

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Statistika merupakan pengetahuan yang berhubungan dengan cara pengumpulan data, penyusunan data, pengolahan data, serta penyajian data berdasarkan kumpulan dan analisis data yang dilakukan. Di sepakbola, statistik digunakan untuk menganalisis kemampuan pemain, perkiraan strategi suatu tim, analisis pertandingan, *scouting* pemain serta kebutuhan data tim dan masih banyak lagi. Di dunia persepakbolaan modern penggunaan statistik sudah marak dilakukan, bahkan sudah menjadi konsumsi publik sehari-hari. Statistik juga merupakan bentuk sederhana dari ilmu sains yang dikombinasikan dengan ilmu olahraga.

Sepakbola merupakan salah satu olahraga yang paling terkenal di dunia dan banyak di gemari oleh masyarakat, orang dewasa, anak-anak, pria dan bahkan wanita pun memainkan dan banyak menyukai olahraga ini. Sepakbola adalah olahraga yang dimainkan oleh dua tim yang masing masing beranggotakan 11 orang (sebelas) orang. Memasuki abad ke-21, olahraga ini telah dimainkan lebih 250 juta orang di 200 lebih negara yang menjadikannya olahraga paling populer di dunia dengan Organisasi sepakbola dunia adalah FIFA (Federation International De Football Association), banyak kejuaran yang telah diselenggarakan oleh berbagai event dunia seperti Piala Dunia, Liga Champion, Piala Eropa, Piala Asia dan lain sebagainya. Dari banyaknya event sepakbola tersebut mampu menghipnotis semua masyarakat untuk menyaksikan untuk menyaksikan pertandingan sepakbola, bukan hanya di Indonesia tapi seluruh lapisan dunia.

Dalam suatu tim membeli pemain tidak menjadi jaminan suatu klub akan memenangkan banyak gelar, sebab banyak faktor lain yang terlibat pada suatu tim dalam memenangkan pertandingan. Hal inilah yang perlu diamati dengan seksama apa yang menjadi faktor yang paling penting bagi setiap klub untuk memenangkan pertandingan. Terdapat banyak sekali kasus diluar dugaan yang terjadi pada

sepakbola, contohnya adalah Yunani yang mampu memenangkan Euro 2004 yang mengalahkan Portugal di final. Dalam konteks prestasi, Leicester City menjadi bukti sah ampuhnya peran analisis data di sepak bola. Sebelumnya, banyak orang yang menganggap kesuksesan Leicester menjuarai Liga Inggris pada 2016 adalah karena keberuntungan. Namun, di balik prestasi fenomenal itu, ada statistika dan mesin rumit yang bekerja. Leicester merupakan salah satu tim di Inggris yang paling getol memakai jasa *Prozone Sports*, perusahaan teknologi dan analisis data olahraga asal Inggris (Harjono, 2018).

Persib atau singkatan dari Persatuan Sepakbola Indonesia Bandung adalah berdiri 14 maret 1933, mereka mempunyai julukan maung Bandung dan Pangeran Biru. Sejak zaman perserikatan permainan Persib Bandung mempunyai ciri khas tersendiri yaitu dalam bahasa sunda “tikteuk” yaitu permainan kaki ke kaki yang mengandalkan kreativitas dan kecepatan penyerangan. Persib Bandung merupakan salah satu klub besar yang ada di Indonesia dan sudah banyak menjuarai liga atau kompetisi yang ada di Indonesia, berkandang di Si Jalak harupat, Soreang, Kabupaten Bandung. Persib memiliki penggemar fanatic yang menyebar di seantero provinsi Jawa barat dan Banten. Memiliki dua fanbase besar yaitu Viking dan Bobotoh yang terkenal di Indonesia dan sangat mencintai Persib Bandung di manapun bermain sehingga sangat menarik untuk dibahas dalam penelitian ini.

Jika berbicara mengenai faktor-faktor yang berpengaruh dalam suatu pertandingan sepak bola, maka disini diperlukan klasifikasi agar memudahkan untuk melihat mana faktor yang paling berpengaruh terhadap kemenangan tim dan salah satu cara yang bisa digunakan adalah metode *random forest*, metode ini dapat mengurutkan mana variabel yang paling berpengaruh dibandingkan dengan variabel yang lain yang tersedia pada data yang digunakan, serta menentukan seberapa besar nilai variabel tersebut untuk membuat peluang kemenangan tim lebih besar dibandingkan peluang tim tidak memenangkan pertandingan.

Banyak sekali jenis metode yang bisa digunakan pada klasifikasi tapi disini peneliti menggunakan *random forest* karena metode ini sangat baik dan mampu mengolah data dalam jumlah besar serta data yang memiliki jumlah variabel yang

banyak. Disisi lain metode ini juga tidak terikat dengan distribusi normal pada datanya sehingga hal tersebut tidak menjadi penghalang dalam melakukan analisis.

Variabel penelitian yang digunakan pada penelitian ini yaitu terdiri dari total tembakan, tembakan *on target*, serangan, serangan berbahaya, sepak pojok, kartu kuning, dan kartu merah dan penguasaan bola. Selain dengan hal itu disini peneliti juga menyesuaikan dengan data-data yang tersedia di lapangan untuk proses analisisnya.

## **1.2 Rumusan Masalah**

Adapun rumusan masalah dalam skripsi ini adalah:

1. Faktor-faktor apa saja yang mempengaruhi kemenangan suatu tim dalam sepakbola?
2. Bagaimana klasifikasi hasil pertandingan sepakbola dengan menggunakan metode random forest?
3. Variabel manakah yang paling berpengaruh dalam pengklasifikasian hasil pertandingan sepakbola?
4. Manakah data yang lebih efektif untuk hasil klasifikasi sepakbola?

## **1.3 Batasan Masalah**

Adapun batasan masalah dalam skripsi ini adalah:

1. Data yang digunakan adalah data hasil pertandingan sepakbola Persib Bandung musim kompetisi 2021-2022
2. Metode yang digunakan dalam penelitian ini adalah Random Forest dengan menggunakan data asli dan data oversampling
3. Data dianalisis menggunakan Microsoft excel dan studio R.

## **1.4 Tujuan dan manfaat**

Tujuan penelitian yang ingin dicapai pada penelitian ini adalah :

1. Untuk mengetahui faktor-faktor apa saja yang mempengaruhi kemenangan suatu tim dalam pertandingan sepakbola
2. Untuk mengetahui seberapa besar faktor tersebut berpengaruh terhadap peluang tim dalam memenangkan pertandingan
3. Untuk mengetahui data yang lebih efektif yang digunakan untuk prediksi hasil sepakbola

Adapun manfaat penelitian ini sebagai berikut :

1. Menambah pengetahuan tentang pengklasifikasian variabel yang ada pada pertandingan sepakbola dengan menggunakan metode random forest.
2. Memberikan pengetahuan tentang besar kecilnya pengaruh suatu variabel terhadap hasil pertandingan sepakbola.

## 1.5 Metode Penelitian

1. Studi Literatur

Tahap Studi Literatur merupakan tahap untuk mengumpulkan data, materi, dan informasi mengenai *Random Forest* dari berbagai sumber, diantaranya buku, jurnal, artikel, video dan lain sebagainya.

2. Analisis

Pada tahap ini, Penulis mengkaji dan menganalisis hasil dari tahap Studi Literatur sesuai dengan masalah yang dipilih dalam Skripsi ini. Kemudian dilakukan pengklasifikasian menggunakan data asli dan data *oversampling* untuk menentukan manakah data yang lebih efektif untuk klasifikasi hasil sepakbola.

3. Simulasi

Sebelum masuk klasifikasi, data akan dibagi terlebih dahulu menjadi data training dan data testing dengan menggunakan bantuan software

R Studio. Kemudian dari hasil tersebut akan dihitung nilai akurasi untuk menentukan data yang terbaik dalam proses klasifikasi.

## 1.6 Sistematika Penulisan

Berdasarkan sistematika penulisannya, skripsi ini terdiri dari lima bab serta daftar pustaka, dimana dalam bab terdapat subbab. Dengan sistematika penulisan sebagai berikut :

### **BAB I : PENDAHULUAN**

Bab ini berisikan tentang pemaparan latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat, metode penelitian dan sistematika penulisan skripsi.

### **BAB II : LANDASAN TEORI**

Bab ini berisi mengenai penjelasan teori-teori yang menjadi landasan pembahasan dan analisis yang dikaji dalam skripsi ini

### **BAB III : METODE PENYELESAIAN PREDIKSI HASIL PERTANDINGAN SEPAKBOLA MENGGUNAKAN RANDOM FOREST**

Bab ini berisi tentang penelitian yang dilakukan dari pengambilan dataset, lalu tahap pengambilan data training dan testing, lalu melakukan *imbalanced data* untuk mencari data *oversampling*. Setelah itu melakukan klasifikasi *Random forest* menggunakan data asli dan data *oversampling*. Dan terakhir mendapatkan hasil klasifikasinya.

**BAB IV : HASIL ANALISIS PERBANDINGAN PERSIB  
BANDUNG DENGAN DATA ASLI DAN DATA  
OVERSAMPLING**

Bab ini berisi penjelasan mengenai hasil pengujian klasifikasi menggunakan Random forest pada hasil pertandingan sepakbola dengan melihat ukuran tingkat variabel terpenting yang menjadi pengaruh tim untuk memenangkan pertandingan.

**BAB V : KESIMPULAN DAN SARAN**

Bab ini berisi penjelasan mengenai beberapa hal yang menjadi kesimpulan atas penelitian yang telah dilakukan serta beberapa saran yang berisi rekomendasi untuk pengembangan tulisan ini.

