

PREDICTION OF THE ENGLISH PREMIER LEAGUE CHAMPION TEAM FOR THE 2021/2022 SEASON USING THE NAÏVE BAYES METHOD

Alesandro Sulthan Yudistira¹, Adi Nugroho²

^{1,2}Program Studi Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia
Email: ¹672018245@student.uksw.edu, ²adi.nugroho@uksw.edu

(Naskah masuk: 16 Mei 2022, Revisi : 25 Mei 2022, diterbitkan: 24 Oktober 2022)

Abstract

The English Premier League is the top division league in England, the most popular football league for football fans. The competition date for the 2021/2022 season was announced on March 25, 2021 and the schedule was released on June 16, 2021. Manchester City won last season (2020/2021). The purpose of this study is to find out the results of Naive Bayes' accuracy in determining which team deserves to win in the English Premier League 2021/2022 season and which team does not deserve to win in the final standings. The method used for prediction is Naive Bayes, which is to predict an opportunity in the future based on events in the past which is then called the Bayes theorem. Combined with Naive's theorem, it can be said that a condition between attributes can be independent. Naive Bayes is an arrangement of a system that calculates a collection of opportunities that are summed at a frequency and a combination of values from the existing data set. Based on the research conducted, it was generated from predictions to determine the winning team quickly and accurately and from the test results it was necessary to process training data and test data with the Naive Bayes method, it was obtained an accuracy of 32.7% for the chances of Manchester City winning and Liverpool with 42.06% accuracy to be able to win the English Premier League season 2021/2022..

Keywords: English Premier, League Champion, Naive Bayes, Season, Team Prediction.

PREDIKSI TIM JUARA LIGA PREMIER INGGRIS MUSIM 2021/2022 MENGGUNAKAN METODE NAÏVE BAYES

Abstrak

Liga Premier Inggris adalah liga divisi teratas di Inggris, liga sepakbola paling digemari para fans sepakbola. Tanggal kompetisi musim 2021/2022 diberitakan pada 25 Maret 2021 dan jadwal dirilis pada 16 Juni 2021. Manchester City menjadi juara di musim lalu (2020/2021). Tujuan penelitian ini adalah untuk mengetahui hasil dari akurasi Naive Bayes dalam menentukan tim mana yang layak juara di Liga Premier Inggris musim 2021/2022 dan tim mana yang tidak layak juara pada klasemen akhir. Metode yang digunakan untuk prediksi adalah Naive Bayes yaitu untuk memprediksi suatu peluang di masa yang datang berdasarkan kejadian di masa lalu kemudian disebut dengan panggilan Teorema Bayes. Teorema tersebut digabungkan dengan Naive, bisa dikatakan sebuah kondisi antara atribut bisa bebas. Naive Bayes adalah susunan suatu sistem yang menghitung kumpulan peluang dijumlahkan pada suatu frekuensi serta gabungan nilai dari data set yang ada. Berdasarkan 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 dengan metode Naive Bayes 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.

Kata kunci: Juara Liga, Premier Inggris, Musim, Naive Bayes, Prediksi Tim.

1. PENDAHULUAN

Pada umumnya manusia mampu memprediksi atau memperkirakan suatu kejadian yang belum terjadi hanya dengan seringnya melihat pola dari suatu kejadian secara berulang. Prediksi adalah suatu proses yang bisa terjadi di masa yang datang melalui suatu informasi atau data dari masa yang berlalu dan

sekarang yang sudah ada, agar kesalahan dapat dicekikan. Contohnya adalah 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 perkomendasi pemilihan pemain kenaikan rata-rata rating tim hanya naik sebesar 0.98%. Kemampuan ini jika diterapkan ke dalam suatu sistem yang berupa *software* maupun *hardware*, contohnya mengklasifikasikan objek atau barang dalam proses industri, pencarian data dalam web, dan lain-lain. *Data mining* sebagai proses meringkasan pola tidak diketahui dan dapat ditambang dengan ukurannya yang 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 yang bisa terjadi di masa yang datang melalui suatu informasi atau data dari masa yang berlalu dan sekarang yang sudah ada, agar kesalahan dapat dicegah. Prediksi tidak harus memberikan jawaban yang benar tapi berusaha untuk menemukan jawaban yang bisa terjadi [2].

Juara merupakan individu maupun perkumpulan yang telah 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 yang 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 yang diusulkan. Ini dikenal sebagai *Teorema Bayes* karena oleh ilmuwan Inggris Thomas Bayes, yang memprediksi peluang masa depan berdasarkan pengalaman masa lalu. *Teorema* tersebut

digabungkan dengan *Naive*, yang 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].

Teorema Bayes adalah fakta dari seorang mengkonfirmasi kemungkinan dari hipotesis hanyalah tingkat kemunculan dari fakta tersebut dapat bisa terjadi dengan asumsi dari hipotesis dibandingkan tanpa asumsi [6].

Teorema Bayes bisa dikatakan dalam statistika untuk menghitung suatu peluang dalam hipotesa [7].

2. METODE PENELITIAN

2.1. Naïve Bayes

Metode yang digunakan adalah *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. Dengan metode ini bisa digunakan untuk memprediksi peluang-peluang dari data set menggunakan aplikasi rapid miner [8].

2.2. Data Mining

Data Mining adalah suatu proses hubungan dari kumpulan data tersimpan dalam penyimpanan dengan teknik statistik dan matematika [9].

2.3. Rumus

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

$$P(h | D) = \frac{P(D | h)P(h)}{P(D)} \quad (1)$$

dimana :

- D adalah kumpulan data pelatihan.
- h adalah asumsi.
- P(h | 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, $P(D) = P(D | h)P(h) + P(D$

$|\neg h)P(\neg h)$, yang dapat dengan mudah dihitung ketika menemukan $P(h | D)$ dan $P(h | \neg D)$

- $P(D|h)$ 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.

3. HASIL DAN PEMBAHASAN

Tabel 1. Data Training

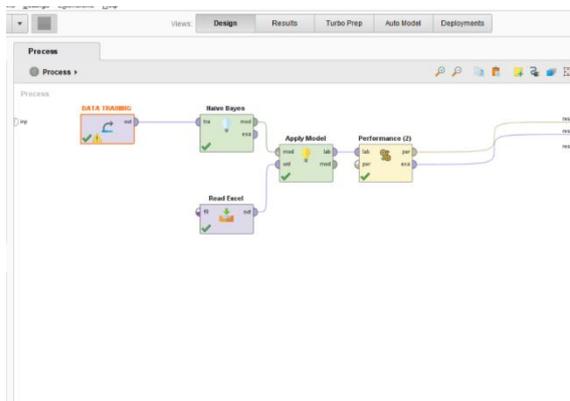
Posisi	Tim	Main	Menang	Seri	Kalah	Poin	Prediksi
1	Manchester City	38	27	5	6	86	Calon Juara
2	Manchester United	38	21	11	6	74	Calon Juara
3	Liverpool	38	20	9	9	69	Calon Juara
4	Chelsea	38	19	10	9	67	Tidak Calon Juara
5	Leicester	38	20	6	12	66	Tidak Calon Juara

Tabel 2. Data Testing

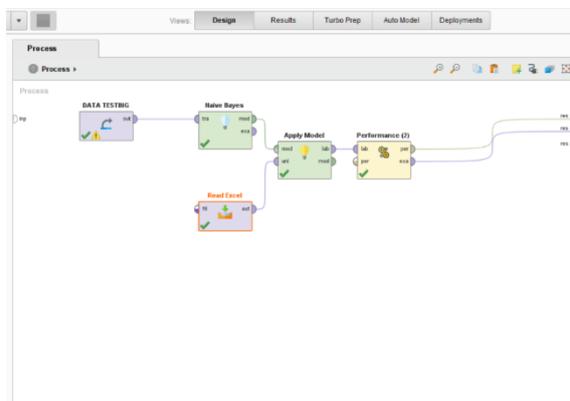
Posisi	Tim	Main	Menang	Seri	Kalah	Poin	Prediksi
1	Manchester City	30	23	4	3	73	Layak Juara
2	Liverpool	30	22	6	2	72	Layak Juara
3	Chelsea	29	17	8	4	59	Layak Juara
4	Spurs	30	17	3	10	54	Tidak Layak Juara
5	Arsenal	28	17	3	8	54	Tidak Layak Juara

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 [10].



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 [11].

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

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

Tabel 3. Head to Head Manchester City vs Liverpool

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

Pada tabel 3 berisi keterangan jadwal, musim, dan tempat serta hasil

Tabel 4. Keterangan atribut yang dihitung

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

26 Juli 2018	HOME	LOSE (1-2)	26 Juli 2018
11 April 2018	HOME	LOSE (1-2)	11 April 2018

Pada tabel 4 berisi pertemuan sebelumnya dari tim Liverpool melawan Manchester City

Tabel 5. Data kekuatan tim dari game FIFA 22

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

Pada tabel 5 berisi kekuatan tim dengan rata-rata nilai dari game FIFA 22

Menghitung setiap peluang juara dari kedua tim di musim sebelumnya.

P(WIN|X)

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

P(DRAW|X)

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

P(LOSE|X)

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

Menghitung setiap peluang juara dari kedua tim di musim ini.

P(WIN|Y)

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

P(DRAW|Y)

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

P(LOSE|Y)

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

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

4. DISKUSI

Dari hasil itu dapat dikatakan bahwa prediksi dengan menggunakan metode *Naive 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.

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

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

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi tim yang juara Liga Premier Inggris musim 2021/2022 menggunakan metode *Naive Bayes* dapat disimpulkan bahwa, 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 bahwa Liverpool juaranya.

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.

DAFTAR PUSTAKA

- [1] M. Sabransyah, Y. N. Nasution, and F. D. T. Amijaya, "Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung," *J. EKSPONENSIAL*, vol. 8, no. 2, pp. 111–118, 2017.
- [2] I. N. Farida, Y. S. Lestyningtyas, and F. S. Clustering, "IMPLEMENTASI FUZZY CLUSTERING UNTUK," vol. 1, no. 1, pp. 24–32, 1978.
- [3] A. F. Efendy, "Tingkat Stress Menjelang

- Perlombaan Terhadap Hasil Perlombaan Pada Atlet Atletik Kejuaraan Nasional Yuniior Dan Remaja Jawa Tengah Di Jakarta Tahun 2013,” *Univ. Negeri Semarang*, pp. 1–23, 2013.
- [4] A. Saleh, “Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” vol. 2, no. 3, pp. 207–217, 2015.
- [5] D. Untuk *et al.*, “Penggunaan data mining dengan metode naïve bayessian classification untuk memprediksi juara liga inggris musim 2006/2007,” 2007.
- [6] M. A. Yaqin, M. Z. Ramadhan, A. F. Jauhari, and A. G. Humami, “Optimasi Pemilihan Posisi Terbaik Pemain Muda Pada Game Football Manager 2018 Dengan Metode Naïve Bayes,” pp. 59–65, 2019.
- [7] N. Sulardi and A. Witanti, “SISTEM PAKAR UNTUK DIAGNOSIS PENYAKIT ANEMIA MENGGUNAKAN,” vol. 1, no. 1, pp. 19–24, 2020.
- [8] T. I. Andini, W. Witanti, and F. Renaldi, “Prediksi Potensi Pemasaran Produk Baru dengan Metode Naïve Bayes Classifier dan Regresi Linear,” pp. 27–32, 2016.
- [9] M. F. Rifai, H. Jatnika, and B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS),” vol. 12, no. 2, pp. 131–144, 2019.
- [10] H. F. Putro, R. T. Vlandari, and W. L. Saptomo, “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan,” vol. 8, no. 2, 2020.
- [11] M. H. Rifqo, A. Wijaya, and J. Pseudocode, “IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM PENENTUAN PEMBERIAN KREDIT,” vol. IV, no. September, pp. 120–128, 2017.
- [12] J. Teknologi, I. Dan, and K. Vol, “TEMATIK - Jurnal Teknologi Informasi Dan Komunikasi Vol. 8, No. 1 Juni 2021,” vol. 8, no. 1, pp. 96–103, 2021.
- [13] R. Wijayatun and Y. Sulisty, “Prediksi Rating Film Menggunakan Metode Naïve Bayes,” vol. 8, no. 2, pp. 60–63, 2016.
- [14] T. Informatika, U. Malikussaleh, and A. Utara, “PENERAPAN ALGORITMA NAIVE BAYES,” vol. 8, no. 1, pp. 884–898, 2014.
- [15] D. Kurniawaty, S. K. Sari, I. Cholissodin, F. Ilmu, and K. Universitas, “IMPLEMENTASI METODE NAÏVE BAYES CLASSIFIER UNTUK SELEKSI,” vol. 3, no. 4, pp. 273–278, 2016.