

Perbandingan Algoritma C4.5 dengan Naive Bayes Untuk Klasifikasi Curah Hujan

Muhammad Wendy Martadiansyah^{1*}, Arya Ghufron², Ryan Alfaret Hidayah³, Dwi Salzabila⁴, Lorenza Amanda⁵

^{1,2,3,4,5}Universitas Bina Sarana Informatika
Jl. Kramat Raya No 98, Jakarta Pusat, Indonesia

e-mail korespondensi : wendy.martadiandsyah75@gmail.com

Submit: 08-12-2024 | Revisi: 24-12-2024 | Terima: 26-12-2024 | Terbit online: 10-01-2025

Abstrak – Penelitian ini membandingkan performa algoritma C4.5 dan Naive Bayes dalam mengklasifikasikan curah hujan dengan menggunakan data curah hujan selama 10 tahun di Kecamatan Jatiwangin, Kabupaten Majalengka. Algoritma C4.5, yang berbasis pada pohon keputusan, menunjukkan kemampuan dalam menangani data yang rumit dengan membagi data berdasarkan aturan tertentu. Di sisi lain, Naive Bayes, yang menggunakan pendekatan probabilitas sederhana, memberikan hasil yang sangat tepat. Hasil dari penelitian ini menunjukkan bahwa Naive Bayes memiliki tingkat akurasi 100%, sedangkan C4.5 mencapai 99,83%. Walaupun Naive Bayes lebih unggul dalam hal akurasi, C4.5 lebih mudah dipahami karena tampilannya yang jelas. Penelitian ini memberikan informasi penting dalam memilih algoritma yang tepat untuk klasifikasi curah hujan, tergantung pada kebutuhan akan akurasi atau kemudahan pemahaman.

Kata Kunci : Curah Hujan, Data Mining, C4.5, Naive Bayes, Klasifikasi

Abstract – *This study compares the performance of the C4.5 and Naive Bayes algorithms in classifying rainfall using rainfall data for 10 years in Jatiwangin District, Majalengka Regency. The C4.5 algorithm, which is based on a decision tree, demonstrates the ability to handle complex data by dividing data based on specific rules. On the other hand, Naive Bayes, who uses a simple probability approach, gives very precise results. The results of this study show that Naive Bayes has an accuracy rate of 100%, while C4.5 reaches 99.83%. While Naive Bayes is superior in terms of accuracy, the C4.5 is easier to understand because of its clear appearance. This study provides important information in choosing the right algorithm for rainfall classification, depending on the need for accuracy or ease of understanding.*

Keywords : Rainfall, Data Mining, C4.5, Naive Bayes, Classification

1. Pendahuluan

Indonesia adalah negara yang dilintasi oleh garis khatulistiwa, sehingga negara yang dilalui oleh garis ini memiliki pola cuaca yang serupa sepanjang tahun. Pola yang paling umum adalah hangat dan lembap atau hangat dan kering sepanjang tahun. Banyak daerah di khatulistiwa juga dikenal sebagai lembap. Karena Indonesia dilintasi garis ini, maka cuaca yang terjadi hanya ada dua, yaitu musim hujan (lembap) dan kemarau (kering) [1]. Cuaca dan iklim adalah dua hal yang berbeda. Cuaca adalah semua kejadian yang terjadi di udara pada waktu dan tempat yang spesifik atau kondisi sesaat dari udara serta perubahan yang berlangsung dalam jangka pendek di suatu lokasi tertentu. Sementara itu, iklim adalah pola cuaca yang berlangsung dari waktu ke waktu (hari ke hari, bulan ke bulan, dan tahun ke tahun) atau keadaan rata-rata cuaca dalam jangka waktu yang cukup panjang atau area yang cukup luas [2].

Dalam ilmu geografi proses terjadinya hujan ialah berasal dari penguapan air laut dan permukaan akibat penyinaran matahari. Kemudian mengalami pengembunan (kondensasi) membentuk titik air yang berkumpul menjadi awan. Jika titik-titik air sudah berat, maka turunlah dalam bentuk hujan [3]. Curah hujan adalah jumlah air hujan yang terkumpul pada permukaan datar yang tidak menguap, meresap, atau mengalir. Satuan curah hujan selalu dinyatakan dalam milimeter atau inci, namun di Indonesia satuan curah hujan adalah milimeter (mm). Curah hujan dalam 1 (satu) milimeter memiliki arti dalam luasan satu meter persegi pada tempat yang datar tertampung air setinggi satu milimeter atau tertampung air sebanyak satu liter [4]. Curah hujan itu sendiri berpengaruh terhadap aktivitas manusia, seperti kegiatan bertani ataupun kegiatan sehari-harinya. Tidak hanya itu, jumlah curah hujan yang tinggi bisa menyebabkan bencana seperti banjir dan tanah longsor [5].



Para peneliti di sektor meteorologi sedang berupaya menciptakan teknik-teknik untuk memprediksi curah hujan. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) bekerja sama dengan para ilmuwan di luar BMKG untuk menganalisis data grid dan teknik pengelompokan curah hujan. Menurut BMKG, intensitas curah hujan dikategorikan menjadi tiga tipe, yaitu hujan ringan, hujan sedang, dan hujan lebat [6].

Untuk memprediksi musim hujan dan curah hujan dengan lebih akurat, salah satu metode yang dapat digunakan adalah *data mining*. *Data mining* adalah sebuah kegiatan semi otomatis untuk menerapkan teknik atau metode statistik, *machine learning*, matematika dan kecerdasan buatan di dalam proses ekstraksi serta identifikasi informasi yang terikat pada bermacam-macam database besar [7].

Data Mining memiliki beberapa teknik, salah satunya adalah klasifikasi. Teknik klasifikasi terdiri dari beberapa metode, dan *decision tree* adalah salah satu subbagian dari metode klasifikasi. Selanjutnya, metode *decision tree* memiliki algoritma, dan algoritma *C4.5* adalah salah satu algoritma yang termasuk dalam *decision tree* [8]. Selain algoritma *C4.5* *Naive Bayes* juga termasuk sebagai salah satu metode dalam teknik klasifikasi.

Dalam penelitian ini, digunakan dua algoritma yang populer dalam *data mining*, yaitu *C4.5* dan *Naive Bayes*, untuk menganalisis dan mengklasifikasikan tingkat curah hujan. Pemilihan algoritma *C4.5* merupakan salah satu pilihan yang paling umum dalam berbagai penelitian untuk menganalisis atau menguji data tentang kepuasan penilaian layanan. Sementara itu, algoritma *Naive Bayes* dipilih karena cara pengolahan nilai atributnya menggunakan kata-kata, yang menjadi salah satu kelebihan *Naive Bayes* dalam menguji hasil klasifikasi. Selain itu, dalam penelitian sebelumnya yang berhubungan dengan kepuasan atau penilaian pelayanan, masih jarang yang menggunakan *Naive Bayes* [9].

Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma *data mining*, yaitu *C4.5* dan *Naive Bayes*, dalam melakukan klasifikasi curah hujan di Kecamatan Jatiwangi, Kabupaten Majalengka, menggunakan software *RapidMiner*. Dalam penelitian ini, dua algoritma tersebut diuji berdasarkan metrik akurasi, precision, recall, dan evaluasi keseluruhan dari klasifikasi curah hujan selama periode 2008-2018. Dataset yang digunakan berisi berbagai atribut cuaca seperti suhu, kelembapan, kecepatan angin, dan kategori hujan. Dengan membandingkan performa kedua algoritma, penelitian ini bertujuan untuk memberikan wawasan tentang keunggulan dan kelemahan masing-masing algoritma dalam menangani masalah klasifikasi curah hujan berdasarkan data meteorologi yang tersedia.

2. Metode Penelitian



Gambar 1 Alur Penelitian

2.1 Pengumpulan Data

Tahap pertama adalah mengumpulkan data dari setiap aspek yang dibutuhkan. Pengumpulan data harus memenuhi beberapa prinsip yaitu mengumpulkan data selengkap lengkapnya, mempertimbangkan ketepatan data dan kebenaran data.[10]. Data yang digunakan dalam penelitian ini adalah data perkiraan curah hujan di Kecamatan Jatiwangi, Kabupaten Majalengka dari tahun 2008 hingga 2018. Data ini berasal dari Kaggle dan telah disesuaikan untuk mencakup wilayah yang relevan dengan penelitian.

Dataset yang dikumpulkan mencakup informasi seperti tanggal, suhu minimum, suhu maksimum, suhu rata-rata, kelembapan rata-rata, lama penyinaran, kecepatan angin terbesar, arah angin saat kecepatan maksimum, curah hujan (mm), kategori hujan, arah angin terbanyak, serta kategori curah hujan. Dataset lengkap dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/kelixo25/forecasting-rainfall>.

2.2 Tahap Seleksi Data

Tahap kedua adalah Tahap seleksi Data. Proses pemilihan data harus dilakukan sebelum tahap pemrosesan data. Pada tahap pemilihan data, data yang sudah ada akan dipilih untuk menemukan atribut yang penting dalam penelitian.[11]. Data yang telah dipilih menggunakan software *RapidMiner* yang dapat dilihat pada Tabel 1

Tabel 1 Pemilihan Atribut dan Label

No.	Atribut	Keterangan
1	Tanggal	Atribut
2	Suhu minimum	Atribut
3	Suhu maksimum	Atribut
4	Suhu rata – rata	Atribut
5	Kelembapan rata – rata	Atribut
6	Lama penyinaran	Atribut
7	Kecepatan angin rata – rata	Atribut
8	Kecepatan angin terbesar	Atribut
9	Arah angin saat kecepatan maksimum	Atribut
10	Curah hujan	Atribut
11	Kategori hujan	Atribut
12	Arah angin terbanyak	Atribut
13	Curah Hujan	Label

2.3 Pengolahan Data

Pada tahap ketiga adalah Pengolahan data. Sebelum memulai pengolahan data, langkah awal yang harus dilakukan adalah *pre-processing* data. *Pre-processing* data bertujuan untuk memperbaiki data yang memiliki nilai yang hilang (missing value) atau kesalahan penulisan. Dalam penelitian ini, normalisasi data yang dilakukan adalah menghapus data yang tidak memiliki informasi yang lengkap dan tidak diperlukan dalam penelitian.[12] Pada tahap ini, data yg telah diseleksi sesuai dengan keperluan penelitian diproses menggunakan *Altair AI Studio* Versi 2024.0.1. Dataset curah hujan yang dapat dilihat pada tabel 2. Dataset lengkap dapat di akses melalui : <https://www.kaggle.com/datasets/kelixo25/forecasting-rainfall>

Tabel 2 Dataset Curah Hujan

Tanggal	Suhu Min (°C)	Suhu Max (°C)	Suhu rata-rata (°C)	Kelembapan rata-rata (%)	Lama penyinaran (Jam)	Kecepatan angin rata-rata	Kecepatan angin terbesar	Arah angin maksimum	Curah hujan	Kategori Hujan	Arah angin terbanyak	Curah Hujan
01/02/2008	24	30,1	26,1	82	2,3	3	5	270	0	0	N	Tidak hujan/berawan
02/02/2008	24	31	26,3	90	1	2	4	180	0	0	S	Tidak hujan/berawan
03/02/2008	22,6	30,8	26	86	0,3	2	3	270	58,2	3	W	Hujan Lebat
04/02/2008	22,8	33,4	26,9	84	0,9	2	3	270	0,6	1	N	Hujan Ringan

Keterangan :

1. Tanggal : Pencatatan Data.
2. Suhu Minimum : Suhu terendah pada hari tersebut (dalam derajat celcius).
3. Suhu Maksimum : Suhu tertinggi pada hari tersebut (dalam derajat celcius).
4. Suhu Rata-rata : Rata-rata suhu harian (dalam derajat celcius).
5. Kelembapan Udara : Peresentase kelembapan udara rata – rata.
6. Lama Penyinaran : Total durasi penyinaran matahari dalam jam.
7. Kecepatan Angin Rata-rata : Kecepatan rata-rata angin (dalam satuan knot).
8. Kecepatan Angin Terbesar : Kecepatan angin maksimum yang tercatat (dalam satuan knot).
9. Arah Angin Saat Kecepatan Maksimum : Arah angin ketika kecepatan maksimum terjadi, dinyatakan dalam derajat.
10. Curah Hujan : Jumlah curah hujan pada hari tersebut (dalam milimeter).
11. Kategori Hujan : Kategori numerik curah hujan :
0 : Tidak ada hujan.
1 : Hujan ringan.
2 : Hujan sedang.
3 : Hujan lebat.
12. Arah Angin Terbanyak : Arah dominan angin pada hari tersebut, dinyatakan dalam derajat.

13. Curah Hujan : Kategori tekstual curah hujan: Tidak Hujan/Berawan, Hujan Sedang, Hujan Ringan, Hujan Lebat.

Tabel 3. Tabel dataset berdasarkan atribut

No	Atribut	Keterangan
1	Tanggal	01/02/2008, 02/02/2008, 03/02/2008, 04/02/2008 dst
2	Suhu minimum	24°C, 24 °C, 26,6°C, 22,8 °C dst
3	Suhu maksimum	30,1°C, 31,1°C, 30,8°C, 33,4°C dst
4	Suhu rata – rata	26,1 °C, 26,3 °C, 26°C, 26,9 °C dst
5	Kelembapan rata – rata	82 %, 90 %, 86 %, 84 % dst
6	Lama penyinaran	2,3 jam, 1 jam, 0,3 jam, 0,9 jam dst
7	Kecepatan angin rata – rata	3 knot, 2 knot, 2 knot, 2 knot dst
8	Kecepatan angin terbesar	5 knot, 4 knot, 3 knot, 3 knot dst
9	Arah angin saat kecepatan Maksimum	270 deg, 180 deg, 270 deg, 270 deg dst
10	Curah Hujan (mm)	0 mm, 0 mm, 58,2 mm, 0,6 mm dst
11	Kategori Hujan	0, 1, 2, 3
12	Arah angin terbanyak	N, S, W, N dst
13	Curah Hujan (Kategori)	Tidak Hujan/Berawa, Hujan Ringan, Hujan Sedang, Hujan Lebat

3. Hasil dan Pembahasan

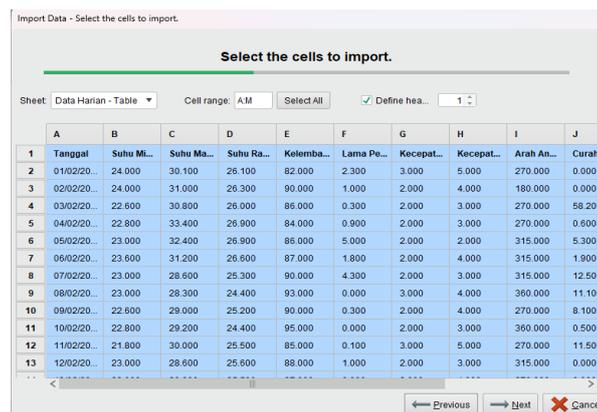
3.1 Analisis Data

Penelitian ini membahas perbandingan algoritma *C4.5* dengan *Navy Bayes* untuk klasifikasi curah hujan selama 10 tahun di Kecamatan Jatiwangin Kabupaten Majalengka. Dataset tersebut mencakup informasi Tanggal, Suhu minimum, suhu maksimum, suhu rata – rata, kelembapan rata – rata, lama penyinaran, kecepatan angin terbesar, arah angin saat kecepatan maksimum, curah hujan(mm), kategori hujan, arah angin terbanyak, curah hujan (kategori).

3.2 Proses Pengujian Menggunakan Metode Algoritma *C4.5*

Algoritma *C.45* merupakan pengembangan dari algoritma ID3, Algoritma *C.45* adalah suatu metode yang digunakan untuk menyelesaikan permasalahan logika untuk menghitung nilai.[13] Yang digunakan untuk membentuk *decision tree*. *Decision tree* dapat diartikan suatu cara untuk memprediksi atau mengklasifikasi yang sangat kuat, klasifikasi disini ialah tentang Curah Hujan. *Decision tree* membagi kumpulan data besar menjadi kumpulan data yang lebih kecil dengan menerapkan serangkaian aturan keputusan[14].

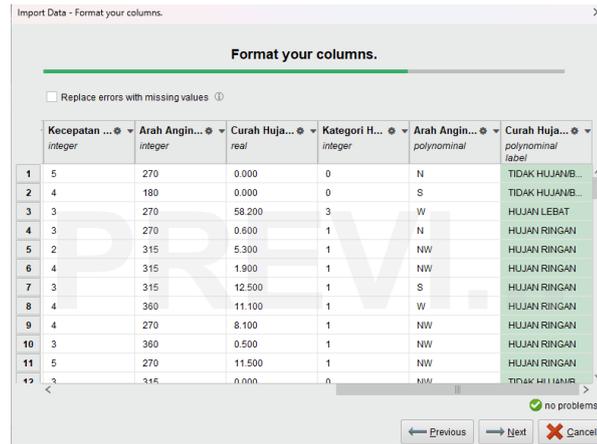
Agar dapat diolah dalam Software Rapidminer pertama kali data yang sudah di olah sebelumnya disimpan dalam format *microsoft excell* file berekstensi *.xls.



Gambar 2 Import Data

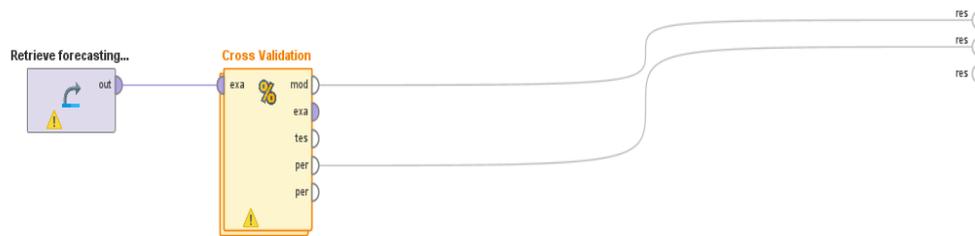
Pada Gambar 2 langkah yang dilakukan adalah mengimport file *excel* ke dalam software *RapidMiner*

Gambar 2 adalah import file dari Microsoft Excel kedalam aplikasi RapidMiner dan Gambar 3 adalah pemberian label atau kategori curah hujan.



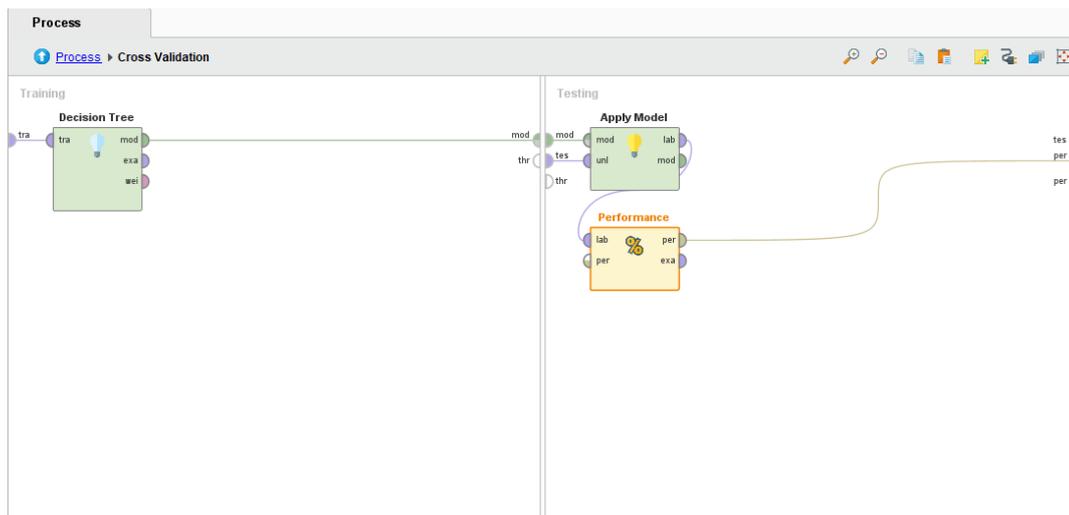
Gambar 3 Pemberian Label

Pada Gambar 3 proses pemberian label untuk menentukan apa yang kita cari yaitu curah hujan(Kategori).



Gambar 4 Cross Validation

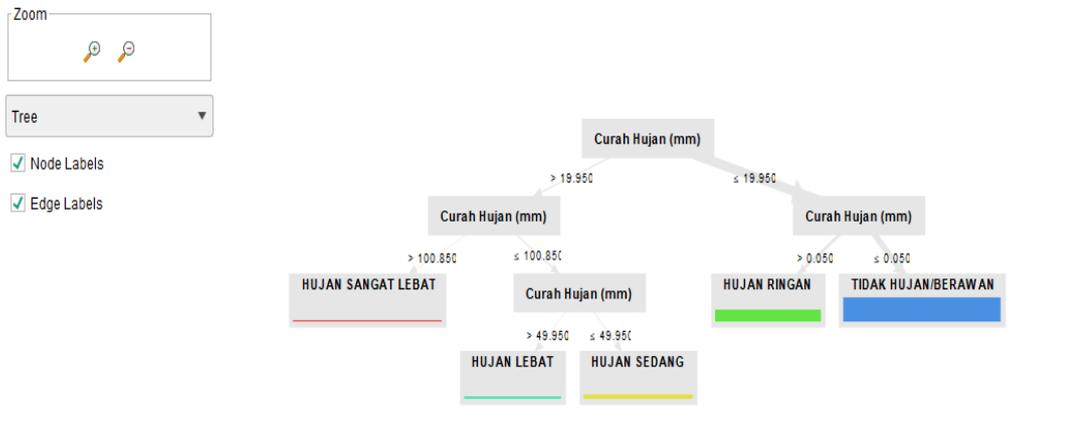
Pada Gambar 4 Proses implementasi algoritma *C4.5* dilakukan dengan menggunakan software rapid miner untuk memastikan model klasifikasi yang dihasilkan memiliki performa yang optimal dan konsisten. Tahapan validasi dilakukan dengan menggunakan metode *Cross Validation*, yang membagi dataset ke dalam beberapa subset untuk pelatihan dan pengujian model secara bergantian. Pendekatan ini bertujuan untuk mengurangi risiko bias dan *overfitting*



Gambar 5 Menentukan Performance Pohon Keputusan

Pada Gambar 4, menunjukkan alur proses *Cross Validation* yang digunakan untuk mengevaluasi model pohon keputusan dalam software RapidMiner. Prosesnya dimulai dengan melatih model menggunakan komponen pohon keputusan dan menggunakan data pelatihan untuk membangun model prediktif. Model yang dihasilkan diteruskan ke komponen *Apply Model* dan diterapkan ke data pengujian untuk menghasilkan prediksi. Hasil prediksi dari *Apply Model* dikirim ke komponen *Performance*, yang menghitung metrik evaluasi seperti akurasi, presisi atau *recall*.

Setelah seluruh langkah telah dikerjakan, RapidMiner akan melakukan proses dan menampilkan hasil keputusan pada *views Result*.



Gambar 6 Pohon Keputusan

Pada Gambar 6, pohon keputusan yang dihasilkan menunjukkan bahwa Curah Hujan (mm) menjadi variabel utama dalam klasifikasi dengan beberapa threshold, yaitu 19,950 mm, 0,050 mm, 100,850 mm, dan 49,950 mm. Data pada level pertama dipisahkan berdasarkan Curah Hujan > 19,950 mm untuk kategori HUJAN Sangat Lebat yang ditandai dengan warna merah, menunjukkan intensitas curah hujan tertinggi. Sementara itu, data dengan Curah Hujan ≤ 19,950 mm dipisahkan lagi menggunakan threshold Curah Hujan ≤ 0,050 mm untuk membedakan kategori Tidak Hujan/Berawan yang ditandai dengan warna biru dan Curah Hujan > 0,050 mm untuk kategori Hujan Ringan yang ditandai dengan warna hijau muda.

Pada cabang data dengan Curah Hujan > 19,950 mm, data dipisahkan lebih lanjut menggunakan threshold Curah Hujan > 100,850 mm untuk kategori Hujan Sangat Lebat yang tetap ditandai dengan warna merah. Sedangkan data dengan Curah Hujan ≤ 100,850 mm dipilah lagi berdasarkan threshold Curah Hujan > 49,950 mm untuk kategori Hujan Lebat yang ditandai dengan warna hijau, dan Curah Hujan ≤ 49,950 mm untuk kategori Hujan Sedang yang ditandai dengan warna kuning.

Penggunaan warna berbeda pada diagram ini membantu visualisasi tingkat intensitas curah hujan, di mana merah menunjukkan kategori Hujan Sangat Lebat, hijau menunjukkan kategori Hujan Lebat, kuning menunjukkan kategori Hujan Sedang, hijau muda menunjukkan kategori Hujan Ringan, dan biru menunjukkan kategori Tidak Hujan/Berawan. Diagram ini memberikan gambaran yang jelas untuk memvisualisasikan klasifikasi curah hujan berdasarkan tingkat intensitasnya.

Table View Plot View

accuracy: 99.83% +/- 0.19% (micro average: 99.83%)

	true TIDAK HUJA...	true HUJAN LEBAT	true HUJAN RING...	true HUJAN SED...	true HUJAN SAN...	class precision
pred. TIDAK HUJ...	2177	0	0	0	0	100.00%
pred. HUJAN LEB...	0	123	0	0	4	96.85%
pred. HUJAN RIN...	0	0	1027	0	0	100.00%
pred. HUJAN SE...	0	1	0	292	0	99.66%
pred. HUJAN SAN...	0	1	0	0	9	90.00%
class recall	100.00%	98.40%	100.00%	100.00%	69.23%	

Gambar 7 Performance Pohon Keputusan

Gambar 7 menunjukkan matriks evaluasi atau confusion matrix dari hasil klasifikasi, lengkap dengan metrik akurasi, precision, dan recall untuk setiap kelas. Model memiliki akurasi sebesar 99.83% ± 0.19%, yang

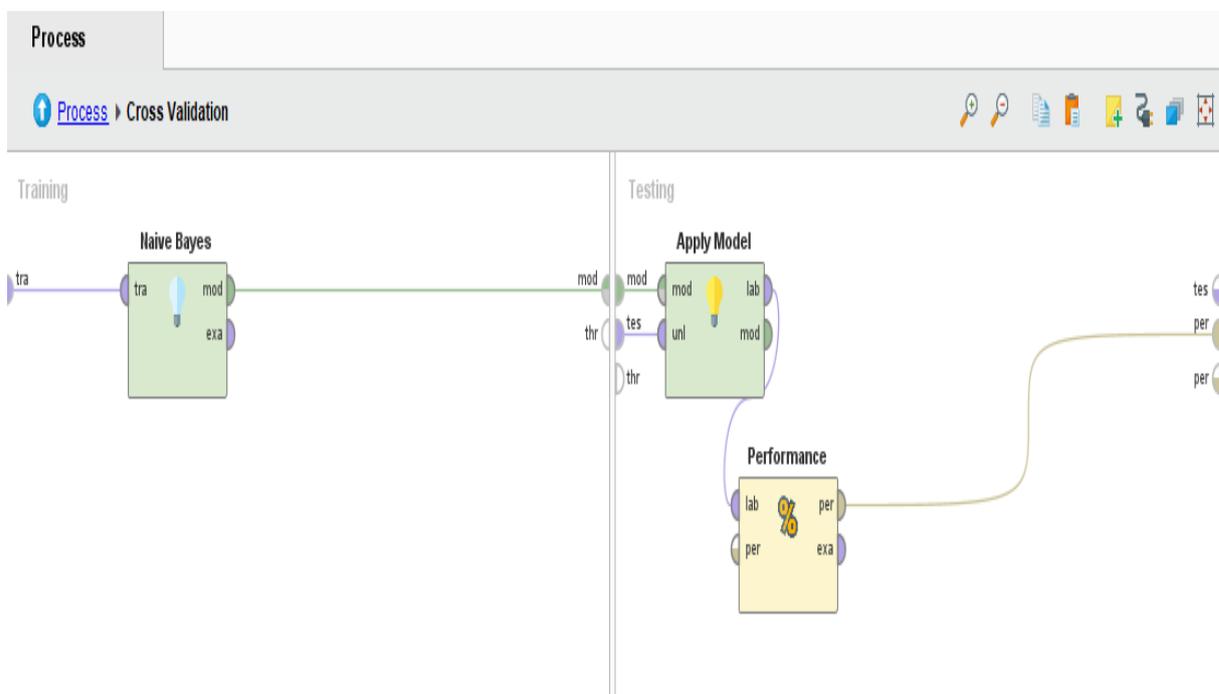
menunjukkan performa yang sangat baik dalam memprediksi kelas dengan benar. Setiap baris dalam matriks merepresentasikan jumlah data yang diprediksi oleh model untuk masing-masing kelas, sedangkan setiap kolom menunjukkan jumlah data yang sebenarnya termasuk dalam kelas tersebut.

Kategori Tidak Hujan/Berawan terklasifikasi dengan sempurna dengan nilai recall dan precision masing-masing sebesar 100%, di mana semua data yang sebenarnya termasuk kategori ini berhasil diprediksi dengan benar sebanyak 2177 data. Kategori HUJAN LEBAT memiliki recall sebesar 98.40% dan precision sebesar 96.85%, dengan sebagian besar data diklasifikasikan dengan benar sebanyak 123, namun terdapat 4 data yang salah diklasifikasikan sebagai HUJAN SANGAT LEBAT. Kategori HUJAN RINGAN juga menunjukkan performa sempurna dengan recall dan precision masing-masing sebesar 100%, tanpa ada kesalahan klasifikasi, di mana 1027 data berhasil diklasifikasikan dengan benar. Pada kategori HUJAN SEDANG, model berhasil mengklasifikasikan 292 data dengan benar, menghasilkan recall sebesar 100% dan precision sebesar 99.66%, meskipun terdapat 1 data yang salah diklasifikasikan sebagai HUJAN SANGAT LEBAT. Untuk kategori HUJAN SANGAT LEBAT, terdapat 9 data yang diklasifikasikan dengan benar, menghasilkan precision sebesar 90% dan recall sebesar 69.23%. Beberapa kesalahan klasifikasi terjadi, di mana 1 data salah diklasifikasikan ke HUJAN LEBAT dan 1 data lainnya ke HUJAN SEDANG.

Secara keseluruhan, model memiliki performa yang sangat baik, terutama pada kategori Tidak Hujan/Berawan, Hujan Ringan, dan Hujan Sedang, dengan nilai precision dan recall yang tinggi. Namun, pada kategori Hujan Sangat Lebat, model masih menunjukkan kesulitan dalam mengklasifikasikan beberapa data, yang terlihat dari nilai recall yang lebih rendah dibandingkan kategori lainnya. Hal ini menunjukkan perlunya perbaikan pada model untuk meningkatkan performa pada kategori tersebut.

3.3 Pengujian dengan menggunakan metode *Naïve Bayes*

Naive Bayes adalah metode klasifikasi probabilitas sederhana yang paling umum. Definisi lain dari *Naive Bayes* adalah teknik pembelajaran mesin yang menggunakan proses komputasi probabilitistik dan statistik. Metode ini dikemukakan oleh ilmuwan Inggris yaitu Thomas Bayes untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [15].



Gambar 8 Menentukan Performance Naive Bayes

Pada Gambar 8 Menunjukkan alur proses *Cross Validation* yang digunakan untuk mengevaluasi model naive bayes dalam software RapidMiner. Prosesnya dimulai dengan melatih model menggunakan komponen pohon keputusan dan menggunakan data pelatihan untuk membangun model prediktif. Model yang dihasilkan diteruskan ke komponen *Apply Model* dan diterapkan ke data pengujian untuk menghasilkan prediksi. Hasil prediksi dari *Apply Model* dikirim ke komponen *Performance*, yang menghitung metrik evaluasi seperti akurasi, presisi atau *recall*.

Table View Plot View

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true TIDAK HUJA...	true HUJAN LEB...	true HUJAN RIN...	true HUJAN SED...	true HUJAN SAN...	class precision
pred. TIDAK HUJ...	2177	0	0	0	0	100.00%
pred. HUJAN LE...	0	125	0	0	0	100.00%
pred. HUJAN RI...	0	0	1027	0	0	100.00%
pred. HUJAN SE...	0	0	0	292	0	100.00%
pred. HUJAN SA...	0	0	0	0	13	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	

Gambar 9 Performance Naive Bayes

Gambar 9 menunjukkan matriks evaluasi atau *confusion matrix* dari hasil klasifikasi, lengkap dengan metrik akurasi, precision, dan recall untuk setiap kelas. Model memiliki akurasi sebesar 100.00% ± 0.00%, yang menunjukkan performa sempurna dalam memprediksi kelas dengan benar. Setiap baris dalam matriks merepresentasikan jumlah data yang diprediksi oleh model untuk masing-masing kelas, sedangkan setiap kolom menunjukkan jumlah data yang sebenarnya termasuk dalam kelas tersebut.

Kategori TIDAK HUJAN/BERAWAN terklasifikasi dengan sempurna dengan nilai recall dan precision masing-masing sebesar 100%, di mana semua data yang sebenarnya termasuk kategori ini berhasil diprediksi dengan benar sebanyak 2177 data. Kategori HUJAN LEBAT juga menunjukkan performa sempurna dengan recall dan precision masing-masing sebesar 100%, di mana 125 data berhasil diklasifikasikan dengan benar tanpa ada kesalahan klasifikasi.

Kategori HUJAN RINGAN memiliki performa yang sama baiknya dengan recall dan precision sebesar 100%, tanpa ada kesalahan klasifikasi, di mana 1027 data berhasil diprediksi dengan benar. Pada kategori HUJAN SEDANG, model juga berhasil mengklasifikasikan 292 data dengan benar, menghasilkan recall dan precision masing-masing sebesar 100%, tanpa kesalahan prediksi. Untuk kategori HUJAN SANGAT LEBAT, model berhasil memprediksi 13 data dengan benar, menghasilkan nilai precision dan recall masing-masing sebesar 100%.

Secara keseluruhan, model memiliki performa yang sempurna pada semua kategori dengan nilai precision, recall, dan akurasi mencapai 100%. Hal ini menunjukkan bahwa model mampu mengklasifikasikan data curah hujan dengan sangat baik tanpa kesalahan prediksi pada dataset yang digunakan.

3.4 Perbandingan performance dua metode

Table View Plot View

accuracy: 99.83% +/- 0.19% (micro average: 99.83%)

	true TIDAK HUJA...	true HUJAN LEBAT	true HUJAN RING...	true HUJAN SED...	true HUJAN SAN...	class precision
pred. TIDAK HUJ...	2177	0	0	0	0	100.00%
pred. HUJAN LEB...	0	123	0	0	4	96.85%
pred. HUJAN RIN...	0	0	1027	0	0	100.00%
pred. HUJAN SE...	0	1	0	292	0	99.66%
pred. HUJAN SAN...	0	1	0	0	9	90.00%
class recall	100.00%	98.40%	100.00%	100.00%	69.23%	

Gambar 10 Performance Pohon Keputusan

● Table View ○ Plot View

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true TIDAK HUJA...	true HUJAN LEB...	true HUJAN RIN...	true HUJAN SED...	true HUJAN SAN...	class precision
pred. TIDAK HUJ...	2177	0	0	0	0	100.00%
pred. HUJAN LE...	0	125	0	0	0	100.00%
pred. HUJAN RI...	0	0	1027	0	0	100.00%
pred. HUJAN SE...	0	0	0	292	0	100.00%
pred. HUJAN SA...	0	0	0	0	13	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	

Gambar 11 Performance Naive Bayes

Dari gambar 10 dan gambar 11, *Decision Tree C4.5* memiliki tingkat akurasi 99,83%, yang berarti terdapat beberapa kesalahan dalam prediksi, khususnya pada kategori tertentu seperti Hujan Sangat Lebat. Algoritma ini bekerja dengan membagi data berdasarkan aturan tertentu yang dihasilkan dari pohon keputusan. Meski mampu menangkap hubungan non-linear antara fitur, model ini rentan terhadap overfitting, terutama jika pohon yang dihasilkan terlalu dalam. Teknik pruning (pemangkasan) biasanya diperlukan untuk mengatasi masalah ini. Keunggulan utama dari *Decision Tree* adalah kemampuannya untuk memberikan penjelasan yang jelas melalui visualisasi pohon keputusan, sehingga sangat membantu dalam analisis data yang membutuhkan interpretasi.

Naive Bayes memiliki tingkat akurasi 100%, menunjukkan bahwa model ini mampu memprediksi semua data dengan benar tanpa kesalahan. Algoritma ini didasarkan pada perhitungan probabilitas dengan asumsi bahwa fitur-fitur bersifat independen, meskipun asumsi ini tidak selalu berlaku di dunia nyata. Performa sempurna pada kumpulan data ini mungkin disebabkan oleh pola data yang sederhana atau distribusi data yang sesuai dengan asumsi *Naive Bayes*. Selain itu, *Naive Bayes* dikenal sangat efisien dan cepat, terutama ketika digunakan pada dataset dengan jumlah fitur yang banyak.

Perbandingan Kedua Algoritma, pada akurasi *Naive Bayes* memiliki akurasi 100%, lebih unggul dibandingkan *C4.5* dengan akurasi 99,83%, Kelebihan Utama a.*C4.5* rentan terhadap overfitting, meskipun teknik pruning dapat membantu mengatasi hal ini. b.*Naive Bayes* mengasumsikan independensi fitur, yang mungkin tidak sesuai jika data memiliki hubungan antar fitur yang kuat, dan pada Keterbatasan a.*C4.5* rentan terhadap overfitting, meskipun teknik pruning dapat membantu mengatasi hal ini. b.*Naive Bayes* mengasumsikan independensi fitur, yang mungkin tidak sesuai jika data memiliki hubungan antar fitur yang kuat.

4. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *Naive Bayes* lebih baik dalam ketepatan dibandingkan algoritma *C4.5* untuk pengenalan curah hujan di Kecamatan Jatiwangin. Dengan ketepatan 100%, *Naive Bayes* dapat memprediksi semua kategori dengan benar tanpa kesalahan. Sementara itu, *C4.5* memberikan ketepatan 99,83%, yang tetap menunjukkan kinerja yang sangat baik meskipun ada beberapa kesalahan dalam prediksi. Kelebihan utama *C4.5* adalah kemampuannya untuk memberikan gambaran visual yang memudahkan analisis lebih lanjut. Berdasarkan hasil tersebut, algoritma *Naive Bayes* disarankan untuk aplikasi yang membutuhkan prediksi yang tepat dengan cepat, sedangkan algoritma *C4.5* lebih sesuai untuk situasi di mana penjelasan model menjadi utama. Penelitian ini juga membuka kesempatan untuk menggabungkan algoritma atau menjelajahi metode lain agar bisa meningkatkan hasil dalam pengelolaan data iklim dan cuaca.

Referensi

- [1] A. K. Hidayat and Empung, "Analisis Curah Hujan Efektif Dan Curah Hujan Dengan Berbagai Periode Ulang Untuk Wilayah Kota Tasikmalaya Dan Kabupaten Garut," *J. Siliwangi*, vol. 2, no. 2, pp. 121–126, 2016.
- [2] R. Ruqoyah, Y. Ruhiat, and A. Saefullah, "Analisis Klasifikasi Tipe Iklim Dari Data Curah Hujan Menggunakan Metode Schmidt-Ferguson (Studi Kasus: Kabupaten Tangerang)," *J. Teor. dan Apl. Fis.*, vol. 11, no. 01, pp. 29–38, 2023, doi: 10.23960/jtaf.v11i1.3076.
- [3] E. dan F. Rika Summalia, "Hujan Perspektif Al-Qur'an dan Sains," *Hadrah J. Keislam. dan Perad.*, vol.

- 17, pp. 65–74, 2023, [Online]. Available: <https://ejournal.uinib.ac.id/jurnal/index.php/hadharah>
- [4] E. Q. Ajr and F. Dwirani, “Dengan Metode Polygon Thiessen Daerah Kabupaten Lebak,” vol. 2, no. 2, pp. 139–146, 2019.
- [5] M. Alviriza Ramadhan, F. Tri Anggraeny, and C. Aji Putra, “Klasifikasi Curah Hujan Harian Menggunakan Metode K-Nearest Neighbor,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3863–3869, 2024, doi: 10.36040/jati.v8i3.9817.
- [6] S. F. Rohmana, A. Rusgiyono, and S. Sugito, “Penentuan Faktor-Faktor Yang Mempengaruhi Intensitas Curah Hujan Dengan Analisis Diskriminan Ganda Dan Regresi Logistik Multinomial (Studi Kasus: Data Curah Hujan Kota Semarang dari Stasiun Meteorologi Maritim Tanjung Emas Periode Oktober 2018 – Maret 201,” *J. Gaussian*, vol. 8, no. 3, pp. 398–406, 2019, doi: 10.14710/j.gauss.v8i3.26684.
- [7] M. Juanda Saputra and M. Izman Herdiansyah, “Penerapan Naive Bayes Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang,” *J. Mantik*, vol. 6, no. 2, pp. 2502–2507, 2022.
- [8] D. Damayanti, “Implementasi Algoritma C4.5 Prediksi Produksi Komoditas Tanaman Perkebunan Berdasarkan Luas Lahan,” *TIN Terap. Inform. Nusant.*, vol. 2, no. 10, pp. 571–579, 2022, doi: 10.47065/tin.v2i10.1026.
- [9] M. Martin and L. Nilawati, “Comparison of C4.5 and Naïve Bayes Algorithms for Assessment of Public Complaints Services,” *J. Informatics Telecommun. Eng.*, vol. 5, no. 1, pp. 101–111, 2021, doi: 10.31289/jite.v5i1.5292.
- [10] A. Khusaeri *et al.*, “Algoritma C4.5 Untuk Pemodelan Daerah Rawan Banjir Studi Kasus Kabupaten Karawang Jawa Barat,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 132–136, 2017, doi: 10.33096/ilkom.v9i2.128.132-136.
- [11] L. Bachtiar and M. Mahradianur, “Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai,” *J. Inform.*, vol. 10, no. 1, pp. 28–36, 2023, doi: 10.31294/inf.v10i1.15115.
- [12] P. P. Haryoto, H. Okprana, and I. S. Saragih, “Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru,” *TIN Terap. Inform. Nusant.*, vol. 2, no. 5, pp. 358–364, 2021, [Online]. Available: <https://ejournal.seminar-id.com/index.php/tin/article/view/919>
- [13] Septri Wanti Siahaan¹ Kristin Daya Rohani Sianipar¹ P.P.P.A.N.W Fikrul Ilmi R.H Zer¹ Dedy Hartama, “Penerapan Algoritma C4.5 dalam Menentukan Faktor yang Dapat Meningkatkan Kemampuan Bahasa Inggris pada Mahasiswa”, *Jurnal Eksplora Informatika* vol. 10, no. 1, pp. 65–67, 2020.
- [14] H. D. Darmawan, D. Yuniarti, and Y. N. Nasution, “Klasifikasi Lama Masa Studi Mahasiswa Menggunakan Perbandingan Metode Algoritma C.45 dan Algoritma Classification and Regression Tree,” *J. Eksponensial*, vol. 8, no. 2, pp. 151–160, 2017.
- [15] M. F. Rozi, “Penerapan Data Mining Menggunakan Metode Naive Bayes Untuk Klasifikasi Data Penentuan Hasil Penjualan Dalam Strategi Pemasaran,” *J. Komput. Teknol. Inf. dan Sist. Inf.*, vol. 2, no. 2, pp. 444–454, 2023, doi: 10.62712/juktisi.v2i2.137.