

Mengimplementasikan Metode Klasifikasi dengan Menggunakan Algoritma Naïve Bayes dalam Menentukan Pemenang UCL 2024

Maulana Fansyuri^{1*}, Andika Arya Pratama², Fready Anggara³, Ikhwanul Maghsauma⁴, Jordi Ricaldo⁵, Mohamad Ryan Syekhan Ramadan⁶

¹⁻⁶Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}dosen02359@unpam.ac.id, ²andikaaryapratama0805@gmail.com,

³Freadyanggara@gmail.com, ⁴ikhwa5575@gmail.com, ⁵jordiricaldo555@gmail.com,

⁶ryansyekhanx@gmail.com.

(* : coresponding author)

Abstrak- Data mining merupakan proses penting dalam analisis data yang mencakup teknik seperti klasifikasi, klustering, dan asosiasi. Dalam penelitian ini, algoritma Naïve Bayes diterapkan untuk memprediksi pemenang Liga titleholders UEFA(UCL) 2024 dengan memanfaatkan data statistik tim, performa pemain, dan hasil pertandingan sebelumnya. Penelitian dilakukan menggunakan perangkat lunak RapidMiner untuk mengelola, menganalisis, dan mengevaluasi data. Dataset dibagi menjadi data pelatihan dan data pengujian dengan teknik validasi split. Model dievaluasi melalui confusion matrix untuk menghitung akurasi, presisi, recall, dan F1-score. Hasil pengolahan data menunjukkan bahwa algoritma Naïve Bayes memiliki potensi dalam memprediksi hasil pertandingan sepak bola, meskipun terdapat kendala dalam akurasi prediksi untuk klub tertentu. Kata kunci Naïve Bayes, Prediksi, Akurasi, Pre-processing data.

Kata Kunci: Naïve Bayes, Prediksi, Akurasi, Pre-Processing Data

Abstract - Data mining is a crucial process in data analysis that encompasses techniques such as classification, clustering, and association. In this study, the Naïve Bayes algorithm is applied to predict the winner of the UEFA Champions League (UCL) 2024 by utilizing team statistics, player performance, and previous match results. The research was conducted using RapidMiner software to manage, analyze, and evaluate the data. The dataset was divided into training and testing data using a split validation technique. The model was evaluated through a confusion matrix to calculate accuracy, precision, recall, and F1-score. The data processing results indicate that the Naïve Bayes algorithm has potential in predicting football match outcomes, although there are challenges in prediction accuracy for certain clubs.

Keywords: Naïve Bayes, Prediction, Accuracy, Data Pre-Processing

1. PENDAHULUAN

Data mining merupakan proses penting dalam analisis data yang melibatkan teknik klasifikasi, klustering, dan asosiasi. Dalam konteks olahraga, khususnya sepak bola, prediksi hasil pertandingan menjadi salah satu aplikasi menarik dari data mining. Salah satu metode yang dapat digunakan untuk klasifikasi adalah Naïve Bayes, yang berbasis pada teorema probabilitas. Penelitian ini bertujuan untuk mengimplementasikan algoritma Naïve Bayes dalam memprediksi pemenang Liga UEFA(UCL) 2024.

2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah proses yang menggunakan statistik, matematika, artificial intellegent dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar(Swastina, 2013). Data mining disisi lain adalah kegiatan yang mencakup pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di waktu yang akan datang.

2.2 RapidMiner

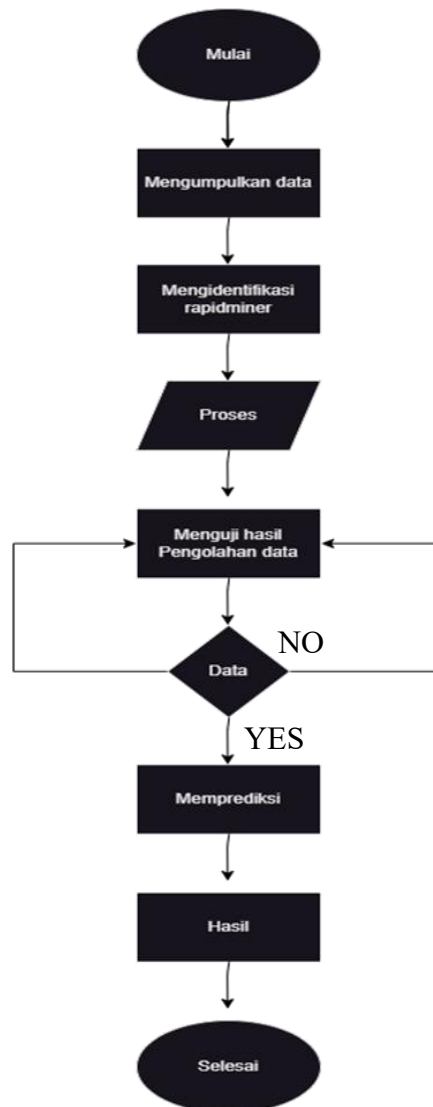
RapidMiner merupakan software yang bersifat terbuka(open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, textbook mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang fading baik. RapidMiner memiliki kurang lebih 500 driver data mining, termasuk driver untuk input, affair, data preprocessing dan visualisasi. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

2.3 Lokasi dan Waktu Penelitian

Adapun yang menjadi tempat penelitian adalah Warkop Niko No. 137 Kel. Petukangan Selatan Kecamatan Pesanggrahan Kota Jakarta Selatan. Kegiatan penelitian ini dilakukan sejak tanggal 12 Desember 2024.

2.4 Rancangan Penelitian

Perancangan penelitian ini digunakan untuk menguraikan dan menyelesaikan masalah dalam penelitian yang dapat dilihat pada *Gambar.1* berikut:



Gambar 1. Langkah Penelitian

Beberapa perancangan pada penelitian diatas, maka masing-masing langkah dapat diuraikan sebagai berikut:

1. Mengumpulkan Data
Pada tahap ini data diperoleh dari akun media social resmi UEFA champions league.
2. Mengidentifikasi Masalah
Tahap ini merupakan langkah untuk memproses tahap konservasi data yang diperoleh sesuai dengan bobot yang sudah ditentukan.
3. Proses Tahap
Ini bertujuan untuk mempermudah pemahaman terhadap isi record.
4. Menguji Hasil Pengolahan Data
Pada tahapan ini melakukan uji coba terhadap hasil pengolahan data dengan menggunakan software Rapid Miner.
5. Memprediksi Prediksi
Dilakukan untuk melihat data akurat dengan Algoritma Naïve bayes.

2.5 Pengumpulan Data

Data yang digunakan dalam penelitian ini mencakup statistik tim, performa pemain, dan hasil pertandingan sebelumnya. Sumber data diambil dari berbagai platform statistik olahraga dan basis data resmi UCL. Variabel yang dipertimbangkan meliputi:

- Jumlah gol yang dicetak
- Jumlah gol yang kebobolan
- Persentase penguasaan bola
- Jumlah kemenangan di babak penyisihan

2.6 Seleksi Data

Seleksi data adalah proses pemilihan subset dari data yang telah dikumpulkan untuk memastikan bahwa hanya informasi yang paling relevan dan bermanfaat yang digunakan dalam analisis. Proses ini meliputi:

1. Identifikasi Fitur Penting:

Menentukan atribut mana yang paling berpengaruh terhadap hasil pertandingan. Misalnya, jumlah gol dicetak dan jumlah kemenangan mungkin lebih relevan dibandingkan dengan atribut lain seperti jumlah kartu kuning.

2. Pembersihan Data:

Menghapus atau memperbaiki nilai yang hilang atau tidak valid. Misalnya, jika ada tim yang tidak memiliki data lengkap tentang performa pemainnya, maka data tersebut harus dihapus atau diisi dengan estimasi berdasarkan rata-rata performa tim lain.

3. Transformasi Data:

Mengubah data ke dalam format yang sesuai untuk analisis. Ini termasuk mengubah data kategorikal menjadi numerik (misalnya, mengubah status kemenangan menjadi 1 dan kekalahan menjadi 0) serta normalisasi nilai untuk memastikan semua fitur berada dalam skala yang sama.

4. Pengurangan Dimensi:

Jika dataset memiliki banyak fitur, teknik pengurangan dimensi seperti Principal Component Analysis (PCA) dapat digunakan untuk mengurangi kompleksitas tanpa kehilangan informasi penting.

5. Pembagian Dataset:

Setelah seleksi dilakukan, dataset dibagi menjadi set pelatihan dan set pengujian untuk memastikan model dapat diuji dengan baik terhadap data baru.

2.7 Proses Klasifikasi dengan Naïve Bayes

- 1. Preprocessing Data:** Data yang diperoleh dibersihkan dan dinormalisasi untuk menghilangkan nilai yang hilang dan outlier.
- 2. Pembagian Data:** Dataset dibagi menjadi dua bagian: data pelatihan (70%) dan data pengujian (30%).
- 3. Implementasi Algoritma Naïve Bayes:** Menggunakan rumus dasar Naïve Bayes:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Gambar 2. Rumus Naïve Bayes

Di mana:

- $P(C|X)P(C|X)$: Probabilitas kelas CC diberikan fitur XX
 - $P(X|C)P(X|C)$: Probabilitas fitur XX diberikan kelas CC
 - $P(C)P(C)$: Probabilitas a priori dari kelas CC
 - $P(X)P(X)$: Probabilitas fitur XX
- 4. Evaluasi Model:** Model dievaluasi menggunakan confusion matrix untuk menghitung akurasi, presisi, recall, dan F1-score. Evaluasi model adalah langkah penting dalam proses pengembangan algoritma klasifikasi, termasuk dalam penggunaan Naïve Bayes untuk memprediksi pemenang Liga Champions UEFA (UCL) 2024. Tahap ini bertujuan untuk menilai seberapa baik model yang telah dibangun dalam memprediksi hasil berdasarkan data yang telah disiapkan.

3. HASIL DAN PEMBAHASAN

Implementasi Pengumpulan Data Dalam penyelesaian masalah prediksi pemenang Uefa Champions League 2024, ada beberapa tahapan yang harus dilakukan, antara lain sebagai berikut:

- Menyiapkan Data training
- Menyiapkan Data set
- Melakukan perhitungan prediksi menggunakan rapidminer

3.1 Pengolahan Data

- Kriteria Data

Kriteria data yang digunakan dapat dilihat pada *Gambar.3* berikut ini:

Klub	T	M	S	K	GM	GK	SG	Poin
1 Liverpool	6	6	0	0	13	1	12	18
2 Barcelona	6	5	0	1	21	7	14	15
3 Arsenal	6	4	1	1	11	2	9	13
4 Leverkusen	6	4	1	1	12	5	7	13
5 Aston Villa	6	4	1	1	9	3	6	13
6 Inter	6	4	1	1	7	1	6	13
7 Brest	6	4	1	1	10	6	4	13
8 LOSC	6	4	1	1	10	7	3	13
9 Dortmund	6	4	0	2	18	9	9	12
10 Bayern	6	4	0	2	17	8	9	12

Gambar 3. Data

- Split Validation

Split Validation adalah teknik yang membagi data menjadi dua bagian secara acak, sebagian sebagai data training dan sebagian sebagai data testing. dengan Pada Gambar.3 dijelaskan bahwa dari hasil data klasemen sementara atau data testing di hitung dengan algoritma naïve bayes. Meskipun split validation lebih umum digunakan dalam konteks algoritma Naïve bayes dan regresi, kita juga dapat menerapkannya dalam konteks algoritma naïve bayes, meskipun dengan beberapa penyesuaian.

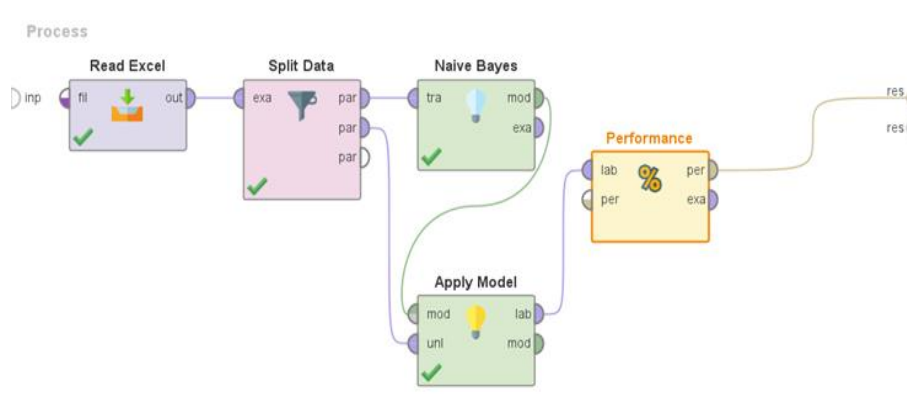
3.2 Proses Pengujian Data Menggunakan RapidMiner

Setelah data dianalisis dan diklasifikasikan menggunakan metode Algoritma Naïve bayes Maka untuk tahap selanjutnya adalah melakukan penghitungan prediksi menggunakan algoritma K-Means. Adapun aplikasi yang digunakan dalam pengujian klasifikasi Penentuan pemenang UCL 2024 ini adalah menggunakan aplikasi RapidMiner seperti ditunjukkan pada *Gambar.4* berikut:

	Jumlah gol ... <i>integer</i>	Jumlah gol ... <i>integer</i>	Jumlah kem... <i>integer</i>	Persentase ... <i>integer</i>	Hasil kemen... <i>polynomial label</i>	hasil <i>polynomial</i>
1	13	1	6	65	Menang	?
2	21	7	5	62	Menang	?
3	11	2	4	58	kalah	?
4	12	5	4	60	kalah	?
5	9	3	4	57	kalah	?
6	7	1	4	59	Menang	?
7	10	6	4	61	kalah	?
8	10	7	4	63	kalah	?
9	18	9	4	55	Menang	?
10	17	8	4	54	Menang	?

Gambar 4. Proses Pengujian Data

Dari gambar data set diatas gunakan format label untuk kolom hasil kemenangan. Lalu hubungkan semua operator untuk mendapatkan hasil prediksi.



Gambar 5. Proses Menghubungkan Operator

Pada Gambar 6 dapat dilihat bahwa Tim pemenang UCL 2024 adalah Borussia Dortmund

Row No.	Hasil keme...	prediction(H...	confidence(...	confidence(...	No	Tim	Jumlah gol ...	Jumlah gol ...	Jumlah kem...	Persentase ...
1	Menang	kalah	0	1	1	Liverpool	13	1	6	65
2	Menang	kalah	0	1	2	Barcelona	21	7	5	62
3	kalah	kalah	0	1	4	Bayer Leverk...	12	5	4	60
4	kalah	kalah	0	1	5	Aston Villa	9	3	4	57
5	Menang	kalah	0	1	6	Inter Milan	7	1	4	59
6	kalah	kalah	0	1	7	Brest	10	6	4	61
7	kalah	kalah	0	1	8	LOSC	10	7	4	63
8	Menang	Menang	1	0	9	Borussia Dort...	18	9	4	55

Gambar 6. Hasil Pengujian Naïve Bayes

Dari hasil perhitungan diatas selanjutnya didapatkan hasil nilai probabilitas $P(T) > P(Y)$ dengan kesimpulan bahwa, dimana Borussia Dortmund dengan perolehan gol 18, kebobolan 9 dan menang sebanyak 4 kali, didapatkan bahwa Klub Gagal atau tidak menjadi juara. Dari contoh hasil pengujian diatas memungkinkan untuk memprediksi apakah klub Borussia dormund dapat menjuarai Gelar di kompetisi UEFA Champions league dengan menggunakan data klasemen mereka pada musim liga yang sedang berjalan.

accuracy: 62.50%

	true Menang	true kalah	class precision
pred. Menang	1	0	100.00%
pred. kalah	3	4	57.14%
class recall	25.00%	100.00%	

Gambar 7. Nilai Akurasi Sistem

Dengan pengolahan data menggunakan aplikasi RapidMiner didapat nilai akurasi sistem sebesar 62.50%. Dari *Gambar.7* di jelaskan bahwa prediksi menang adalah 1 dan prediksi kalah adalah 4 dengan nilai precision sebesar 100% dan nilai recall sebesar 25%.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Naïve Bayes dapat digunakan untuk memprediksi pemenang Liga Champions UEFA dengan mengolah data historis seperti jumlah gol, kebobolan, penguasaan bola, dan jumlah kemenangan. Hasil analisis menggunakan RapidMiner menunjukkan bahwa Borussia dormund menjadi salah satu kandidat pemenang berdasarkan data statistik. Namun, hasil ini masih membutuhkan validasi lebih lanjut karena beberapa hasil prediksi, seperti kemungkinan kemenangan Borussia dormund, memberikan hasil yang tidak akurat. Dengan demikian, penggunaan metode ini dapat diperbaiki dengan penambahan data lebih lengkap dan metode evaluasi yang lebih robust untuk meningkatkan keakuratan model prediksi.

REFERENCES

- A. P. Wibawa, M. G. (2018). *Metode-metode klasifikasi*.
 Defiyanti, S. (2017). *Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik*.
 Lestari, P. I. (2018). *IMPLEMENTASI ALGORITME K-MEANS CLUSTERING DAN NAIVE BAYES CLASSIFIER UNTUK KLASIFIKASI DIAGNOSA PENYAKIT PADA KUCING*.
 Linita Sulistina, S. S. (2022). *Perbedaan tingkat akurasi metode k-means dan hierarchical clustering di bidang peramalan dan klasifikasi*.
 Narasati, T. S. (2024). *“Implementasi Algoritma Naïve Bayes Untuk Analisis Sentimen Ulasan produk makanan dan minuman di Tokopedia*.

- R. Mubarak, M. H. (2024). KLIK: Kajian Ilmiah Informatika dan Komputer komparasi Performa naive bayes gaussian dan k-nn untuk prediksi kelulusah mahasiswa dengan CRISP-DM.
- Retno. (2017). *Data Mining Teori dan Aplikasi Rapidminer*.
- S. Siska, G. A. (2023). *Implementasi Metode Naive Bayes pada Prediksi penyakit seliak*.
- Swastina, R. (2013). *Pengantar Data Mining: Teori dan Aplikasi*. Jakarta: Andi.
- Y. Yuliana, P. P. (2021). "Sistem Pakar Diagnosa Penyakit ISPA Menggunakan Metode Naive Bayes Classifier Berbasis Web.
- ZHALSABILLA, S. (2022). *INTEGRASI ALGORITMA KLUSTERISASI K-MEANS DAN KLASIFIKASI NAIVE BAYES PADA PENGELOMPOKAN JUMLAH SEKOLAH, MURID, DAN GURU DI PROVINSI JAWA BARAT*.