

## 1. Pendahuluan

Pada umumnya manusia mampu memprediksi atau memperkirakan suatu kejadian belum terjadi hanya dengan seringnya melihat pola dari suatu kejadian secara berulang. Contohnya adalah prediksi terhadap pertandingan sepakbola antara tim A melawan tim B, dilihat dari sejarah pertandingan kedua tim tersebut dari 2 tahun terakhir, tim A lebih sering menang terhadap tim B. Dengan kenyataan tersebut maka bisa memprediksi pertandingan berikutnya dimenangkan oleh tim A. Kemampuan ini jika diterapkan ke dalam suatu sistem berupa *software* maupun *hardware* sangat berguna dalam banyak pekerjaan, contohnya mengklasifikasikan objek atau barang dalam proses industri, pencarian data dalam web, dan lain-lain. *Data mining* sebagai proses meringkaskan pola tidak diketahui dan dapat ditambang dengan ukuran besar. Ada beberapa macam pengklasifikasi dalam data mining yaitu pohon keputusan (*decision tree*), *Naive Bayes*, *Support Vector Machine* (SVM), dan lainnya [1].

Prediksi adalah suatu proses bisa terjadi di masa mendatang melalui suatu informasi atau data dari masa berlalu dan sekarang telah ada, agar kesalahan dapat dikecilkan. Prediksi tidak harus memberikan jawaban benar tapi berusaha untuk menemukan jawaban bisa terjadi [2].

Juara merupakan individu maupun perkumpulan memenangkan suatu kompetisi liga, olimpiade, atau kontes dalam bidang lain, misal bidang seni, olahraga, bela diri, dan lain-lain. Individu atau tim juga dapat menjadi juara di divisi berbeda [3].

Liga Premier Inggris musim ini adalah musim ketiga puluh dari Liga teratas Inggris, liga sepakbola *professional* divisi paling atas di Inggris sejak dibentuk 1992. Musim 2021/2022 adalah musim ketiga penggunaan sistem yang bisa merekam kejadian di lapangan untuk asisten wasit video (*Video Assistant Referee*). Kompetisi awal musim ini diberitakan 25 Maret 2021 dan jadwal diberitakan 16 Juni 2021. Manchester City adalah juara musim lalu, yaitu musim 2020/2021. Sebanyak 20 klub saling bertanding musim ini, jumlahnya yaitu ada 17 klub klasemen paling atas musim 2020/2021 dan 3 klub promosi divisi kedua liga Inggris musim 2020/2021.

Algoritma *Naive Bayes* merupakan salah satu algoritma dalam teknik klasifikasi. *Naive Bayes* adalah klasifikasi menggunakan probabilitas dan metode statistik ini dikenal sebagai *Teorema Bayes* karena oleh ilmuwan Inggris Thomas Bayes, telah memprediksi peluang masa depan berdasarkan pengalaman masa lalu. *Teorema* tersebut digabungkan dengan *Naive*, mengasumsikan bahwa kondisi antara atribut adalah independen. Klasifikasi *Naive Bayes* mengasumsikan bahwa ada tidaknya fitur tertentu dalam satu kelas tidak ada hubungannya dengan fitur kelas lain [4].

*Naive Bayes* merupakan sebuah metode yang membutuhkan sejumlah kecil data latih untuk menentukan suatu estimasi parameter yang dibutuhkan oleh proses klasifikasi. *Naive Bayes* sering tampil jauh lebih baik dari yang diharapkan dalam situasi kehidupan nyata yang paling kompleks [5].

## 2. Studi Pustaka

Pada jurnal yang berisi prediksi hasil pertandingan Liga Inggris berupa *Home Win* (H), *Away Win* (A), dan *Draw* (D) menggunakan metode *Naive Bayes* membutuhkan jumlah data latih yang kecil untuk menentukan estimasi parameter yang diperlukan pada proses pengklasifikasian. Pada penelitian itu menggunakan 3 skenario pembagian data latih dan data uji yaitu skenario 1 (75%-25%), skenario 2 (80%-20%), dan skenario 3 (90%-10%). Pada keseluruhan skenario menghasilkan nilai akurasi *Naive Bayes* 1 (60,5%), *Naive Bayes* 2 (60,5%), *Naive Bayes* 3 (63,5%) [6].

Pada jurnal yang membahas menentukan *starting eleven* pada sepakbola menggunakan metode *naive bayes* berisi sistem yang menerima input berupa nilai atribut kemampuan dan kondisi pemain yang diproses untuk menghasilkan output berupa rekomendasi untuk dijadikan *starting eleven*. Dalam proses menentukan pemain, nilai atribut kemampuan pemain dilakukan proses *profile matching* untuk menentukan posisi ideal bagi pemain, dari tiap kelompok posisi dilakukan proses identifikasi menggunakan *Naive Bayes* untuk menentukan pemain yang cocok untuk dijadikan *starting eleven*. Pengujian rekomendasi posisi dilakukan dengan hasil akurasi sebesar 65%, sedangkan pengujian pemilihan *starting eleven* dilakukan menggunakan *game football manager* dengan melakukan pertandingan dengan pemilihan pemain secara *default* dan pemilihan pemain hasil rekomendasi masing-masing sebanyak sepuluh kali melawan tim dengan komposisi pemain yang sama, hasil dari pertandingan tersebut dihitung selisih (%) dari rata-rata rating pemain. Hasil yang diberikan setelah digunakan perekomendasi pemilihan pemain kenaikan rata-rata rating tim hanya naik sebesar 0.98% [7].

Pada jurnal yang membahas penentuan posisi ideal pemain dalam sepakbola dengan menggunakan metode *naive bayes* berisi kriteria penilaian yang digunakan yaitu *dribbling*, *passing*, *crossing*, *shooting*, *acceleration*, *agility*, *stamina*, *jumping*, *agression*, *composure*, *creativity*, *decisioning*, *positioning*, *marking* dan *teamwork*. Hasil akhir sistem ini adalah menghasilkan rekomendasi daftar peringkat pemain dari setiap posisi. Peneliti melakukan pengujian *white box* dengan teknik *basis path testing* membuat perkiraan logika yang kompleks untuk mendefinisikan aliran eksekusi dan pengujian *black box* dengan teknik *equivalence partitioning* dengan teknik pengujian yang membagi domain *input*, menentukan kasus pengujian dengan mengungkapkan kelas-kelas kesalahan. Dari pengujian kelayakan sistem yang dilakukan menggunakan kuesioner didapatkan persentase perkategori yaitu 88.25% untuk variabel tampilan, 85.5% untuk variabel kemudahan pengguna, dan 87% untuk variabel kerja sistem [8].

Pada jurnal yang membahas Optimasi Pemilihan Posisi Terbaik Pemain Muda Pada Game Football Manager 2018 Dengan Metode *Naive Bayes* yang mengoptimalkan peran posisi pemain terbaik setiap pemain muda. Kontribusi yang dihasilkan dari penelitian itu yaitu data atribut pemain yang digunakan dalam proses perhitungan lebih banyak dari penelitian-penelitian sebelumnya dan penelitian ini memberikan proses pengujian setelah dilakukan proses perhitungan menggunakan *naive bayes* dimana proses tersebut belum terdapat pada penelitian sebelumnya. Alur kerja pada penelitian tersebut meliputi studi pustaka, penentuan data training, perhitungan jumlah posisi pemain, perhitungan probabilitas tiap posisi pada pemain, penentuan posisi pemain, penentuan kriteria optimal berdasarkan perbandingan rating, pengujian dengan melakukan lima kali pertandingan dengan musuh yang sama, dan perhitungan selisih rating (%). Alur kerja tersebut

dapat memberikan hasil yang cukup memuaskan, yaitu pemain memiliki rating yang lebih banyak ketika menggunakan metode *naïve bayes* dan rata-rata pemain mengalami kenaikan rating sebanyak 8.7% [9].

Perbedaan penelitian pada peneliti di atas adalah mereka menggunakan data dan atribut masing-masing untuk diperhitungkan menjadi akurasi atau nilai.

### 3. Metode Penelitian

#### a. Naïve Bayes

*Naive Bayes* adalah susunan sistem sederhana yang menghitung kumpulan probabilitas dijumlahkan frekuensi dan gabungan nilai dari data set yang telah ada, sedangkan, *Teorema Bayes* merupakan kumpulan atribut independen atau hubungan non-mutual yang dikasih oleh nilai pada variabel kelas [10].

#### b. Rumus Naïve Bayes

Di bawah ini adalah rumus dari metode *Naïve Bayes* :

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$$

dimana :

- D adalah kumpulan data pelatihan.
- h adalah asumsi.
- $P(h \mid D)$  yaitu probabilitas posterior, Contoh : kondisi probabilitas hipotesis h setelah data pelatihan (bukti) ada.
- $P(h)$  yaitu probabilitas awal dari hipotesis h. Kuantitas non-klasik ini biasanya ditemukan lalu memeriksa data masa lalu (data pelatihan).
- $P(D)$  adalah probabilitas sebelumnya dari data pelatihan D. Kuantitas ini biasanya berupa nilai yang konstan,  $P(D) = P(D \mid h)P(h) + P(D \mid \neg h)P(\neg h)$ , yang dapat dengan mudah dihitung ketika menemukan  $P(h \mid D)$  dan  $P(h \mid \neg D)$
- $P(D \mid h)$  adalah peluang D yang berasal dari hipotesis h, dan disebut kemungkinan terjadi. Jumlah ini mudah untuk dihitung asalkan bernilai nilai 1 saat D dan h cocok, dan mengembalikan nilai 0 ketika tidak konsisten.

#### 4. Hasil Penelitian dan Pembahasan

**Tabel 1.** Data Training

POSISI	TIM	MAIN	MENANG	SERI	KALAH	POIN	PREDIKSI
1	MANCHESTER CITY	30	23	4	3	73	LAYAK JUARA
2	LIVERPOOL	30	22	6	2	72	LAYAK JUARA
3	CHELSEA	29	17	8	4	59	LAYAK JUARA
4	SPURS	30	17	3	10	54	TIDAK LAYAK JUARA
5	ARSENAL	28	17	3	8	54	TIDAK LAYAK JUARA

Pada **Tabel 1** menggunakan data klasemen akhir Liga Premier Inggris musim 2020/2021

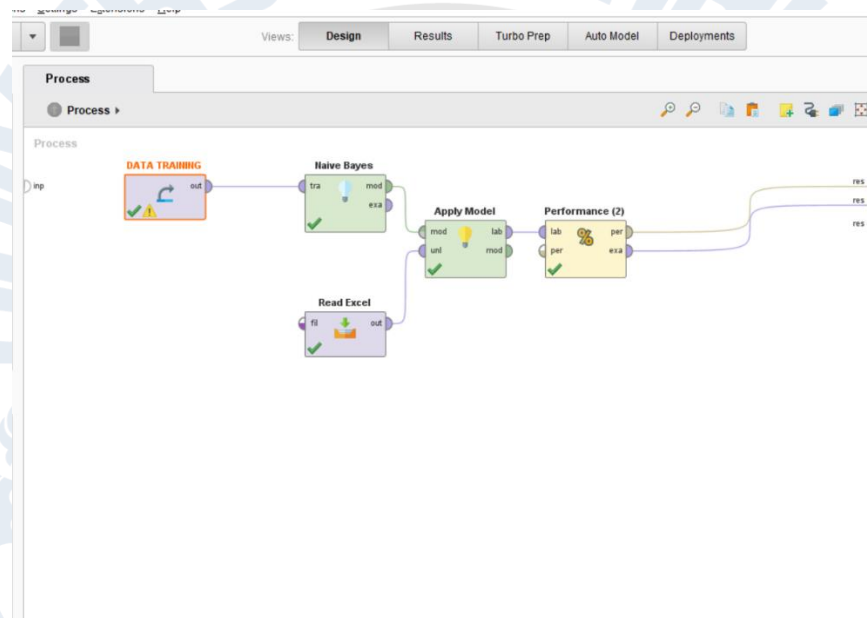
**Tabel 2.** Data Testing

POSISI	TIM	MAIN	MENANG	SERI	KALAH	POIN	PREDIKSI
1	MANCHESTER CITY	38	27	5	6	86	CALON JUARA
2	MANCHESTER UNITED	38	21	11	6	74	CALON JUARA
3	LIVERPOOL	38	20	9	9	69	CALON JUARA
4	CHELSEA	38	19	10	9	67	TIDAK CALON JUARA
5	LEICESTER	38	20	6	12	66	TIDAK CALON JUARA

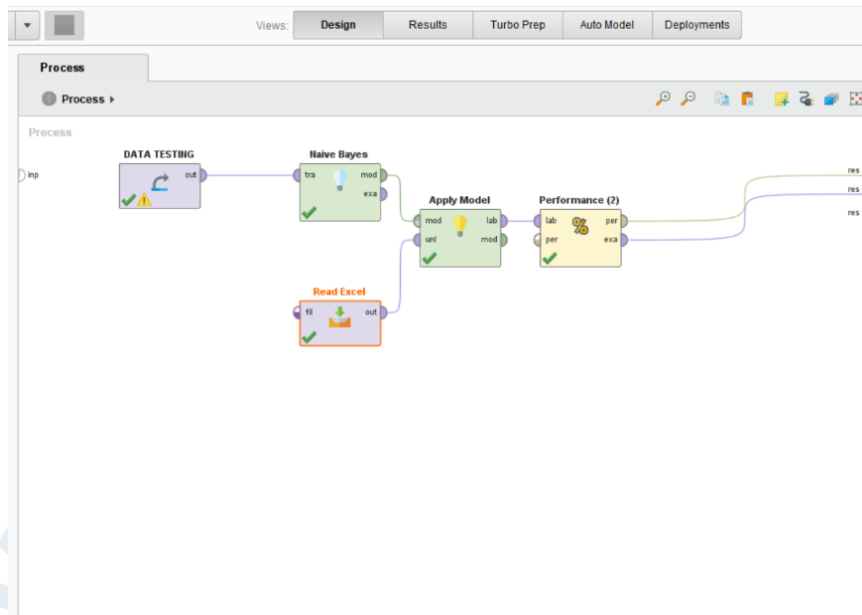
Pada **Tabel 2** menggunakan data klasemen Liga Premier Inggris musim 2021/2022 tanggal 4 April 2022

## A. Implementasi software rapid miner

Kegunaan software ini mengimpor informasi yang terdapat dari berbagai macam basis data untuk dianalisa di dalam sebuah *software*. Rapid Miner adalah sebagai pilihan untuk memprediksi dan analisa dari suatu pekerjaan komputer [11].



Gambar 1 Proses Rapid Miner



Gambar 2 Proses Rapid Miner

Keterangan Gambar 1 dan Gambar 2 :

1. Membutuhkan *read excel* untuk memasukkan data pelatihan dan data tes [12].
2. Membutuhkan *naïve bayes* untuk berfungsi sebagai metode penelitian [13].
3. Membutuhkan *retrieve* ( *data training* dan *data testing*) untuk memuat data ke dalam proses [14].
4. Membutuhkan *apply model* buat uji data pelatihan dan data tes yang perlu ditentukan hasil prediksi itu [15].
5. Membutuhkan *performance* untuk menentukan jenis tugas pembelajaran dan perhitungan paling umum untuk jenis itu [16].



Penentuan tim yang menjadi juara Liga Premier Inggris musim 2021/2022 dengan menggunakan rumus *Naïve Bayes* :

**Tabel 1**  
Tabel Head to Head Manchester City vs Liverpool

Jadwal	Musim (M)	Tempat (T)	Hasil (H)
X	SUMMER	AWAY	?
Y	SUMMER	HOME	?

Pada **Tabel 1** menggunakan data klasemen akhir Liga Premier Inggris musim 2020/2021

**Tabel 2**  
Keterangan atribut yang dihitung

Musim (M)	Tempat (T)	Hasil (H)
3 Oktober 2021	AWAY	DRAW (2-2)
7 Februari 2021	AWAY	WIN (1-4)
8 November 2020	HOME	DRAW (1-1)
3 Juli 2020	HOME	WIN (4-0)
10 November 2019	AWAY	LOSE (3-1)
4 Agustus 2019	AWAY	DRAW (1-1)
4 Januari 2019	HOME	WIN (2-1)
7 Oktober 2018	AWAY	DRAW (0-0)
26 Juli 2018	HOME	LOSE (1-2)
11 April 2018	HOME	LOSE (1-2)

Pada **Tabel 2** menggunakan data klasemen Liga Premier Inggris musim 2021/2022 tanggal 4 April 2022

**Tabel 3**  
 Kekuatan tim dengan rata-rata nilai dari game FIFA 22

Nama Tim	Attack	Defence	Midfield	Rata-Rata
Manchester City	84	86	87	85,5
Liverpool	86	85	84	85

Tabel 3 adalah Data kekuatan tim dari game FIFA 22

- a. Menghitung setiap peluang juara dari kedua tim di musim sebelumnya.

$$\begin{aligned}
 &P(\text{WIN}|\text{X}) \\
 &= P(\text{X}|\text{WIN}) * P(\text{H}=\text{WIN}) \\
 &= P(\text{M}=\text{SUMMER}|\text{H}=\text{WIN}) * P(\text{T}=\text{HOME}|\text{H}=\text{WIN}) * \\
 &P(\text{H}=\text{WIN}|\text{H}=\text{WIN}) * P(\text{H}=\text{WIN}) \\
 &= 4/10 * 6/10 * 7/10 * 9/10 \\
 &= 0,1512
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{DRAW}|\text{X}) \\
 &= P(\text{X}|\text{DRAW}) * P(\text{H}=\text{DRAW}) \\
 &= P(\text{M}=\text{SUMMER}|\text{H}=\text{DRAW}) * P(\text{T}=\text{HOME}|\text{H}=\text{DRAW}) * \\
 &P(\text{H}=\text{DRAW}|\text{H}=\text{DRAW}) * P(\text{H}=\text{DRAW}) \\
 &= 4/10 * 1/4 * 4/8 * 2/10 \\
 &= 0,01
 \end{aligned}$$

$$\begin{aligned}
 &P(\text{LOSE}|\text{X}) \\
 &= P(\text{X}|\text{LOSE}) * P(\text{H}=\text{LOSE}) \\
 &= P(\text{M}=\text{SUMMER}|\text{H}=\text{LOSE}) * P(\text{T}=\text{HOME}|\text{H}=\text{LOSE}) * \\
 &P(\text{H}=\text{LOSE}|\text{H}=\text{LOSE}) * P(\text{H}=\text{LOSE}) \\
 &= 3/5 * 1/2 * 4/3 * 2/10 \\
 &= 0,00375
 \end{aligned}$$



- b. Menghitung setiap peluang juara dari kedua tim di musim ini.

$$\begin{aligned} P(\text{WIN}|Y) &= P(Y|\text{WIN}) * P(H=\text{WIN}) \\ &= P(M=\text{SUMMER}|H=\text{WIN}) * P(T=\text{HOME}|H=\text{WIN}) * \\ &\quad P(H=\text{WIN}|H=\text{WIN}) * P(H=\text{WIN}) \\ &= 2/10 * 8/10 * 9/10 * 9/10 \\ &= 0,1296 \end{aligned}$$

$$\begin{aligned} P(\text{DRAW}|Y) &= P(Y|\text{DRAW}) * P(H=\text{DRAW}) \\ &= P(M=\text{SUMMER}|H=\text{DRAW}) * P(T=\text{HOME}|H=\text{DRAW}) * \\ &\quad P(H=\text{DRAW}|H=\text{DRAW}) * P(H=\text{DRAW}) \\ &= 2/4 * 1/2 * 4/2 * 2/10 \\ &= 0,1 \end{aligned}$$

$$\begin{aligned} P(\text{LOSE}|Y) &= P(Y|\text{LOSE}) * P(H=\text{LOSE}) \\ &= P(M=\text{SUMMER}|H=\text{LOSE}) * P(T=\text{HOME}|H=\text{LOSE}) * \\ &\quad P(H=\text{LOSE}|H=\text{LOSE}) * P(H=\text{LOSE}) \\ &= 1/3 * 1/2 * 1/2 * 2/10 \\ &= 0,01667 \end{aligned}$$

Jadi untuk prediksi jadwal yang datang yaitu :

- Presentasi prediksi WIN =  $0,1296 / (0,1296 + 0,1 + 0,01667) * 100\% = 32,7\%$
- Presentasi prediksi DRAW =  $0,1 / (0,1296 + 0,1 + 0,01667) * 100\% = 25,23\%$
- Presentasi prediksi LOSE =  $0,01667 / (0,1296 + 0,1 + 0,01667) * 100\% = 42,06\%$

Perbedaan pada penelitian lain adalah mereka menggunakan data dan atribut masing-masing untuk diperhitungkan menjadi akurasi atau nilai.

Dari hasil itu dapat dikatakan bahwa prediksi dengan menggunakan metode *Naïve Bayes* untuk menentukan tim yang juara Liga Premier Inggris musim 2021/2022 menghasilkan akurasi sebesar 32,7 % untuk peluang Manchester City dan 42,06 % untuk peluang Liverpool.

Jika mengacu pada perhitungan game FIFA 22, Manchester City unggul dari Liverpool dari segi kekuatan tim.

Dan tim yang layak juara tersebut ialah Manchester City dan Liverpool.

Sedangkan tim yang kemungkinan besar tidak juara ialah Chelsea, Spurs dan Arsenal.

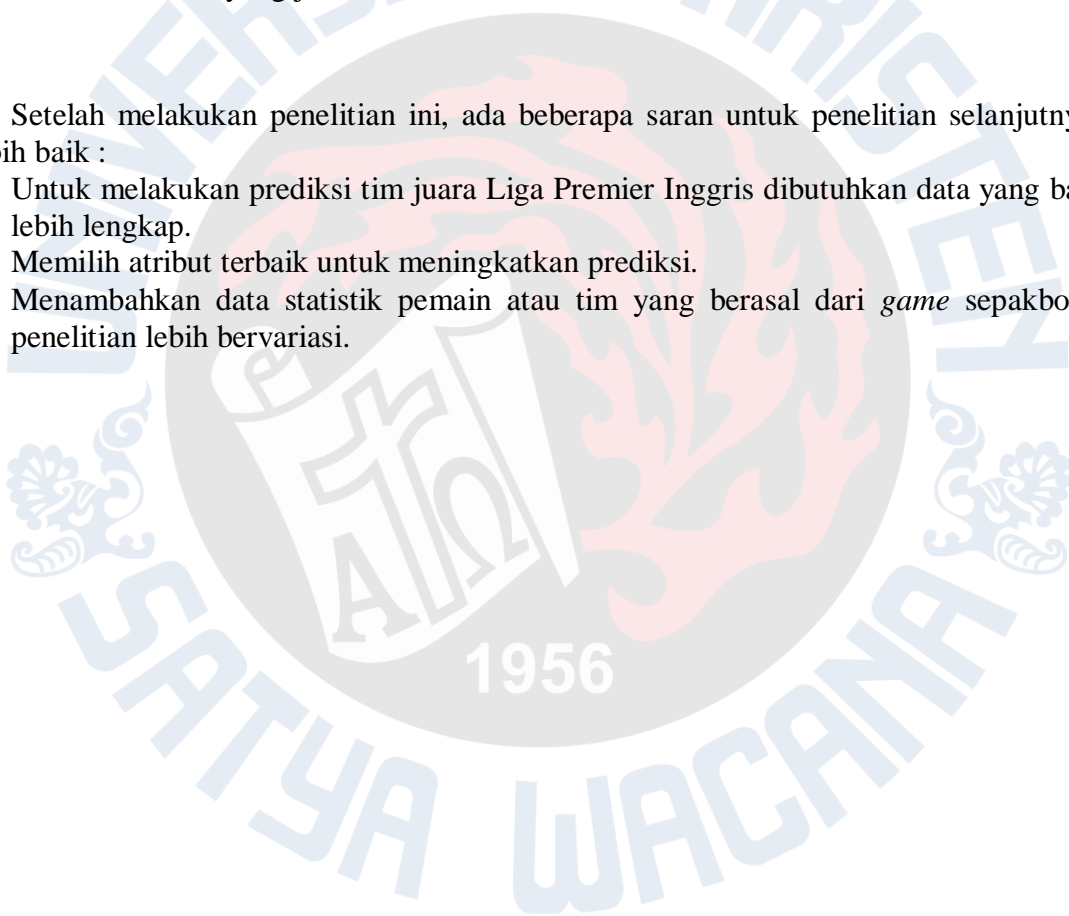
## 5. Simpulan dan Saran

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi tim yang juara Liga Premier Inggris musim 2021/2022 menggunakan metode *Naïve Bayes* dapat disimpulkan :

1. Pada penelitian yang dilakukan, dihasilkan dari prediksi untuk menentukan tim yang juara dengan cepat dan tepat dan dari hasil uji itu diperlukan proses mengolah data pelatihan dan data tes didapatkan akurasi 32,7% untuk peluang Manchester City bisa juara dan Liverpool dengan akurasi 42,06% untuk bisa juara Liga Premier Inggris musim 2021/2022. Dengan hasil tersebut bisa disimpulkan bahwa Liverpool juaranya.
2. Pengolahan data metode menggunakan data pelatihan dan data tes untuk menghasilkan peluang untuk kelas yang beda, bisa dikatakan nilai itu dapat dihasilkan untuk menentukan tim yang juara.

Setelah melakukan penelitian ini, ada beberapa saran untuk penelitian selanjutnya agar lebih baik :

1. Untuk melakukan prediksi tim juara Liga Premier Inggris dibutuhkan data yang baru dan lebih lengkap.
2. Memilih atribut terbaik untuk meningkatkan prediksi.
3. Menambahkan data statistik pemain atau tim yang berasal dari *game* sepakbola agar penelitian lebih bervariasi.



## 6. Daftar Pustaka

- [1] M. Sabransyah, Y. N. Nasution, F. D. T. Amijaya. 2017. *Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung*. Samarinda:Universitas Mulawarman
- [2] I. N. Farida, Y. S. Lestyaningtyas. 2015. *Implementasi Fuzzy Clustering untuk Prediksi Pemilihan Ketua OSIS*. Kediri:Universitas Nusantara PGRI Kediri
- [3] A. F. Efediy. 2013. *Tingkat Stress Menjelang Perlombaan Terhadap Hasil Perlombaan Pada Atlet Atletik Kejuaraan Nasional Yuniior Dan Remaja Jawa Tengah Di Jakarta Tahun 2013*. Semarang:Universitas Negeri Semarang
- [4] A. Saleh. 2015. *Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga*. Medan:Universitas Potensi Utama
- [5] M. Salah. 2007. *Penggunaan data mining dengan metode naïve bayessian classification untuk memprediksi juara liga inggris musim 2006/2007*. Yogyakarta:Universitas Sanata Dharma
- [6] D. Prabowo. 2020. *Prediksi hasil pertandingan sepakbola english premier league dengan menggunakan algoritma k-nearest neighbors dan naive bayes classifier*. Yogyakarta:Universitas Islam Indonesia
- [7] F. F. Pratama, Y. I. Nurhasanah. 2020. *Penggunaan Metode Profile Matching Dan Naïve Bayes Untuk Menentukan Starting Eleven Pada Sepak Bola*. Bandung:Institut Teknologi Nasional Bandung
- [8] D. Sartika, D. Andreswari, K. Anggriani. 2016. *Penentuan Posisi Ideal Pemain Dalam Cabang Olahraga Sepak Bola Dengan Menggunakan Pendekatan Dua Metode Naïve Bayes & Profile Matching*. Bengkulu:Universitas Bengkulu
- [9] M. A. Yaqin, M. Z. Ramadhan, A. F. Jauhari, A. G. Humami. 2019. *Optimasi Pemilihan Posisi Terbaik Pemain Muda Pada Game Football Manager 2018 Dengan Metode Naïve Bayes*. Malang:Universitas Islam Negeri Maulana Malik Ibrahim Malang
- [10] T. I. Andini, W. Witanti, and F. Renaldi. 2016. *Prediksi Potensi Pemasaran Produk Baru dengan Metode Naïve Bayes Classifier dan Regresi Linear*. Cimahi:Universitas Jenderal Achmad Yani
- [11] H. F. Putro, R. T. Vulandari, and W. L. Saptomo. 2020. *Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan*. Surakarta:STMIK Sinar Nusantara
- [12] M. H. Rifqo, A. Wijaya, and J. Pseudocode. 2017. *Implementasi Algoritma Naive Bayes Dalam Penentuan Pemberian Kredit*. Bengkulu:Universitas Muhammadiyah Bengkulu
- [13] S. J. S. Tyas. 2021. *Analisis Perbandingan Algoritma Naïve Bayes dan C.45 dalam Klasifikasi Data Mining untuk memprediksi kelulusan*. Bandung:Universitas Pendidikan Indonesia
- [14] R. Wijayatun and Y. Sulisty. 2016. *Prediksi Rating Film Menggunakan Metode Naïve Bayes*. Surakarta:Universitas Muhammadiyah Surakarta
- [15] Bustami. 2014. *Penerapan Algoritma Naive Bayes untuk mengklasifikasi Data Nasabah Asuransi*. Aceh:Universitas Malikussaleh
- [16] D. Kurniawaty, S. K. Sari, I. Cholissodin, Maryamah, Asikin M. F. 2016. *Implementasi Metode Naïve Bayes Clasffier untuk seleksi Asisten Praktikum pada simulasi Hadoop Multinode Cluster*. Malang:Universitas Brawijaya