

PENERAPAN ALGORITME *NAÏVE BAYES* DALAM MEMPREDIKSI JUARA LIGA PRIMER INGGRIS MUSIM 2022/2023

Guzti Eka Putra^{1*}, Titin Fatimah²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}guztiekaputra@gmail.com, ²titin.fatimah@budiluhur.ac.id
(* : corresponding author)

Abstrak- Liga Primer Inggris adalah kompetisi sepak bola yang digadang-gadang sebagai kompetisi yang paling sulit untuk diprediksi juaranya setiap tahun. Setiap musim, para pendukung klub dan pengamat sepakbola berdebat dan memprediksi klub mana yang akan keluar sebagai juara Liga Primer Inggris. Permasalahan yang muncul adalah prediksi yang dilakukan secara manual dan hanya berdasarkan pengalaman terkadang tidak dapat memberikan hasil yang akurat. Penelitian ini dilakukan dengan harapan dapat memberikan kontribusi pada bidang *data mining* dan pengolahan data, khususnya dalam penerapan algoritme *Naïve Bayes* dalam memprediksi klub juara Liga Primer Inggris. Penelitian ini bertujuan untuk memprediksi dengan tepat klub mana yang akan menjadi juara Liga Primer Inggris musim 2022/2023 dengan mengimplementasikan algoritme *Naïve Bayes* menggunakan atribut yang diambil dari data statistik historis 10 besar klub selama 20 tahun terakhir. Pengujian dilakukan dengan membagi data latihan dan data uji dengan perbandingan 7:3 dan menghasilkan nilai akurasi sebesar 91,7%, presisi 84,2%, dan *recall* sebesar 88,9% dari model latihan tersebut. Sementara hasil prediksi dari data yang diuji pada minggu ke-26 kompetisi Liga Primer Inggris menunjukkan bahwa hanya ada 2 klub yang berpeluang menjadi calon juara pada kompetisi musim 2022/2023. Arsenal yang berada di puncak klasemen hanya memiliki 6,67% peluang sedangkan klub Manchester City yang berada di peringkat dua memiliki peluang sebesar 93,33% untuk bisa menjadi juara Liga Primer Inggris musim 2022/2023.

Kata Kunci: Prediksi, Klub Juara, Liga Primer Inggris, *Data Mining*, *Naïve Bayes*.

IMPLEMENTATION OF *NAÏVE BAYES* ALGORITHM IN PREDICTING THE CHAMPION OF ENGLISH PREMIER LEAGUE 2022/2023 SEASON

Abstract- *The English Premier League is a football competition that is predicted to be the most difficult competition to predict the winner every year. Every season, club supporters and football observers debate and predict which club will come out as Premier League champions. The problem that arises is that predictions are made manually and only based on experience sometimes cannot provide accurate results. This research was conducted with the hope of contributing to the field of data mining and data processing, especially in the application of the Naïve Bayes algorithm in predicting the English Premier League champions. This study aims to predict exactly which club will win the English Premier League in the 2022/2023 season by implementing the Naïve Bayes algorithm using attributes taken from historical statistical data of the top 10 clubs for the last 20 years. The test was carried out by dividing the training data and test data with a ratio of 7:3 and producing an accuracy value of 91.7%, 84.2% precision and 88.9% recall of the training model. While the prediction results from the data tested in the 26th week of the English Premier League show that there are only 2 clubs that have the chance to become champions in the 2022/2023 season competition. Arsenal, who are at the top of the standings, only have a 6.67% chance, while Manchester City, who are ranked second, have a 93.33% chance of winning the 2022/2023 Premier League season.*

Keywords: Prediction, Club Champion, English Premier League, *Data Mining*, *Naïve Bayes*.

1. PENDAHULUAN

Liga Primer Inggris atau biasa dikenal dengan nama *English Premier League* adalah salah satu kompetisi sepakbola yang paling banyak penggemarnya. Setiap musim, para pendukung klub dan pengamat sepakbola berdebat dan memprediksi klub mana yang akan keluar sebagai juara Liga Primer Inggris. Permasalahan yang muncul adalah prediksi yang dilakukan secara manual dan hanya berdasarkan pengalaman terkadang tidak dapat memberikan hasil yang akurat. Maka dari itu, rumusan masalah dalam penelitian ini adalah bagaimana cara yang tepat untuk memprediksi klub yang akan menjadi juara Liga Primer Inggris di musim 2022/2023.

Dengan menggunakan data historis Liga Primer Inggris, dapat dilakukan analisis untuk menentukan faktor-faktor apa saja yang mempengaruhi tim untuk menjadi juara pada musim yang akan datang. Implementasi *data mining* dalam memprediksi klub juara Liga Primer Inggris dapat memberikan manfaat dalam mengoptimalkan strategi tim sepakbola dan membantu para pengamat sepakbola dalam membuat prediksi yang lebih akurat. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan *data mining* dengan menggunakan algoritme *Naïve Bayes* dalam memprediksi klub juara Liga Primer Inggris di musim 2022/2023.

Penelitian tentang prediksi juara Liga Primer Inggris menggunakan *data mining* dengan algoritme *Naïve Bayes* sendiri sebelumnya sudah pernah dilakukan oleh [1]. Perbedaan penelitian ini dari penelitian sebelumnya adalah dari atribut yang digunakan untuk penambangan datanya. Hasil dari jumlah kemenangan, jumlah kekalahan, hasil seri, jumlah gol, kebobolan, selisih gol dan poin menjadi atribut yang digunakan, sementara penelitian sebelumnya menggunakan atribut *overall* sebuah tim yang berasal dari *game* FIFA 22.

Algoritme *Naïve Bayes* dipilih pada penelitian ini karena berdasarkan penelitian yang dilakukan oleh [2], algoritme *Naïve Bayes* terbukti menjadi metode klasifikasi terbaik untuk memprediksi hasil pertandingan sepakbola jika dibandingkan dengan metode klasifikasi lainnya.

Klasifikasi merupakan salah satu poin utama dalam pengolahan *data mining*. Klasifikasi merupakan proses yang terdiri dari dua tahap, yaitu tahap pembelajaran dan tahap pengklasifikasian [3]. Klasifikasi juga merupakan pengelompokan data dimana data yang digunakan mempunyai kelas atau target yang akan diklasifikasi [4].

Prediksi diartikan sebagai ramalan; prakiraan. Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil [5].

Machine learning adalah cabang ilmu dari kecerdasan buatan (AI) yang berfokus pada algoritme dan pengembangan teknik yang memungkinkan sistem komputer untuk beradaptasi dan belajar secara otomatis. *Machine learning* adalah mesin pembelajaran yang sangat membantu dalam menyelesaikan masalah [6].

Data mining adalah proses mencari informasi menarik atau sebuah pola unik dalam data terpilih dengan menggunakan metode atau teknik tertentu [7]. *Data mining* berisi pencarian *trend* atau pola yang diinginkan dalam database besar untuk membantu pengambil keputusan di waktu yang akan datang [8]. *Data mining* biasa juga disebut dengan “*knowledge discovery*” atau menemukan pola tersembunyi pada data. *Data mining* adalah proses dari menganalisa data dari prespektif yang berbeda dan menyimpulkannya ke dalam informasi yang berguna [9]. *Data mining* diartikan sebagai menambang data atau upaya untuk menggali informasi yang berharga dan berguna pada database yang sangat besar [10]. *Data Mining* adalah sebuah proses penemuan pola, hubungan, dan informasi yang berharga dari sekumpulan data besar dan kompleks. Tujuan utama dari *data mining* adalah untuk menggali pengetahuan yang tersembunyi dalam data untuk mendukung pengambilan keputusan yang lebih baik. Tahapan dalam proses penggalian pengetahuan tersebut diantaranya: *data cleaning*, *data integration*, *data selection*, *data transformation*, *data mining*, *pattern evaluation*, dan *knowledge presentation*. Empat tahapan pertama disebut sebagai *preprocessing data*, dimana data terlebih dahulu disiapkan sebelum akhirnya dilakukan proses penambangannya [11].

Metode *Naïve Bayes* adalah algoritme klasifikasi yang populer dalam *data mining*. *Naïve Bayes* adalah salah satu metode pengklasifikasian sederhana yang sering digunakan karena mudah diterapkan dan memiliki hasil yang baik saat diterapkan pada banyak kasus [12]. Algoritme *Naïve Bayes* memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa lalu. Metode *Naïve Bayes* menerapkan teknik *supervised* klasifikasi objek di masa depan dengan menetapkan *label* kelas ke instance dengan menggunakan probabilitas bersyarat. Probabilitas bersyarat adalah ukuran peluang suatu peristiwa yang terjadi berdasarkan peristiwa lain yang telah terjadi. Sementara istilah *supervised* sendiri merujuk pada klasifikasi data training yang sudah diberi *label* dengan kelas [13]. Dasar dari teorema *Naïve Bayes* menggunakan rumus dari persamaan (1).

$$P(A/B) = (P(B/A) * P(A))/P(B) \quad (1)$$

Keterangan:

$P(A/B)$: probabilitas A terjadi dengan bukti bahwa B telah terjadi

$P(B/A)$: probabilitas B terjadi dengan bukti bahwa A telah terjadi

$P(A)$: peluang terjadinya A

$P(B)$: peluang terjadinya B

Akurasi menggambarkan sejauh mana model dapat melakukan prediksi yang benar terhadap data yang diamati. Akurasi dapat didefinisikan sebagai tingkat kedekatan antara nilai aktual dan nilai prediksi [14]. Presisi dan *Recall* adalah metrik evaluasi yang penting dalam klasifikasi, dan termasuk dalam konteks metode *Naïve Bayes*. Presisi menunjukkan tingkat ketelitian atau ketepatan dalam pengklasifikasian. Sedangkan *recall* berfungsi untuk mengukur proporsi positif aktual yang benar diidentifikasi. Untuk mengukur akurasi, presisi, dan *recall* biasanya digunakan *confusion matrix* [14]. Akurasi adalah metrik evaluasi yang digunakan untuk mengukur kinerja model klasifikasi, termasuk dalam konteks metode *Naïve Bayes*. Sementara presisi dan *recall* sangat berguna ketika kelas yang diidentifikasi memiliki perbedaan signifikan dalam dampaknya.

Akurasi merupakan tingkat kedekatan antara nilai prediksi dan nilai sebenarnya. Akurasi dapat dihitung dengan persamaan (2).

$$Akurasi = \frac{(TP+TN)}{(TP + TN + FP + FN)} \quad (2)$$

Presisi merupakan tingkat ketepatan antara data yang diminta dan jawaban yang diberikan oleh sistem persamaan. Presisi dapat dihitung dengan persamaan (3).

$$Presisi = \frac{TP}{(TP + FP)} \quad (3)$$

Recall merupakan pengukur tingkat keberhasilan suatu sistem dalam menemukan kembali sebuah informasi. Recall dapat dihitung dengan persamaan (4).

$$Presisi = \frac{TP}{(TP + FN)} \quad (4)$$

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Pada pengukuran kinerjanya, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Nilai *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif [15]. *Confusion Matrix* adalah sebuah metode evaluasi yang umum digunakan dalam klasifikasi dan prediksi dalam bentuk tabel. Tabel ini menggambarkan hasil prediksi dari model dengan membandingkan kelas prediksi dengan kelas sebenarnya pada data uji. *Confusion Matrix* terdiri dari empat matrik evaluasi utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Bentuk sebuah *Confusion Matrix* dapat dilihat pada Gambar 1.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	TP	FP
	Negative	FN	TN

Gambar 1. *Confusion Matrix*

2. METODE PENELITIAN

2.1 Data Penelitian

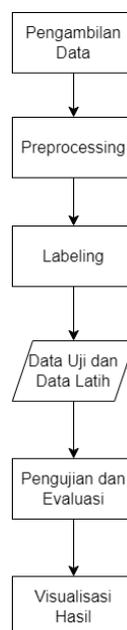
Dataset yang digunakan dalam penelitian ini berasal dari Kaggle, sebuah situs *Data Science* yang menyediakan banyak kumpulan data yang kompleks dan penting. Untuk data 10 besar klub Liga Primer Inggris dalam 20 tahun ke belakang sendiri didapat dari url: <https://www.kaggle.com/datasets/kozylama/english-premier-league-team-statistics-20002022>. Data awal berupa *file* .csv yang berisi atribut-atribut yang diperlukan untuk penelitian, beberapa contoh datanya dapat dilihat pada Gambar 2.

2009,1,Manchester Utd,38,28,6,4,68,24,44,90	
2009,2,Liverpool,38,25,11,2,77,27,50,86	
2009,3,Chelsea,38,25,8,5,68,24,44,83	
2009,4,Arsenal,38,20,12,6,68,37,31,72	

Gambar 2. Contoh *Dataset Awal*

2.2 Penerapan Metode Klasifikasi *Naïve Bayes*

Dalam mengimplementasikan *data mining* dengan algoritme *Naïve Bayes* pada penelitian ini, dilakukan beberapa langkah. Langkah-langkah tersebut mewakili setiap proses penelitian yang dilakukan dari awal sampai akhir aplikasi bekerja. Gambar 3 memperlihatkan setiap tahapan penerapan metode yang digunakan.



Gambar 3. Tahapan Metode Klasifikasi *Naïve Bayes*

2.2.1 *Preprocessing*

Tahap *preprocessing* merupakan tahapan yang paling penting saat *dataset* sudah tersedia. Tahapan *preprocessing* pada penelitian ini terdiri atas beberapa proses yaitu *text-to-column*, *data selection* dan *data transformation*.

a. *Text-to-column*

Dataset dengan format .csv bukan merupakan sebuah tabel, untuk itu perlu diubah terlebih dahulu ke dalam bentuk tabel sehingga data yang disajikan nantinya akan lebih mudah dibaca dan dimengerti. Dengan fitur *text-to-column* pada aplikasi Excel 2016, data dengan format .csv dapat dengan mudah diubah menjadi bentuk tabel.

b. *Data Selection*

Pada tahapan ini, data yang telah diubah menjadi bentuk tabel melalui proses *text-to-column* dipilih atribut dan data mana saja yang akan diambil untuk nantinya menjadi data latih. Atribut yang dipakai pada penelitian ini adalah Jumlah Kemenangan, Hasil Seri, Jumlah Kekalahan, Gol, Kebobolan, Selisih Gol dan Poin.

c. *Data Transformation*

Setelah atribut dipilih, nilai yang ada pada setiap atribut dikelompokkan menjadi hanya dua sampai tiga kelas saja. Proses transformasi ini dilakukan agar nantinya proses penambangan data yang berjalan dapat dilakukan dengan lebih mudah dan dapat memberikan hasil akurasi yang baik pada saat klasifikasi *Naïve Bayes* dilakukan.

2.2.2 Labeling

Tahapan selanjutnya adalah *labeling* yakni pemberian *label* pada *dataset* yang sudah melalui proses sebelumnya sehingga didapatkan sebuah atribut baru yang akan menjadi penentu atau *output* dari penambahan data yang dilakukan. *Label* yang akan diberikan pada data yang sudah siap adalah *label* Status yang akan mempunyai nilai ‘Calon Juara’ dan ‘Bukan Calon Juara’. Pemberian *label* ini didasarkan pada atribut peringkat per musimnya. Gambar 4 memperlihatkan hasil akhir dari data yang sudah siap untuk diproses ke tahap selanjutnya.

	JumlahKemenangan	JumlahHasilSeri	JumlahKekalahan	JumlahGol	JumlahKebobolan	SelisihGol	Poin	Status
2	> 20	< 10	< 10	> 60	30-50	>30	>75	Calon Juara
3	> 20	< 10	< 10	> 60	30-50	>30	>75	Calon Juara
4	> 20	< 10	>= 10	> 60	30-50	10-30	<=75	Calon Juara
5	<= 20	>= 10	< 10	> 60	30-50	10-30	<=75	Bukan Calon Juara
6	<= 20	>= 10	>= 10	> 60	30-50	10-30	<=75	Bukan Calon Juara
7	<= 20	>= 10	>= 10	40-60	30-50	<10	<=75	Bukan Calon Juara
8	<= 20	< 10	>= 10	40-60	30-50	<10	<=75	Bukan Calon Juara
9	<= 20	>= 10	>= 10	40-60	30-50	<10	<=75	Bukan Calon Juara
10	<= 20	< 10	>= 10	40-60	>50	<10	<=75	Bukan Calon Juara

Gambar 4. Hasil Akhir Data

2.2.3 Pembagian Data Latih dan Data Uji

Setelah tahap *labeling*, data telah siap untuk dilakukan pembagian. Data akan dibagi menjadi dua bagian yaitu data latih dan data uji. Pembagian data dilakukan dengan perbandingan 7:3 atau 70% data latih dan 30% data uji. Dalam penelitian ini, total data yang ada berjumlah 200 data, maka dengan rasio 7:3, yang akan menjadi data latih adalah sebanyak 140 data sedangkan 60 data lainnya akan menjadi data uji.

2.2.4 Klasifikasi Naïve Bayes

Langkah pertama yang dilakukan dalam klasifikasi *Naïve Bayes* adalah menghitung probabilitas kemunculan setiap atribut dan *label* menggunakan persamaan (1) pada data latih. Tabel 1 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Jumlah Kemenangan. Probabilitas didapatkan dengan membandingkan nilai setiap jumlah kemenangan *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 1. Probabilitas Kemunculan Setiap Nilai untuk Atribut Jumlah Kemenangan

C1	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara
> 20	39	7	13/14	7/98
<= 20	3	91	1/14	91/98
Jumlah	42	98	1	1

Tabel 2 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Hasil Seri. Probabilitas didapatkan dengan membandingkan nilai setiap hasil seri *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 2. Probabilitas Kemunculan Setiap Nilai untuk Atribut Hasil Seri

C2	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara
< 10	32	42	16/21	6/14
>= 10	10	56	5/21	8/14
Jumlah	42	98	1	1

Tabel 3 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Jumlah Kekalahan. Probabilitas didapatkan dengan membandingkan nilai setiap kekalahan *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 3. Probabilitas Kemunculan Setiap Nilai untuk Atribut Jumlah Kekalahan

C3	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara
< 10	39	21	13/14	3/14
>= 10	3	77	1/14	11/14
Jumlah	42	98	1	1

Tabel 4 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Gol. Probabilitas didapatkan dengan membandingkan nilai setiap gol *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 4. Probabilitas Kemunculan Setiap Nilai untuk Atribut Gol

C4	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara
> 60	38	23	19/21	23/98
40-60	4	73	2/21	73/98
< 40	0	2	0	1/49
Jumlah	42	98	1	1

Tabel 5 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Kebobolan. Probabilitas didapatkan dengan membandingkan nilai setiap kebobolan *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 5. Probabilitas Kemunculan Setiap Nilai untuk Atribut Kebobolan

C5	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara
< 30	16	1	8/21	1/98
30-50	26	71	13/21	71/98
> 50	0	26	0	13/49
Jumlah	42	98	1	1

Tabel 6 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Selisih Gol. Probabilitas didapatkan dengan membandingkan nilai setiap selisih gol *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 6. Probabilitas Kemunculan Setiap Nilai untuk Atribut Selisih Gol

C6	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara

< 30	36	5	6/7	5/98
30-50	6	34	1/7	17/49
> 50	0	59	0	59/98
Jumlah	42	98	1	1

Tabel 7 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Poin. Probabilitas didapatkan dengan membandingkan nilai setiap jumlah poin *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 7. Probabilitas Kemunculan Setiap Nilai untuk Atribut Poin

C7	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara
> 75	31	3	31/42	3/98
<= 75	11	95	11/42	95/98
Jumlah	42	98	1	1

Tabel 8 menunjukkan probabilitas kemunculan setiap nilai untuk atribut Status. Probabilitas didapatkan dengan membandingkan nilai setiap jumlah *label* positif dan *label* negatif dengan jumlah total semua *label*.

Tabel 8. Probabilitas Kemunculan Setiap Nilai untuk Atribut Status

C8	Label		Probabilitas	
	Calon Juara	Bukan Calon Juara	Calon Juara	Bukan Calon Juara
Jumlah	42	98	3/10	7/10

3. HASIL DAN PEMBAHASAN

3.1 Pengujian

Pada penelitian ini, pengujian akan dilakukan dengan mencari nilai akurasi, presisi dan *recall* dari implementasi algoritme *Naïve Bayes* dalam memprediksi *label* untuk data uji yang sebelumnya telah disiapkan. Selain pada sisi akurasi, presisi dan *recall*, tahap pengujian ini juga akan melakukan prediksi untuk juara Liga Primer Inggris musim 2022/2023 yang pada saat penelitian ini dilakukan, yaitu minggu ke-26 di musim 2022/2023, klub Arsenal, Manchester City, Manchester United dan Newcastle United sedang berada di peringkat 4 besar. Data dari 4 klub inilah yang akan menjadi data uji dalam memprediksi klub juara Liga Primer Inggris di musim 2022/2023.

3.1.1 Akurasi Data Uji

Sebelum dilakukan perhitungan dengan klasifikasi, Tabel 9 menyajikan ringkasan 60 baris data uji yang akan dilakukan pengujian.

Tabel 9. Data Uji

No.	C1	C2	C3	C4	C5	C6	C7	C8
1.	> 20	< 10	< 10	> 60	30-50	> 30	> 75	Calon Juara
2.	> 20	< 10	< 10	> 60	< 30	> 30	> 75	Calon Juara
3.	> 20	< 10	< 10	> 60	30-50	> 30	> 75	Calon Juara

...
60.	<= 20	< 10	>= 10	< 40	30-50	< 10	<= 75	Bukan Calon Juara	

Dari data uji tersebut, gunakan rumus persamaan klasifikasi *Naïve Bayes* untuk mendapatkan *Label* Prediksi dari masing-masing baris data. Probabilitas kemunculan setiap nilai untuk masing-masing atribut mengacu pada data latih dan tabel yang sudah dihitung sebelumnya. Berikut ini adalah perhitungan klasifikasi untuk data uji nomor 1:

$$\begin{aligned}
 \text{Calon Juara} &= P(\text{Calon Juara} \mid \text{JumlahKemenangan} = >20) \cdot P(\text{Calon Juara} \mid \text{JumlahHasilSeri} = <10) \\
 &\cdot P(\text{Calon Juara} \mid \text{JumlahKekalahan} = <10) \cdot P(\text{Calon Juara} \mid \text{JumlahGol} = >60) \cdot P(\text{Calon Juara} \mid \text{JumlahKebobolan} = 30-50) \\
 &\cdot P(\text{Calon Juara} \mid \text{SelisihGol} = >30) \cdot P(\text{Calon Juara} \mid \text{Poin} = >75) \cdot P(\text{Calon Juara}) \\
 &= 13/14 \cdot 16/21 \cdot 13/14 \cdot 19/21 \cdot 13/21 \cdot 6/7 \cdot 31/42 \cdot 3/10 \\
 &= 0,06983555
 \end{aligned}$$

$$\begin{aligned}
 \text{Bukan Calon Juara} &= P(\text{Bukan Calon Juara} \mid \text{JumlahKemenangan} = >20) \cdot P(\text{Bukan Calon Juara} \mid \text{JumlahHasilSeri} = <10) \\
 &\cdot P(\text{Bukan Calon Juara} \mid \text{JumlahKekalahan} = <10) \cdot P(\text{Bukan Calon Juara} \mid \text{JumlahGol} = >60) \cdot P(\text{Bukan Calon Juara} \mid \text{JumlahKebobolan} = 30-50) \\
 &\cdot P(\text{Bukan Calon Juara} \mid \text{SelisihGol} = >30) \cdot P(\text{Bukan Calon Juara} \mid \text{Poin} = >75) \cdot P(\text{Bukan Calon Juara}) \\
 &= 1/14 \cdot 3/7 \cdot 3/14 \cdot 23/98 \cdot 71/98 \cdot 5/98 \cdot 3/98 \cdot 7/10 \\
 &= 0,00000122
 \end{aligned}$$

$$P \mid \text{Calon Juara} = \frac{0,06983555}{0,06983555 + 0,00000122} = 0,9999825$$

$$P \mid \text{Bukan Calon Juara} = \frac{0,00000122}{0,06983555 + 0,00000122} = 0,0000175$$

Dari hasil perhitungan menggunakan klasifikasi *Naïve Bayes* yang dilakukan, data uji nomor 1 mendapatkan *Label* Prediksi Calon Juara. Maka, berdasarkan perhitungan sebelumnya, setiap baris data uji dilakukan perhitungan menggunakan klasifikasi *Naïve Bayes* dan hasilnya disajikan pada Tabel 10.

Tabel 10. Hasil Klasifikasi Data Uji

No.	C1	C2	C3	C4	C5	C6	C7	C8	C9
1.	> 20	< 10	< 10	> 60	30-50	> 30	> 75	Calon Juara	Calon Juara
2.	> 20	< 10	< 10	> 60	< 30	> 30	> 75	Calon Juara	Calon Juara
3.	> 20	< 10	< 10	> 60	30-50	> 30	> 75	Calon Juara	Calon Juara
...
60.	<= 20	< 10	>= 10	< 40	30-50	< 10	<= 75	Bukan Calon Juara	Bukan Calon Juara

Keterangan: C9 = *Label* Prediksi

Keseluruhan hasil klasifikasi data uji pada Tabel 10 kemudian direpresentasikan ke dalam bentuk *confusion matrix* untuk dihitung nilai akurasi, presisi dan *recall* nya. Tabel 11 menyajikan representasi *confusion matrix*.

Tabel 11. Confusion Matrix Hasil Klasifikasi Data Uji

Nilai Aktual

	Calon Juara	Bukan Calon Juara
Calon Juara	16	2
Nilai Prediksi		
Bukan Calon Juara	3	39

Maka, berdasarkan *confusion matrix* pada Tabel 11 dan penjabaran tentang akurasi, presisi dan *recall*, dapat dihitung nilai akurasi, presisi dan *recall* yang didapatkan dari hasil pengujian ini yang tersaji pada Tabel 12.

Tabel 12. Hasil Pengujian

Pengujian		
Akurasi	$\frac{(16 + 39)}{(16 + 39 + 3 + 2)}$	91,7%
Presisi	$\frac{16}{(16 + 3)}$	84,2%
Recall	$\frac{16}{(16 + 2)}$	88,9%

3.1.2 Prediksi Juara Liga Primer Inggris Musim 2022/2023

Pengujian berikutnya adalah melakukan prediksi untuk klub juara Liga Primer Inggris musim 2022/2023. Pada saat penelitian dilakukan, kompetisi liga belum selesai dan belum diketahui siapa juaranya. Prediksi untuk menebak siapa yang akan menjadi juara Liga Primer Inggris musim 2022/2023 dilakukan dengan melihat data klasemen 4 besar klub pada minggu ke-26 dengan data uji pada Gambar 5 yang diambil dari tautan <https://www.worldfootball.net/schedule/eng-premier-league-2022-2023-spieltag/26/>.

#	Team	M.	W	D	L	goals	Dif.	Pt.
1	Arsenal FC	26	20	3	3	59:25	34	63
2	Manchester City	26	18	4	4	66:25	41	58
3	Manchester United	26	16	4	6	42:35	7	52
4	Newcastle United	26	12	11	3	44:19	25	47

Gambar 5. Klasemen 4 Besar Liga Primer Inggris Minggu ke-26

Dengan rumus perhitungan klasifikasi *Naïve Bayes* dan menggunakan *model* latih dari *dataset* sebelumnya, maka perhitungan probabilitas masing-masing klub untuk bisa menjadi juara Liga Primer Inggris musim 2022/2023 dapat dihitung seperti pengujian sebelumnya. Hasil perhitungan probabilitasnya tersaji pada Tabel 13.

Tabel 13. Hasil Perhitungan Probabilitas Klub 4 Besar Musim 2022/2023

Klub	Perhitungan Probabilitas Calon Juara	Perhitungan Probabilitas Bukan Calon Juara	Hasil Prediksi
Arsenal	0,00010565 (77,84%)	0,00003006 (22,16%)	Calon Juara
Manchester City	0,00147915 (99,08%)	0,00001372 (0,92%)	Calon Juara
Manchester United	0 (0%)	0,01487998 (100%)	Bukan Calon Juara
Newcastle United	0,00000584 (2,67%)	0,00021289 (97,33%)	Bukan Calon Juara

Pada Tabel 13 dapat disimpulkan bahwa sampai minggu ke-26 kompetisi musim 2022/2023, hanya ada 2 klub yang menjadi Calon Juara liga yaitu Arsenal dan Manchester City. Perhitungan peluang antara kedua klub tersebut

dapat dilakukan dengan menggunakan cara yang sama yakni dengan rumus probabilitas. Perhitungan peluang juara secara persentase dapat dilihat pada Tabel 14.

Tabel 14. Persentase Peluang Juara Liga Inggris Musim 2022/2023

Pengujian		
Arsenal	$0,00010565(0,00010565+0,00147915) \cdot 100\%$	6,67%
Manchester City	$0,00147915(0,00010565+0,00147915) \cdot 100\%$	93,33%
Total		100%

4. KESIMPULAN

Hasil pengujian dan perhitungan dari program penelitian yang dibuat dengan menggunakan *dataset* dan metode yang diusulkan menunjukkan hasil yang menarik. Dalam pengujian dengan 200 baris data yang telah ditransformasi menjadi model latih dengan perbandingan 70% data latih dan 30% data uji, diperoleh akurasi sebesar 91,7%, presisi sebesar 84,2%, dan *recall* sebesar 88,9%. Hal ini menunjukkan bahwa metode klasifikasi dengan algoritme *Naïve Bayes* dapat memberikan prediksi yang cukup akurat dalam memprediksi klub juara Liga Primer Inggris musim 2022/2023. Dari hasil prediksi, ditemukan bahwa peluang juara Liga Primer Inggris musim 2022/2023 akan diperebutkan oleh dua klub, yaitu Arsenal dan Manchester City. Klub lainnya yang juga dihitung peluangnya, seperti Manchester United dan Newcastle United, diprediksi bukan menjadi calon juara. Menariknya, meskipun klub Arsenal saat itu berada di puncak klasemen, prediksinya hanya memiliki peluang sebesar 6,67% untuk menjadi juara. Sebaliknya, klub Manchester City yang saat itu berada di peringkat kedua klasemen memiliki peluang lebih tinggi, yakni sebesar 93,33%, untuk menjadi juara. Pengembangan lebih lanjut di masa depan disarankan untuk menambahkan lebih banyak jumlah baris data dalam melakukan pengujian. Dengan lebih banyak data dan atribut yang bervariasi, hal ini akan memberikan gambaran yang lebih komprehensif dan menghasilkan nilai akurasi yang lebih dapat diandalkan. Proses *preprocessing* data pun perlu diperhatikan dengan lebih baik, terutama pada tahapan *data transformation*. Perubahan bentuk data asli yang berlebihan dapat mengurangi kualitas prediksi, oleh karena itu, perlu dilakukan dengan hati-hati. Pengujian menggunakan metode klasifikasi lainnya juga bisa menjadi langkah yang menarik. Hal ini dapat membuka peluang untuk mengeksplorasi metode lain yang mungkin dapat memberikan nilai akurasi yang lebih baik dibandingkan dengan algoritme *Naïve Bayes* yang telah digunakan pada penelitian ini. Di masa depan, diharapkan penelitian sejenis dapat menjadi lebih kuat dan relevan dalam memprediksi klub juara Liga Primer Inggris di musim-musim mendatang. Pengetahuan yang diperoleh dari penelitian ini juga diharapkan dapat memberikan kontribusi pada dunia *data mining* dan pengolahan data dalam dunia olahraga serta berbagai industri lainnya yang memanfaatkan prediksi sebagai bagian dari aktivitas mereka.

DAFTAR PUSTAKA

- [1] A. S. Yudistira and A. Nugroho, "Prediksi Tim Juara Liga Premier Inggris Musim 2021/2022 Menggunakan Metode Naive Bayes," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1239–1243, 2022, doi: 10.20884/1.jutif.2022.3.5.328.
- [2] İ. H. KINALIOĞLU and C. KUŞ, "Prediction of UEFA Champions League Elimination Rounds Winners Using Machine Learning Algorithms," *Cunhur. Sci. J.*, vol. 41, no. 4, pp. 951–967, 2020, doi: 10.17776/csj.780391.
- [3] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.
- [4] J. Mariyanto *et al.*, "Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Implementasi Algoritma Naive Bayes Untuk Memprediksi zona dan Tingkat Penyebaran Covid-19 Di Provinsi Jambi," vol. 2, no. September, pp. 248–253, 2022, [Online]. Available: <https://ejournal.unama.ac.id/index.php/jakakom>.
- [5] M. Kafil, "Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 3, no. 2, pp. 59–66, 2019, doi: 10.36040/jati.v3i2.860.
- [6] F. D. Telaumbanua, P. Hulu, T. Z. Nadeak, R. R. Lumbantong, and A. Dharma, "Penggunaan Machine Learning," *J. Teknol. dan Ilmu Komput.*, vol. 3, no. 1, pp. 57–64, 2019.
- [7] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [8] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes," *Technol. J. Ilm.*, vol. 13, no. 3, p. 207, 2022, doi: 10.31602/tji.v13i3.7312.

- [9] I. K. Juni Arta, G. Indrawan, and G. R. Dantes, “Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode Technique for Others Reference By Similarity To Ideal Solution,” *JST (Jurnal Sains dan Teknol.*, vol. 5, no. 2, pp. 11–21, 2017, doi: 10.23887/jstundiksha.v5i2.8549.
- [10] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, “Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang,” *J. Janitra Inform. dan Sist. Inf.*, vol. 2, no. 1, pp. 9–19, 2022, doi: 10.25008/janitra.v2i1.142.
- [11] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2011.
- [12] N. Chatrina, Siregar, R. Ruli, A. Siregar, and M. Yoga, Distra, Sudirman, “Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ),” *J. Teknol.*, vol. 34, no. 1, pp. 102–110, 2020, [Online]. Available: <https://aperti.e-journal.id/teknologia/article/view/67>.
- [13] P. Domingos, “A Few Useful Things to Know About Machine Learning,” *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012, doi: 10.1145/2347736.2347755.
- [14] Qadrini L, Sepperwali A, and Aina A, “Decision Tree dan Adaboostpada Klasifikasi Penerima Program Bantuan Sosial,” *Decis. Tree Dan Adab. Pada Klasifikasi Penerima Progr. Bantu. Sos.*, vol. 2, no. 7, pp. 1959–1966, 2021.
- [15] Karsito and S. Susanti, “Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia,” *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.