

Implementasi Metode Klasifikasi *Naïve Bayes* Untuk Memprediksi Kemenangan Klub Bundesliga Pada Pekan ke-15 Menggunakan Rapidminer

Wama Albertho Povay¹, Nurhaeni Irianti², Julius Rifaldo Putra³, Lodi Tekam⁴, Beto Ayun⁵,
 Heru Sutejo⁶,

^{1,2,3,4,5,6} Universitas Sepuluh Nopember Papua, Indonesia

* Correspondence e-mail; wamaalbertho@gmail.com

Article history

Submitted: 2024/12/01; Revised: 2024/12/05; Accepted: 2024/12/24

Abstract

Prediksi hasil kompetisi olahraga menjadi salah satu fokus penelitian yang menarik, khususnya dengan penerapan teknologi pembelajaran mesin. Studi ini bertujuan mengaplikasikan metode klasifikasi *Naïve Bayes* untuk memprediksi jumlah kemenangan klub-klub Bundesliga pada pekan ke 15 musim 2024-2025 dengan memanfaatkan aplikasi RapidMiner. Data yang dianalisis mencakup tabel klasemen Bundesliga musim berjalan serta data hasil pertandingan pekan ke 13 dan 14. Penelitian ini melibatkan beberapa tahapan utama, seperti pengolahan data, ekstraksi fitur, dan pelatihan model melalui RapidMiner. Hasil menunjukkan bahwa metode ini mampu mengidentifikasi pola performa tim secara efektif dan memberikan hasil prediksi dengan akurasi yang baik. Dengan demikian, penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem prediksi berbasis data untuk mendukung analisis di bidang olahraga.

Keywords

Naïve Bayes, RapidMiner, Bundesliga, Prediksi, Classification,



© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY SA) license, <https://creativecommons.org/licenses/by-sa/4.0/>.

PENDAHULUAN

Prediksi hasil kompetisi olahraga menjadi salah satu topik yang semakin diminati, terutama dengan kemajuan teknologi dalam analisis data dan pembelajaran mesin. Kemampuan untuk memperkirakan hasil kompetisi tidak hanya memiliki nilai akademik tetapi juga aplikasi praktis, seperti mendukung perencanaan strategi tim, evaluasi performa, dan kebutuhan komersial lainnya (Purnama & Wahyudi, 2024).

Sebagai salah satu liga sepak bola paling populer di dunia, Bundesliga menyediakan data yang kaya untuk dianalisis. Dengan memanfaatkan data historis yang tersedia, pola performa tim dapat dipelajari secara lebih mendalam. Salah satu metode analisis yang efektif dalam menangani data klasifikasi adalah algoritma *Naïve Bayes*, yang dikenal sederhana namun andal dalam menghasilkan prediksi (Saputra et al., 2023).

Penelitian ini mengimplementasikan algoritma Naïve Bayes untuk memprediksi klub-klub Bundesliga yang akan memenangkan pertandingan pada pekan ke 15 dengan memanfaatkan aplikasi RapidMiner (Kirana et al., 2024). Data yang digunakan mencakup hasil pertandingan pekan ke 13 dan 14 serta klasemen sementara musim 2024-2025. Analisis dilakukan melalui serangkaian tahapan, mulai dari pengolahan data, ekstraksi fitur, hingga pelatihan model.

Tujuan utama penelitian ini adalah mengevaluasi kemampuan Naïve Bayes dalam menganalisis data historis guna memberikan prediksi yang akurat (Julkarnain, 2024). Hasil penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem prediksi berbasis data di bidang olahraga dan menjadi dasar untuk penelitian selanjutnya.

METODE

Penelitian ini dilaksanakan melalui tahapan yang terstruktur dengan mencakup pengumpulan data, seleksi data, pengolahan data, dan implementasi Naïve Bayes (Rahmadanti et al., 2024). Penjelasan rinci dari setiap tahapan adalah sebagai berikut:



Gambar 1. Langkah Penelitian

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini meliputi tabel klasemen Bundesliga yang sedang berlangsung pada musim 2024-2025. Data yang dikumpulkan mencakup statistik tim, seperti jumlah kemenangan, kekalahan, selisih gol, dan total poin yang diperoleh tim.

2. Seleksi Data

Pada tahap seleksi data, data yang telah dikumpulkan dipilih untuk memastikan hanya informasi yang relevan yang digunakan dalam analisis. Proses ini mencakup pemilihan fitur yang berkontribusi signifikan terhadap prediksi, seperti performa tim secara keseluruhan, hasil pertandingan terakhir, dan kontribusi pemain utama.

3. Proses Data

Tahap ini berfokus pada pengolahan data agar siap digunakan dalam model klasifikasi. Beberapa langkah yang dilakukan pada tahap ini antara lain:

- **Preprocessing:** Menangani data yang tidak lengkap (missing values) dan menormalkan data agar konsisten.
- **Transformasi Data:** Memastikan data berada dalam format yang sesuai untuk diterapkan dalam algoritma Naïve Bayes.

4. Implementasi Naïve Bayes

Tamilis Synex: Multidimensional Collaboration

Setelah data diproses, algoritma Naïve Bayes diterapkan untuk membangun model prediksi berdasarkan data yang telah disiapkan (Fitriani, 2019). Implementasi model dilakukan menggunakan perangkat lunak RapidMiner untuk memastikan efisiensi dan akurasi dalam menghasilkan hasil prediksi. Model yang dibangun diuji menggunakan data klasemen sementara musim 2024-2025 untuk memprediksi klub Bundesliga yang akan meraih kemenangan pada pekan ke 15 (Triawan et al., 2024).

HASIL DAN PEMBAHASAN

Implementasi Pengumpulan Data

Data diambil langsung dari situs *transfermarkt.co.id*, yang menyediakan informasi detail terkait performa tim, klasemen, hasil pertandingan, dan statistik pemain. Situs ini dipilih karena kredibilitasnya dalam menyediakan data sepak bola yang lengkap dan akurat.

The screenshot displays the transfermarkt.co.id website interface. On the left, there's a navigation menu with options like 'BERANDA', 'TRANSFER & RUMOR', 'HARGA PASARAN', 'KOMPETISI', and 'STATISTIK'. The main content area shows 'GAMES OF MATCHDAY 13 BUNDESLIGA 24/25' with a list of matches including SC Freiburg vs VfL Wolfsburg (3:2), FC Augsburg vs Bayer 04 Leverkusen (0:2), 1.FSV Mainz 05 vs Bayern Munich (2:1), Borussia Mönchengladbach vs Holstein Kiel (4:1), 1.FC Union Berlin vs VfL Bochum (1:1), FC St. Pauli vs SV Werder Bremen (0:2), 1.FC Heidenheim 1846 vs VfB Stuttgart (1:3), and Borussia Dortmund vs Eintracht Frankfurt (1:1). On the right, the 'TABEL BUNDESLIGA' table is visible, listing clubs like Bayern Munich, E. Frankfurt, B. Leverkusen, RB Leipzig, VfL Wolfsburg, etc., with columns for matches played (M), wins (S), losses (K), goals scored (Gol), goal difference (+/-), and points (Pts).

Gambar 2. Hasil & Klasmen Pekan 13

The screenshot shows the transfermarkt.co.id website for Week 14. The 'GAMES OF MATCHDAY 14 BUNDESLIGA 24/25' section lists matches such as SC Freiburg vs VfL Wolfsburg (3:2), FC Augsburg vs Bayer 04 Leverkusen (0:2), 1.FSV Mainz 05 vs Bayern Munich (2:1), Borussia Mönchengladbach vs Holstein Kiel (4:1), 1.FC Union Berlin vs VfL Bochum (1:1), FC St. Pauli vs SV Werder Bremen (0:2), 1.FC Heidenheim 1846 vs VfB Stuttgart (1:3), Borussia Dortmund vs Eintracht Frankfurt (1:1), and RB Leipzig vs Eintracht Frankfurt (2:1). The 'TABEL BUNDESLIGA' table on the right shows updated standings, with Bayern Munich at the top (14 matches, 10 wins, 30 points) and VfL Bochum at the bottom (14 matches, 0 wins, -24 points).

Gambar 3. Hasil & Klasmen Pekan 14

Implementasi Proses Data

Setelah data diambil dari situs *transfermarkt.co.id*, informasi tersebut diekstraksi dan disusun ulang ke dalam format Excel menggunakan perangkat lunak Microsoft Excel. Setiap atribut data dimasukkan secara terstruktur ke dalam kolom yang sesuai,

sehingga memudahkan analisis dan proses klasifikasi menggunakan algoritma Naïve Bayes. Data ini kemudian dibagi menjadi dua kategori, yaitu data training dan data testing. Data training mencakup hasil pertandingan dan klasemen pada pekan ke 13 musim 2024/2025 yang digunakan untuk melatih algoritma, sementara data testing berisi hasil dan klasemen pada pekan ke 14 klasemen sementara musim 2024/2025 yang digunakan untuk menguji akurasi prediksi model (Yudistira & Nugroho, 2022).

Sebelum digunakan lebih lanjut, data yang telah dimasukkan ke dalam Excel divalidasi untuk memastikan tidak ada kesalahan seperti nilai kosong (missing values) atau duplikasi. Format data juga diatur sedemikian rupa agar kompatibel dengan perangkat lunak RapidMiner yang digunakan dalam analisis. Data yang telah divalidasi kemudian disimpan dalam dua file Excel terpisah, satu untuk data training dan satu untuk data testing, sehingga mempermudah proses implementasi algoritma Naïve Bayes (Metode et al., 2024).

Pada tahap ini data yang telah didapat akan dibuat sebagai data *training* dan data *testing* yang dimana dari data tersebut kemudian diambil ditentukan dan menghasilkan beberapa data atribut, yaitu:

1. Urutan Posisi.
2. Klub.
3. Main (Jumlah pertandingan yang telah di mainkan)
4. Poin.

Menang (Disini peneliti mengganti isi atribut apabila post = 1st maka Menang = Ya, dan apabila post ≠ 1st maka Menang = Tidak)

Tabel 1. Data Training

	A	B	C	D	E	F
	Urutan Posisi	Klub	Main	Menang	Poin	
1	1	Bayern Munich	13	Ya	33	
2	2	Eintracht Frankfurt	13	Tidak	27	
3	3	Bayer 04 Leverkusen	13	Ya	26	
4	4	RB Leipzig	13	Ya	24	
5	5	VfL Wolfsburg	13	Ya	21	
6	6	Borussia Dortmund	13	Tidak	21	
7	7	SC Freiburg	13	Tidak	21	
8	8	VfB Stuttgart	13	Ya	20	
9	9	1.FSV Mainz 05	13	Tidak	19	
10	10	SV Werder Bremen	13	Ya	19	
11	11	Borussia Mönchengladbach	13	Tidak	18	
12	12	1.FC Union Berlin	13	Tidak	16	
13	13	FC Augsburg	13	Tidak	16	
14	14	TSG 1899 Hoffenheim	13	Tidak	13	
15	15	FC St. Pauli	13	Tidak	11	
16	16	1.FC Heidenheim 1846	13	Tidak	10	
17	17	Holstein Kiel	13	Tidak	5	
18	18	VfL Bochum	13	Tidak	2	
19						
20						
21						

Tabel 1 menunjukkan informasi tentang urutan posisi, nama klub, jumlah pertandingan yang telah dimainkan, status kemenangan, dan total poin yang diperoleh setiap klub. Bayern Munich memimpin klasemen dengan 33 poin, sementara VfL Bochum berada di posisi terakhir dengan 6 poin. Semua klub tercatat telah memainkan 13 pertandingan, dengan poin yang tercatat berkisar antara 6 hingga 33, mencerminkan persaingan yang cukup ketat. Beberapa klub memiliki poin yang sama, seperti VfL Wolfsburg dan Borussia Dortmund dengan 21 poin, tetapi posisi

mereka dibedakan oleh kriteria tambahan. Data ini memberikan gambaran umum tentang performa klub-klub di Bundesliga musim ini.

Klasifikasi Algoritma Naïve Bayes

Algoritma Naïve Bayes didasarkan pada Teorema Bayes dan mengasumsikan bahwa setiap fitur saling independen. Rumus utamanya adalah sebagai berikut:

$$P(x) = \frac{P(C) \cdot P(C)}{P(x)} \quad (1)$$

X : Vector Masukan.

C : Kelas Tertentu.

$P(C|x_1, x_2, \dots, x_n)$: Probabilitas Kedua Kelas Berdasarkan Vector Masukan yang Diketahui (probabilitas posterior)

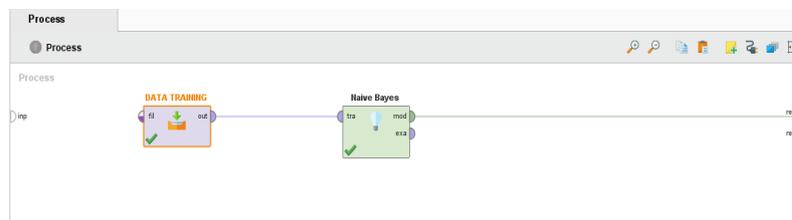
$P(C)$: Probabilitas Kelas yang Dicari (probabilitas prior) dari Seluruh Data.

$P(X|C)$: Probabilitas Setiap Masukan Berdasarkan Kondisi Kelas.

$P(x_1, x_2, x_3, \dots, x_n)$: Probabilitas Suatu Masukan dari Seluruh Data.

Penerapan Algoritma Naïve Bayes

Data training selanjutnya dilakukan pemodelan menggunakan algoritma *Naïve Bayes* pada aplikasi RapidMiner, Berikut Gambar 6 pemodelan data_training Menggunakan algoritma *Naïve Bayes*.



Gambar 4. Pemodelan



Gambar 5. Pemodelan Hasil Distribusi

Berdasarkan hasil dari pemodelan data_training menggunakan operator *Naïve Bayes* pada aplikasi RapidMiner seperti pada Gambar 6 diatas, maka didapatkan hasil berupa:

Distribution model for label attribute Menang:

Class Tidak (0.333)

4 distributions

Class Ya (0.667)

4 distributions

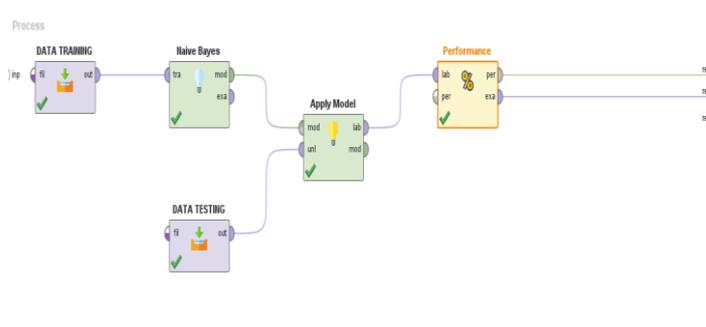
Pengujian Algoritma Naïve Bayes

Tabel 2. Data Testing

Urutan Posisi	Klub	Main	Menang	Poin
1	Bayern Munich	14	Tidak	33
2	Leverkusen	14	Ya	29
3	Frankfurt	14	Tidak	27
4	Leipzig	14	Tidak	27
5	Freiburg	14	Ya	24
6	Stuttgart	14	Ya	23
7	SV Mainz 05	14	Ya	22
8	Bor. Dortmund	14	Tidak	22
9	Werder Bremen	14	Ya	22
10	Wolfsburg	14	Tidak	21
11	Bor. M'gladbach	14	Ya	21
12	Union Berlin	14	Tidak	17
13	Augsburg	14	Tidak	16
14	SG Hoffenheim	14	Tidak	14
15	St. Pauli	14	Tidak	11
16	Heidenheim	14	Tidak	10
17	Holstein Kiel	14	Tidak	6
18	VfL Bochum	14	Tidak	3

Tabel 2 merupakan data testing adalah subset data yang menampilkan urutan posisi klub berdasarkan performa hingga pekan ke-14 data ini tidak digunakan selama proses pelatihan, yang berfungsi untuk menguji kemampuan model dalam memprediksi hasil yang belum pernah dilihat sebelumnya. Dengan menggunakan data_testing, dapat dievaluasi seberapa baik model *Naïve Bayes* dapat menghasilkan prediksi yang akurat mengenai potensi kemenangan yang akan di raih klub-klub bundesliga pada pekan ke 15 mendatang.

Selanjutnya menggunakan data_training & data_testing, Peneliti mencoba menerapkan data_training & data_testing tersebut dengan menggunakan algoritma *Naïve Bayes* pada aplikasi RapidMiner (Abisono Punkastyo et al., 2024). Pemodelan pada aplikasi RapidMiner tersebut dapat dilihat pada Gambar 6 di bawah:



Gambar 6. Pemodelan algoritma *Naïve Bayes*

Gambar 6 di atas ini menunjukkan pemodelan data_training (data pada Tabel 1) dan data_testing (data pada Tabel 2) pemodelan di aplikasi *RapidMiner* ini menggambarkan proses penerapan algoritma *Naïve Bayes* untuk klasifikasi. Proses dimulai dengan memuat data pelatihan menggunakan komponen "Data Training", dilanjutkan dengan pelatihan model menggunakan blok "Naïve Bayes". Model yang

Tamilis Synex: Multidimensional Collaboration

telah dibuat kemudian diterapkan pada data uji dari blok "Data Testing" melalui komponen "Apply Model". Setelah model menghasilkan prediksi, kinerjanya dievaluasi di blok "Performance" menggunakan berbagai metrik seperti akurasi, presisi, atau F1-score dengan membandingkan prediksi dengan label asli data uji (Penelitian et al., 2023). Gambar 6 di atas memberikan gambaran lengkap tentang bagaimana model dibangun dan diuji menggunakan aplikasi RapidMiner untuk menghasilkan prediksi mengenai potensi kemenangan yang akan di raih oleh klub-klub Bundesliga pada pekan ke 15.



Row No.	Menang	prediction(M...	confidence(...	confidence(...	Urutan Posisi	Klub	Main	Poin
1	Tidak	Ya	1.000	0.000	1	Bayern Munich	14	33
2	Ya	Ya	0.864	0.136	2	B. Leverkusen	14	29
3	Tidak	Tidak	0.251	0.749	3	E. Frankfurt	14	27
4	Tidak	Ya	0.864	0.136	4	RB Leipzig	14	27
5	Ya	Ya	0.864	0.136	5	SC Freiburg	14	24
6	Ya	Ya	0.992	0.008	6	VfB Stuttgart	14	23
7	Ya	Tidak	0.251	0.749	7	1.FSV Mainz 05	14	22
8	Tidak	Ya	0.864	0.136	8	Bor. Dortmund	14	22
9	Ya	Ya	0.864	0.136	9	Werder Brem...	14	22
10	Tidak	Ya	0.984	0.016	10	VfL Wolfsburg	14	21
11	Ya	Ya	0.766	0.234	11	Bor. Mgladba...	14	21
12	Tidak	Ya	0.864	0.136	12	Union Berlin	14	17
13	Tidak	Tidak	0.009	0.991	13	FC Augsburg	14	16
14	Tidak	Ya	0.864	0.136	14	TSG Hoffenh...	14	14
15	Tidak	Tidak	0.251	0.749	15	FC St Pauli	14	11
16	Tidak	Tidak	0.251	0.749	16	Heidenheim	14	10
17	Tidak	Tidak	0.251	0.749	17	Holstein Kiel	14	5
18	Tidak	Tidak	0.251	0.749	18	VfL Bochum	14	3

Gambar 7. Hasil Prediksi

Hasil pengujian yang dapat dilihat pada Gambar 7 diatas, didapatkan hasil prediksi dimana dari 18 klub bundesliga yang akan bertanding pada pekan ke 15 terdapat 11 klub di prediksi akan menang sementara 7 klub lainnya di prediksi tidak menang.

KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Naïve Bayes untuk memprediksi hasil pertandingan Bundesliga musim 2024-2025 menggunakan aplikasi RapidMiner. Berdasarkan klasemen sementara hingga pekan ke-14 dan data historis, model Naïve Bayes mampu memprediksi 11 klub yang akan menang pada pekan ke-15, sementara 7 klub lainnya diprediksi tidak menang. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes efektif dalam mengenali pola performa tim dari data historis, serta menunjukkan akurasi yang baik dalam prediksi. Penelitian ini memberikan kontribusi terhadap penggunaan metode berbasis data dalam menganalisis performa olahraga.

REFERENSI

Abisono Punkastyo, D., Septian, F., & Syaripudin, A. (2024). Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa.

- In Journal of System and Computer Engineering (JSCE) ISSN (Vol. 5, Issue 1).
- Fitriani, A. S. (2019). JTAM (Jurnal Teori dan Aplikasi Matematika) Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur. 3(2), 98–104. <https://doi.org/10.31764/jtam.v3i2.995>
- Julkarnain, M. (2024). Penerapan Algoritma Naive Bayes dalam Memprediksi Lulus Tepat Waktu Mahasiswa. 4(2). <https://doi.org/10.47709/digitech.v4i2.4963>
- Kirana, A. N., Nurhakim, B., Permana, S. E., Prihartono, W., & Dwilestari, G. (2024). IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI CUACA MENGGUNAKAN RAPIDMINER. In Jurnal Mahasiswa Teknik Informatika (Vol. 8, Issue 2). <https://www.kaggle.com/datasets/ananthr1/weather->
- Metode, I., Naïve, K., Untuk, B., Juara, M., Liga, L., Firman, A., Yanto, A., & Testiana, G. (2024). Implementation of Naïve Bayes Classification Method To Predict La Liga Champion. Jurnal Teknologi Sistem Informasi E-ISSN, 5(2), 128–139. <https://doi.org/10.35957/jtsi.v5i2.8028>
- Penelitian, L., Kepada, P., Universitas, M., Ali, D., & Nasrullah, A. (2023). Data Mining Algoritma Naive Bayes Untuk Memprediksi Jumlah Siswa Baru. EJECTS: E-Journal Computer, Technology and Informations System, 2(2).
- Purnama, L., & Wahyudi, T. (2024). Analisa Sentimen Tentang Piala Dunia u-20 Indonesia Menggunakan Algoritma K-Nearest Neighbor. Jurnal Sains Dan Teknologi, 6(2), 217–222. <https://doi.org/10.55338/saintek.v6i2.1397>
- Rahmadanti, R., Febriani, S. S., Syaputra, M. I., & Laraswati, D. (2024). PENERAPAN ALGORITMA NAÏVE BAYES UNTUK PREDIKSI TREN PERTUMBUHAN UMKM BERDASARKAN JENIS USAHA DI KOTA DEPOK. Jurnal INSAN (Journal of Information Systems Management Innovation, 4(2).
- Saputra, A., Subing, M., & Pratama, R. (2023). Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022 COMPARISON OF NAIVE BAYES TAXONOMY AND SUPPORT VECTOR MACHINES FOR ANALYZING TWITTER USER SENTIMENT ON FIFA WORLD CUP 2022. TEKNOMATIKA, 13(01).
- Triawan, E., Suarna, N., & Rinaldi Dikananda, A. (2024). KLASIFIKASI TIPE PENYERANG SEPAK BOLA LIGA INGGRIS BERDASARKAN DATA STATISTIK PEMAIN MENGGUNAKAN METODE NAIVE BAYES. In Jurnal Mahasiswa Teknik Informatika (Vol. 8, Issue 2).
- Yudistira, A. S., & Nugroho, A. (2022). PREDICTION OF THE ENGLISH PREMIER LEAGUE CHAMPION TEAM FOR THE 2021/2022 SEASON USING THE NAÏVE BAYES METHOD. Jurnal Teknik Informatika (Jutif), 3(5), 1239–1243.

Tamilis Synex: Multidimensional Collaboration

<https://doi.org/10.20884/1.jutif.2022.3.5.328>