebih objektif dan berbasis data.

Di tingkat nasional, penelitian mengenai penerapan *machine learning* dalam sepak bola mulai menunjukkan perkembangan. (Triawan et al., 2024) menerapkan metode *Naïve Bayes* untuk mengklasifikasikan tipe penyerang sepak bola Liga Inggris berdasarkan data statistik pemain. Penelitian tersebut berhasil menunjukkan bahwa data statistik dapat dimanfaatkan untuk mengidentifikasi karakteristik pemain menggunakan pendekatan *machine learning*. Namun, penelitian ini masih hanya membahas klasifikasi jenis pemain dan belum menganalisis klasifikasi formasi tim berdasarkan karakteristik tim lawan. Selain itu, penelitian tersebut

menggunakan data statistik individu pemain, sedangkan penelitian ini berfokus pada data pertandingan dan karakteristik lawan sebagai dasar klasifikasi formasi Persebaya Surabaya. Oleh karena itu, masih terdapat peluang penelitian untuk mengembangkan penerapan *machine learning* dalam analisis taktik sepak bola, khususnya pada klasifikasi formasi tim menggunakan metode *Random Forest*.

Berdasarkan penjelasan tersebut, penelitian ini bertujuan untuk membangun model klasifikasi formasi Persebaya Surabaya berdasarkan karakteristik lawan menggunakan metode *Random Forest*. Dengan memanfaatkan data historis pertandingan Liga 1 Indonesia musim 2020–2025, penelitian ini berupaya mengidentifikasi pola hubungan antara karakteristik lawan dengan formasi yang digunakan oleh Persebaya Surabaya. Hasil penelitian diharapkan dapat memberikan gambaran tentang elemen yang memengaruhi penggunaan formasi dalam pertandingan serta memberikan kontribusi dalam pengembangan analisis sepak bola berbasis data, khususnya pada penerapan *machine learning* untuk klasifikasi formasi di Indonesia.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, masalah utama yang akan diteliti dalam penelitian ini dapat dirumuskan seperti berikut:

1. Bagaimana merancang dan membangun model klasifikasi formasi Persebaya Surabaya berdasarkan karakteristik lawan menggunakan algoritma *Random Forest* dengan memanfaatkan data historis pertandingan Liga 1 Indonesia musim 2020–2025?
2. Bagaimana tingkat kemampuan model klasifikasi yang dihasilkan berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* dalam mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan?
3. Fitur karakteristik lawan apa saja yang paling berpengaruh *feature importance* terhadap hasil klasifikasi formasi Persebaya Surabaya menggunakan metode *Random Forest*?

1.3 Batasan Masalah

Agar penelitian ini lebih jelas dan fokus pada tujuan utamanya, penulis menetapkan beberapa batasan dalam penelitian ini. Berikut penjelasannya:

1. Data yang digunakan merupakan data sekunder berupa data historis pertandingan Liga 1 Indonesia musim 2020–2025 yang diperoleh dari platform penyedia statistik sepak bola, yaitu *Sofascore*. Data yang digunakan meliputi karakteristik lawan seperti lokasi pertandingan *home/away*, posisi lawan di klasemen, rata-rata gol yang dicetak lawan, rata-rata gol yang kebobolan lawan, serta informasi lain yang relevan dengan penelitian.
2. Penelitian ini hanya berfokus pada klub Persebaya Surabaya sebagai objek penelitian. Analisis dilakukan untuk mengklasifikasikan formasi yang digunakan Persebaya Surabaya berdasarkan karakteristik lawan yang dihadapi pada kompetisi Liga 1 Indonesia.
3. Metode yang digunakan dalam penelitian ini adalah algoritma *Random Forest* sebagai metode klasifikasi. Penelitian tidak melakukan perbandingan performa dengan algoritma *machine learning* lainnya.
4. Kelas output pada penelitian ini dibatasi pada tiga formasi utama Persebaya Surabaya, yaitu 4-3-3, 4-2-3-1, dan 4-4-2, sesuai hasil seleksi label pada tahap pra-pemrosesan data.
5. Penelitian ini tidak membahas prediksi susunan pemain *starting eleven*, pergantian pemain, strategi permainan, maupun hasil akhir pertandingan. Penelitian hanya berfokus pada klasifikasi formasi berdasarkan karakteristik lawan menggunakan data historis pertandingan.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang telah dibahas sebelumnya, tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun model klasifikasi formasi Persebaya Surabaya berdasarkan karakteristik lawan menggunakan algoritma *Random Forest* dengan memanfaatkan data historis pertandingan Liga 1 Indonesia musim 2020–2025.

2. Mengevaluasi kinerja model klasifikasi yang dihasilkan menggunakan parameter *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui kemampuan model dalam mengklasifikasikan formasi Persebaya Surabaya.
3. Menganalisis tingkat kepentingan fitur *feature importance* untuk mengidentifikasi karakteristik lawan yang paling berpengaruh terhadap hasil klasifikasi formasi Persebaya Surabaya.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Bagi tim kepelatihan Persebaya Surabaya
 - a. Memberikan informasi berbasis data mengenai pola penggunaan formasi Persebaya Surabaya berdasarkan karakteristik lawan yang dihadapi.
 - b. Menjadi bahan pendukung dalam proses analisis pertandingan melalui pemanfaatan pendekatan *machine learning* untuk mengidentifikasi kecenderungan penggunaan formasi.
2. Bagi analis pertandingan
 - a. Membantu proses analisis pertandingan melalui pemanfaatan data historis yang diolah secara otomatis sehingga dapat meningkatkan efisiensi dalam pengumpulan dan pengolahan informasi.
 - b. Memberikan informasi tambahan berupa hasil klasifikasi formasi Persebaya Surabaya pada pertandingan berikutnya berdasarkan karakteristik lawan, sehingga dapat digunakan sebagai pendukung dalam proses analisis pertandingan.
3. Bagi penulis
 - a. Sebagai bentuk penerapan ilmu Informatika, khususnya pada bidang *Data Mining*, *Machine Learning*, dan Kecerdasan Buatan dalam menyelesaikan permasalahan nyata di bidang olahraga.
 - b. Menambah pengetahuan dan pengalaman dalam melakukan penelitian yang memanfaatkan data historis pertandingan sepak bola sebagai objek kajian.

- c. Memenuhi salah satu syarat akademik yang dibutuhkan untuk menyelesaikan pendidikan Strata-1 (S1) di Program Studi Informatika.

1.6 Metode Penelitian

Data yang digunakan dalam penelitian ini berasal dari data historis pertandingan Persebaya Surabaya di Liga 1 Indonesia musim 2020–2025 yang diperoleh dari sumber sekunder, yaitu platform Sofascore, dan diklasifikasikan berdasarkan karakteristik lawan menggunakan algoritma Random Forest.

Untuk memastikan bahwa proses penelitian berjalan secara terstruktur, tahapan penelitian disusun secara sistematis, mulai dari pengumpulan data hingga implementasi sistem berbasis web. Tahapan-tahap yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan Data

Pada saat ini, data dari semua pertandingan Persebaya Surabaya di Liga 1 Indonesia musim 2020–2025 dikumpulkan. Data disusun ke dalam dataset setelah dikumpulkan secara manual dari platform *Sofascore*. Ini termasuk informasi pertandingan, statistik performa tim, dan karakteristik lawan yang digunakan dalam penelitian.

2. Pra-pemrosesan Data

Sebelum dataset digunakan dalam proses pembangunan model klasifikasi menggunakan algoritma Random Forest, tahap pra-pemrosesan data dilakukan. Tahap ini mencakup penerapan injeksi logika berbasis aturan, seleksi label formasi, seleksi fitur, transformasi data kategorikal menggunakan kode label, dan pemisahan antara variabel input fitur dan variabel keluaran label. Tujuan dari upaya ini adalah untuk menghasilkan dataset yang memenuhi kebutuhan model sehingga dapat digunakan semaksimal mungkin selama proses pelatihan dan evaluasi.

3. Pembagian Dataset

Setelah melewati tahap pra-pemrosesan, dataset tersebut kemudian dibagi menjadi dua bagian, yaitu data latih yang digunakan untuk melatih model dan data uji yang digunakan untuk menguji kinerja model. Pembagian data dilakukan dengan cara *train_test_split*, yaitu 80% digunakan untuk data latih dan 20% untuk data uji. Data latih digunakan untuk membuat model,

sedangkan data uji digunakan untuk mengetahui seberapa baik model tersebut bekerja.

4. Pembangunan Model Random Forest

Pada tahap ini, proses pembuatan model dilakukan dengan menggunakan algoritma *Random Forest*. Model tersebut dilatih dengan data latihan agar bisa mempelajari pola hubungan antara karakteristik lawan dengan formasi yang dipakai oleh Persebaya Surabaya. Hasil dari tahap ini adalah model klasifikasi yang sudah bisa digunakan untuk mengklasifikasikan formasi.

5. Evaluasi Model

Dalam tahap evaluasi, kemampuan model untuk mengklasifikasikan formasi Persebaya Surabaya dinilai dengan menggunakan metrix *accuracy*, *precision*, *recall*, *F1-score*, dan confusion matrix. Hasil evaluasi digunakan untuk menentukan tingkat performa model yang telah dibangun.

6. Analisis Feature Importance

Setelah model dinilai, dilakukan pengecekan pentingnya setiap fitur untuk mengetahui seberapa besar pengaruhnya terhadap hasil klasifikasi. Analisis ini bertujuan untuk mengetahui fitur-fitur apa saja yang paling berpengaruh dalam membantu model *Random Forest* membuat keputusan.

7. Implementasi Sistem Berbasis Web

Model *random forest* yang sudah dibuat dan dievaluasi selanjutnya diaplikasikan ke dalam aplikasi berbasis web dengan menggunakan *framework Streamlit*. Sistem yang dibuat memungkinkan pengguna menginput karakteristik lawan dan mendapatkan hasil klasifikasi formasi Persebaya Surabaya secara interaktif melalui situs web.

8. Lingkungan Pengembangan

Tahap ini membahas *software* dan *library* yang digunakan dalam penelitian tersebut. Proses pengolahan data, pembuatan model, evaluasi, hingga pemasangan sistem dilakukan di *Google Colaboratory* menggunakan bahasa pemrograman *Python* dan beberapa *library* pendukung seperti *Pandas*, *NumPy*, *Scikit-Learn*, *Joblib*, serta *Streamlit*.

Melalui tahapan penelitian tersebut, diharapkan model yang dibangun mampu mengklasifikasikan formasi Persebaya Surabaya berdasarkan karakteristik lawan

yang dihadapi serta dapat diimplementasikan dalam bentuk aplikasi berbasis web yang mudah digunakan.

1.7 Sistematika Penulisan

Sistematika penulisan dalam proposal penelitian ini dibuat agar dapat menjelaskan dengan jelas urutan pembahasan yang terdapat pada setiap bab. Sistematika penulisan yang digunakan adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab ini membahas awal-awal tentang topik yang diteliti, termasuk latar belakang masalah, pernyataan masalah, batasan penelitian, tujuan dan manfaat yang ingin dicapai, serta penjelasan mengenai cara penyusunan proposal penelitian secara sistematis.

BAB II : TINJAUAN PUSTAKA

Bab ini menceritakan berbagai teori, konsep, dan hasil penelitian sebelumnya yang berkaitan dengan topik yang dibahas. Semua teori yang dibahas digunakan sebagai dasar yang kokoh untuk menjelaskan isu yang diteliti dan membantu proses analisis dalam penelitian ini.

BAB III : METODE PENELITIAN

Dalam bab ini dijelaskan secara rapi dan terstruktur tahapan-tahapan yang dilakukan selama proses penelitian. Setiap langkah dalam proses penelitian dijelaskan secara rinci dan bertahap agar memberikan pemahaman yang jelas mengenai cara pelaksanaan penelitian secara keseluruhan.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini membahas hasil dari penelitian yang sudah dilakukan, mencakup cara mengolah data, proses pembelajaran model menggunakan algoritma *Random Forest*, pengujian kemampuan model dengan bantuan *confusion matrix*, serta perhitungan *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, juga dijelaskan arti dan pentingnya setiap variabel yang digunakan dalam penelitian tersebut. Selain itu, bab ini juga menjelaskan bagaimana menerapkan model tersebut ke dalam aplikasi web dengan menggunakan Streamlit, serta membahas hasil yang berhasil diperoleh dari penerapan tersebut.

BAB V : PENUTUP

Bab ini menjelaskan kesimpulan yang didapat dari hasil penelitian yang sudah dilakukan, serta beberapa saran yang bisa dipakai sebagai acuan untuk mengembangkan penelitian selanjutnya agar hasilnya lebih baik dan lebih maksimal.