

# Prediksi Pola Curah Hujan di Kota Semarang Menggunakan Algoritma *Naive Bayes*

<sup>1</sup>Fadhil Ibnu Is'ad, <sup>2</sup>Jihan Amirah Munif, <sup>3</sup>Aliefian Hendra Wardana, <sup>4</sup>Aprilisa Arum Sari, S.T, M.Kom

<sup>1</sup>Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa

<sup>1</sup>fadhilibnuisad@gmail.com

<sup>2</sup>Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa

<sup>2</sup>amunifjihan03@gmail.com

<sup>3</sup>Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa

<sup>3</sup>aliefianhendra661@gmail.com

<sup>4</sup>Teknik Informatika/Fakultas Ilmu Komputer

Universitas Duta Bangsa

<sup>4</sup>Aprilisa\_arumsari@udb.ac.id.

**Abstrak**— Curah hujan merupakan faktor penting dalam menentukan kondisi cuaca harian yang berdampak pada aktivitas sosial, ekonomi, dan lingkungan. Prediksi cuaca yang akurat dibutuhkan untuk mengantisipasi potensi bencana seperti banjir atau kekeringan. Penelitian ini bertujuan untuk mengklasifikasikan kondisi cuaca harian di Kota Semarang berdasarkan curah hujan menggunakan metode klasifikasi *Naive Bayes*. Data yang digunakan berasal dari BMKG berupa data iklim harian periode Januari hingga Juni 2025, dengan atribut yang digunakan meliputi curah hujan (RR), suhu minimum (Tn), suhu maksimum (Tx), suhu rata-rata (Tavg), kelembaban rata-rata (RH\_avg), dan kecepatan angin rata-rata (FF\_avg). Label cuaca dibentuk berdasarkan kaidah klasifikasi BMKG, yaitu cerah (RR=0), hujan ringan (0.1–5 mm), hujan sedang (5.1–20 mm), dan hujan lebat (>20 mm). Algoritma *Naive Bayes* diterapkan untuk mempelajari hubungan antar atribut dan label cuaca. Hasil pengujian menunjukkan bahwa model mampu melakukan klasifikasi cuaca harian dengan akurasi yang cukup baik, sehingga dapat digunakan sebagai sistem prediksi cuaca sederhana berbasis data historis.

**Kata kunci:** *Naive Bayes*, Klasifikasi Cuaca, Data Mining, Prediksi Curah Hujan, BMKG

**Abstract**— *Rainfall is a crucial factor in determining daily weather conditions, which impact social, economic, and environmental activities. Accurate weather prediction is essential to anticipate potential disasters such as floods or droughts. This study aims to classify daily weather conditions in Semarang City based on rainfall using the Naive Bayes classification method. The data used was obtained from BMKG and consists of daily climate data from January to June 2025, with attributes including rainfall (RR), minimum temperature (Tn), maximum temperature (Tx), average temperature (Tavg), average humidity (RH\_avg), and average wind speed (FF\_avg). Weather labels were generated based on BMKG classification rules: clear (RR=0), light rain (0.1–5 mm), moderate rain (5.1–20 mm), and heavy rain (>20 mm). The Naive Bayes algorithm was applied to learn the relationships between the attributes and the weather labels. The evaluation results show that the model is capable of classifying daily weather conditions with reasonably good accuracy, making it suitable as a simple weather prediction system based on historical data.*

**Keywords:** *Naive Bayes, Weather Classification, Data Mining, Rainfall Prediction, BMKG*

## I. PENDAHULUAN

Perubahan iklim yang semakin tidak menentu berdampak signifikan terhadap pola curah hujan di Indonesia, terutama dalam sektor pertanian, tata kelola sumber daya air, dan mitigasi bencana hidrometeorologi seperti banjir dan kekeringan [1]. Oleh karena itu, kemampuan untuk memprediksi curah hujan secara akurat menjadi kebutuhan yang mendesak dalam perencanaan berbasis iklim.

Dalam beberapa tahun terakhir, pendekatan data mining semakin banyak digunakan dalam pemodelan cuaca dan iklim karena kemampuannya dalam menganalisis data berskala besar secara efisien. Salah satu algoritma yang populer dalam klasifikasi data adalah *Naive Bayes*, yang dikenal memiliki keunggulan dalam hal kesederhanaan model, kecepatan proses, dan akurasi yang kompetitif [2]. Algoritma ini telah berhasil

diterapkan dalam berbagai studi, termasuk prediksi tingkat kepuasan pembelajaran daring [2] dan analisis sentimen terhadap informasi cuaca dari media sosial seperti Twitter BMKG [3].

Dalam konteks meteorologi, penelitian menunjukkan bahwa algoritma *Naive Bayes* mampu mengklasifikasikan curah hujan berdasarkan parameter iklim dengan akurasi tinggi serta waktu komputasi yang lebih cepat dibandingkan algoritma *Random Forest* [4]. Selain itu, metode ini juga terbukti efektif dalam estimasi curah hujan harian menggunakan pendekatan klasifikasi data [5].

Perbandingan antara algoritma *Naive Bayes* dan C4.5 menunjukkan bahwa meskipun C4.5 unggul dalam interpretasi aturan klasifikasi, *Naive Bayes* tetap lebih efisien dan sederhana [6]. Implementasi praktis menggunakan perangkat lunak RapidMiner juga telah dilakukan dalam konteks

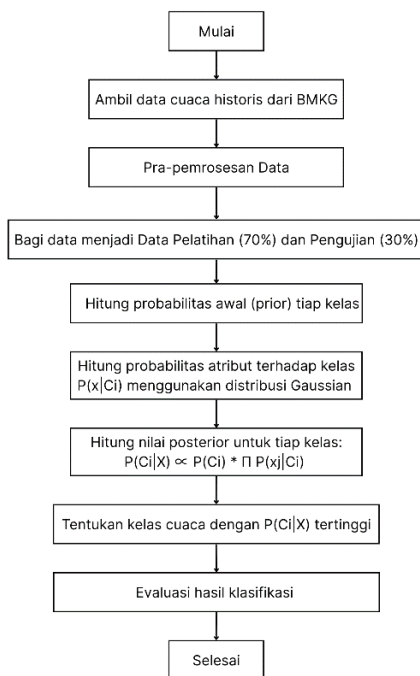
prediksi cuaca, menunjukkan fleksibilitas dan kemudahan penerapan metode ini [7].

Selain itu, pendekatan klasifikasi seperti algoritma C4.5 juga telah digunakan untuk mengestimasi potensi curah hujan ekstrem di wilayah rawan banjir [8]. Meskipun demikian, beberapa studi menunjukkan bahwa *Naive Bayes* tetap memberikan performa klasifikasi yang kompetitif jika dibandingkan dengan metode lain seperti *Artificial Neural Network* (ANN) dan *K-Nearest Neighbors* (K-NN) [9]. Bahkan, penggunaan gabungan algoritma seperti *Naive Bayes* dan *Decision Tree* sering kali memberikan hasil yang lebih optimal [10].

Sebagai tambahan, studi literatur memberikan landasan konseptual dan teknis mengenai penerapan *decision tree* dalam klasifikasi, yang relevan sebagai pembanding dengan metode *Naive Bayes* dalam konteks klasifikasi iklim dan cuaca [11].

Penelitian ini mengembangkan model klasifikasi curah hujan harian di Semarang berbasis *Naive Bayes* untuk prediksi cuaca yang efisien dan mendukung mitigasi risiko iklim lokal.

## II. METODOLOGI PENELITIAN



Gambar 1. Alur Penelitian

### 1. Studi Literatur

Penelitian ini diawali dengan studi literatur mengenai klasifikasi cuaca menggunakan algoritma *Naive Bayes*. Prediksi curah hujan menjadi penting karena dampaknya terhadap sektor pertanian di Indonesia [1]. *Naive Bayes* dipilih karena bersifat sederhana, cepat, dan efektif dalam menangani data kategorikal [2], [3]. Beberapa studi telah berhasil menerapkannya dalam prediksi cuaca dan klasifikasi curah hujan [4], [6], [7], serta membandingkannya dengan algoritma lain seperti Random Forest, C4.5, KNN, dan ANN [5], [8], [9], [10], [11]. Hasil studi tersebut mendukung pemilihan *Naive Bayes* sebagai algoritma utama dalam penelitian ini.

### 2. Pengumpulan Data Cuaca Historis

Penelitian ini menggunakan data iklim harian Kota Semarang yang diperoleh dari BMKG, sebagaimana juga digunakan dalam penelitian terkait klasifikasi data cuaca [3]. Variabel yang dikumpulkan meliputi curah hujan (RR), suhu minimum (Tn), suhu maksimum (Tx), suhu rata-rata (Tavg), kelembaban rata-rata (RH\_avg), dan kecepatan angin rata-rata (FF\_avg). Data diperoleh untuk periode Januari hingga Juni 2025 dan digunakan sebagai dasar dalam proses klasifikasi.

### 3. Pra-pemrosesan Data

Data mentah seringkali mengandung inkonsistensi, missing values, atau outlier, sehingga perlu dilakukan pra-pemrosesan sebelum digunakan dalam klasifikasi [4], [6]. Langkah-langkah pra-pemrosesan mencakup penanganan nilai hilang, normalisasi data, deteksi dan penghapusan outlier, serta pemilihan fitur (feature selection) untuk meningkatkan efisiensi dan akurasi model [9], [11].

Pada penelitian ini, data dibersihkan dari nilai tidak valid seperti RR = 8888, kemudian dilakukan normalisasi dan pengecekan konsistensi. Proses labelisasi dilakukan berdasarkan nilai curah hujan (RR), yang dibagi menjadi empat kelas cuaca:

- 0 mm → Cerah
- 0.1–5 mm → Hujan Ringan
- 5.1–20 mm → Hujan Sedang
- >20 mm → Hujan Lebat

Kategorisasi ini mempermudah proses klasifikasi oleh algoritma *Naive Bayes* dalam tahap selanjutnya [7].

#### 4. Pembagian Data (*Training dan Testing*)

Setelah pra-pemrosesan, data dibagi menjadi dua bagian yaitu data *training* untuk melatih model dan data *testing* untuk mengevaluasi performa model terhadap data yang belum pernah dilihat. Pembagian ini penting untuk mengukur kemampuan generalisasi algoritma [3]. Umumnya, rasio pembagian yang digunakan adalah 70% untuk training dan 30% untuk testing, sebagaimana diterapkan dalam penelitian serupa [3], [7].

#### 5. Pemilihan dan Implementasi Model *Data Mining*

Beberapa algoritma data mining dapat digunakan untuk klasifikasi atau prediksi cuaca, seperti K-Means untuk clustering [5], K-Nearest Neighbor (KNN) [8], *Decision Tree* [7], [9], Random Forest [10], dan *Gradient Boosting* [11]. Namun, pemilihan model disesuaikan dengan karakteristik data dan tujuan penelitian, yaitu klasifikasi curah hujan [3].

Algoritma *Naive Bayes* dipilih karena sederhana, efisien, dan efektif untuk data klasifikasi dengan jumlah atribut terbatas serta distribusi yang tidak seimbang [4], [6]. Implementasi dilakukan menggunakan perangkat lunak Weka versi 3.8, dengan data dalam format *Excel* yang dikonversi menjadi file CSV agar kompatibel dengan sistem [7].

#### 6. Pelatihan Model

Model yang dipilih, yaitu *Naive Bayes*, dilatih menggunakan 70% data training yang telah diproses. Selama pelatihan, Weka secara otomatis mengestimasi parameter probabilistik berdasarkan distribusi atribut terhadap kelas target [7]. Untuk menghindari *overfitting*, dapat diterapkan teknik validasi silang (*cross-validation*) guna mengevaluasi kestabilan model [3], [4].

#### 7. Evaluasi Model

Setelah pelatihan, model dievaluasi menggunakan 30% data testing. Metrik evaluasi yang digunakan mencakup akurasi [3], presisi, recall, F1-score, dan *confusion matrix* untuk klasifikasi [2], serta MAE, RMSE, dan  $R^2$  untuk regresi [10]. Hasil pengujian menunjukkan bahwa model *Naive Bayes* mencapai akurasi 93,33% dan kappa statistic 0,9054, dengan performa klasifikasi

yang sangat baik dalam membedakan kelas cuaca berdasarkan *confusion matrix* [3], [7].

### III. HASIL DAN PEMBAHASAN

Model klasifikasi cuaca harian yang dibangun menggunakan algoritma *Naive Bayes* berhasil diimplementasikan dengan baik melalui perangkat lunak Weka versi 3.8. Data yang digunakan dalam pelatihan dan pengujian terdiri dari enam parameter iklim yaitu curah hujan harian (RR), suhu minimum (Tn), suhu maksimum (Tx), suhu rata-rata (Tavg), kelembaban rata-rata (RH\_avg), dan kecepatan angin rata-rata (FF\_avg). Kelas target dibagi menjadi empat kategori yaitu Cerah, Hujan Ringan, Hujan Sedang, dan Hujan Lebat berdasarkan nilai RR.

Pengujian model dilakukan menggunakan metode percentage split dengan pembagian 70 persen data untuk pelatihan dan 30 persen untuk pengujian. Hasil evaluasi model menunjukkan tingkat akurasi sebesar 93.33 persen, dengan total 56 dari 60 data uji berhasil diklasifikasikan dengan benar. Nilai kappa statistic yang diperoleh sebesar 0.9054 menunjukkan tingkat kesepakatan yang sangat tinggi antara prediksi model dan kelas aktual. Selain itu, nilai mean absolute error sebesar 0.072 dan root mean squared error sebesar 0.179 menunjukkan bahwa kesalahan prediksi tergolong rendah.

Detail performa model untuk masing-masing kelas menunjukkan hasil yang sangat baik. Kelas Hujan Lebat dan Hujan Sedang memiliki true positive rate sebesar 1.000, yang berarti seluruh data uji pada kelas tersebut berhasil dikenali dengan benar oleh model. Kelas Cerah memperoleh nilai precision sebesar 0.958 dan recall sebesar 0.920, sedangkan kelas Hujan Ringan memiliki precision dan recall yang seimbang di angka 0.882. Nilai f-measure yang konsisten tinggi pada semua kelas menunjukkan bahwa model tidak hanya akurat, tetapi juga seimbang dalam mengenali semua kelas cuaca.

*Confusion matrix* dari hasil pengujian menunjukkan distribusi klasifikasi sebagai berikut:

- 10 data Hujan Sedang diklasifikasikan dengan benar.
- 15 dari 17 data Hujan Ringan diklasifikasikan dengan benar, dengan 1 data salah sebagai Cerah dan 1 sebagai Hujan Sedang.

- c) 8 data Hujan Lebat diklasifikasikan secara tepat seluruhnya.  
 d) 23 dari 25 data Cerah berhasil diklasifikasikan dengan benar, sementara 2 data diklasifikasikan sebagai Hujan Ringan.

Hasil ini menunjukkan bahwa model mampu mengenali pola karakteristik setiap kelas berdasarkan parameter cuaca dengan sangat baik. Nilai ROC Area untuk seluruh kelas berkisar antara 0.949 hingga 1.000, yang menandakan bahwa kemampuan pemisahan antar kelas oleh model sangat tinggi.

Secara keseluruhan, algoritma *Naive Bayes* terbukti mampu memberikan performa klasifikasi yang tinggi dan efisien untuk kasus klasifikasi cuaca harian. Keunggulan metode ini terletak pada kecepatan proses, efektivitas terhadap data numerik, dan kinerja stabil meskipun terjadi ketidakseimbangan jumlah data antar kelas. Hasil ini menunjukkan potensi penerapan *Naive Bayes* sebagai solusi prediksi cuaca berbasis data historis yang praktis dan ringan secara komputasi.

#### A. Pengujian menggunakan Algoritma Naïve Bayes dengan Bantuan Weka versi 3.8

Pada tahap ini, informasi yang dikumpulkan dari situs web BMKG akan diproses menjadi format yang diperlukan dan dilakukan untuk mengidentifikasi atribut data yang akan digunakan. Proses pengolahan data awal terdiri dari langkah-langkah berikut:

##### 1) Data Selection

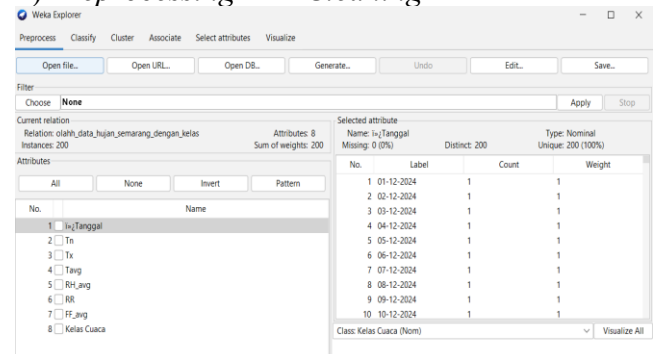
Tabel 1.1 *Dataset* Prediksi Cuaca

Tanggal	Tn	Tx	Tavg	RH_avg	RR	FF_avg
01/12/2024	25	32.4	28.3	83	7.4	1
02/12/2024	26.6	31.2	26.7	93	0.8	1
03/12/2024	25.2	31.6	28.1	83	27.3	1
04/12/2024	24.8	32.2	28.5	82	13.2	1
05/12/2024	26.2	32.2	29.2	77	0.6	2
06/12/2024	24.4	30.2	26.8	87	14.5	1
...	...	...	...	...	...	...
11/06/2025	24.4	32.6	28.5	81	35.4	2

12/06/2025	24.2	33.8	29.4	72	3.4	2
13/06/2025	24.2	34.2	28.4	79	0	2
14/06/2025	25.6	33.6	28.2	83	0	3
15/06/2025	25.2	34.2	29.8	73	5.4	2
16/06/2025	27.2	34.2	29	77	0	2
17/06/2025	25	31.4	27.4	87	83.4	1
18/06/2025	23.8	32.2	27.1	83	19.6	1

Tabel 1.1 merupakan dataset prediksi cuaca harian Kota Semarang yang diambil dari situs resmi BMKG, dengan rentang waktu mulai dari tanggal 1 Desember 2024 hingga 18 Juni 2025. Dataset ini terdiri dari enam atribut utama yang digunakan sebagai variabel input dalam proses klasifikasi, yaitu tanggal (Tanggal), suhu minimum (Tn), suhu maksimum (Tx), suhu rata-rata (Tavg), kelembaban relatif rata-rata (RH\_avg), curah hujan harian (RR), dan kecepatan angin rata-rata (FF\_avg).

##### 2) Preprocessing atau Cleaning

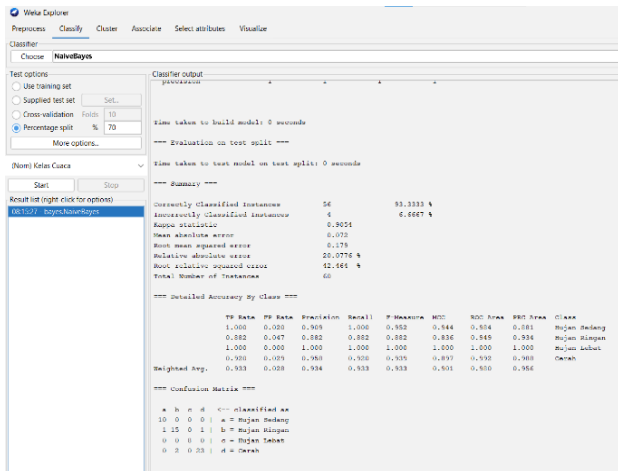


Gambar 2. Proses *Preprocessing*

Proses preprocessing dilakukan menggunakan perangkat lunak *Microsoft Excel* untuk tahap awal, seperti menghapus nilai tidak valid, mengecek konsistensi data, dan menambahkan kolom label cuaca berdasarkan klasifikasi curah hujan. Setelah data bersih dan siap, dataset dikonversi ke format .CSV dan diimpor ke perangkat lunak **Weka versi 3.8** untuk proses klasifikasi dan evaluasi model.

## 2) Klasifikasi Data

Data yang telah dipra-pemroses dan dikonversi ke format CSV diimpor ke Weka versi 3.8, kemudian diklasifikasikan menggunakan algoritma *Naive Bayes* dengan metode *percentage split 70:30*. Gambar 3 menampilkan antarmuka Weka saat proses klasifikasi berlangsung beserta hasil awalnya.



Gambar 3. Proses Klasifikasi Data

Gambar tersebut memperlihatkan bahwa proses klasifikasi berjalan secara otomatis dengan menampilkan ringkasan hasil klasifikasi, termasuk jumlah *instance* yang diklasifikasikan dengan benar dan metrik evaluasi lainnya.

### B. Hasil yang Dicapai

#### 1. Hasil Evaluasi Model *Naive Bayes*

Tabel 1.2 Hasil Evaluasi Model *Naive Bayes*

Kelas	Precision	Recall	F-Measure	ROC Area
<b>Cerah</b>	0.958	0.920	0.939	0.992
<b>Hujan Ringan</b>	0.882	0.882	0.882	0.949
<b>Hujan Sedang</b>	0.909	1.000	0.952	0.984
<b>Hujan Lebat</b>	1.000	1.000	1.000	1.000

Tabel 1.2 menunjukkan performa model *Naive Bayes* dalam mengklasifikasikan empat kelas cuaca berdasarkan data uji sebesar 30% dari total dataset. Nilai *precision*, *recall*, dan *f-measure* pada semua kelas berada di atas 0.88, menunjukkan performa

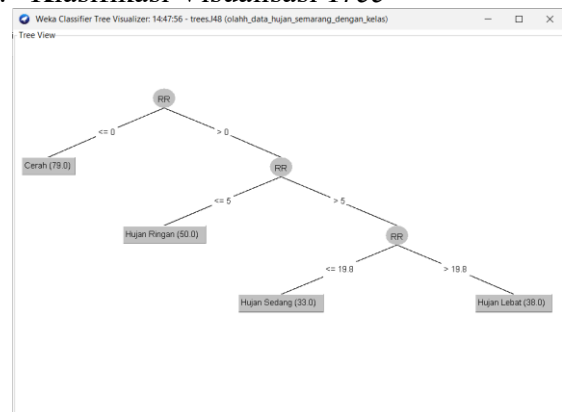
model yang tinggi. ROC Area yang mendekati 1 pada semua kelas mengindikasikan bahwa model mampu membedakan antar kelas dengan sangat baik.

Tabel 1.3 Confusion Matrix

Kelas Aktual	Cerah	Hujan Ringan	Hujan Sedang	Hujan Lebat
<b>Cerah</b>	23	2	0	0
<b>Hujan Ringan</b>	1	15	1	0
<b>Hujan Sedang</b>	0	0	10	0
<b>Hujan Lebat</b>	0	0	0	8

Berdasarkan hasil confusion matrix diatas, model berhasil mengklasifikasikan seluruh data pada kelas Hujan Sedang (10 data) dan Hujan Lebat (8 data) dengan akurasi 100%. Untuk kelas Hujan Ringan, 15 dari 17 data diklasifikasikan dengan benar, sedangkan sisanya keliru sebagai Cerah dan Hujan Sedang, namun akurasi masih relatif tinggi. Pada kelas Cerah, 23 dari 25 data berhasil diklasifikasikan dengan tepat, dan 2 data salah diprediksi sebagai Hujan Ringan, meskipun tidak sempurna tetapi hasil masih menunjukkan angka prediksi yang mendekati akurat. Hasil ini menunjukkan bahwa model memiliki akurasi tinggi, terutama pada kelas Hujan Sedang dan Hujan Lebat, dengan kesalahan klasifikasi yang rendah pada kelas lainnya.

#### 2. Klasifikasi Visualisasi *Tree*



Gambar 4. Hasil Klasifikasi Visualisasi Tree

Gambar 4 menunjukkan hasil visualisasi pohon keputusan dari proses klasifikasi menggunakan

algoritma *Naive Bayes Tree* pada data curah hujan di Kota Semarang. Atribut *RR* (curah hujan dalam milimeter) menjadi variabel utama dalam pemisahan kelas. Cabang pertama memisahkan data dengan nilai  $RR \leq 0$  yang diklasifikasikan sebagai **Cerah**, sedangkan nilai  $RR > 0$  diarahkan ke percabangan selanjutnya.

Jika  $RR \leq 5$ , maka data diklasifikasikan sebagai **Hujan Ringan**. Nilai  $RR > 5$  akan diperiksa kembali terhadap batas  $RR \leq 19.8$ , di mana data dengan nilai tersebut diklasifikasikan sebagai **Hujan Sedang**, dan nilai  $RR > 19.8$  diklasifikasikan sebagai **Hujan Lebat**. Setiap node daun menampilkan label kelas dan jumlah *instance* yang diklasifikasikan pada kelas tersebut. Visualisasi ini menggambarkan bahwa algoritma mampu mengidentifikasi ambang batas curah hujan yang signifikan dalam proses klasifikasi.

### C. Penjelasan Rumus *Naive Bayes*

Algoritma *Naive Bayes* menggunakan pendekatan probabilistik berdasarkan Teorema Bayes, dengan asumsi bahwa antar atribut saling independen terhadap kelas. Rumus umum dari probabilitas posterior suatu kelas  $C_i$  terhadap *instance* data  $X$  dengan  $n$  buah atribut adalah sebagai berikut:

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

Karena nilai  $P(X)$  konstan untuk semua kelas, maka perhitungan dapat disederhanakan menjadi:

$$P(C_i|X) \propto P(C_i) \cdot \prod_{j=1}^n P(x_j|C_i)$$

Dengan keterangan:

- $P(C_i)$  adalah probabilitas awal kelas  $C_i$  (*prior probability*), dihitung dari proporsi jumlah data pada kelas tersebut terhadap total data,
- $P(x_j|C_i)$  adalah probabilitas atribut  $x_j$  muncul dalam kelas  $C_i$ ,
- $X=(x_1, x_2, \dots, x_n)$  adalah *instance* data dengan  $n$  atribut.

Untuk atribut numerik seperti suhu, curah hujan, dan kelembaban, nilai  $P(x_j|C_i)$  dihitung menggunakan distribusi Gaussian (Normal), dengan rumus:

$$P(x_j|C_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp \left( -\frac{(x_j - \mu)^2}{2\sigma^2} \right)$$

Penjelasan rumus:

- $\mu$  adalah nilai rata-rata dari atribut  $x_j$  pada kelas  $C_i$ ,
- $\sigma$  adalah standar deviasi dari atribut  $x_j$  pada kelas  $C_i$ ,
- $\exp$  menyatakan fungsi eksponensial dengan basis  $e$ .

## IV. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa algoritma *Naive Bayes* efektif dalam mengklasifikasikan cuaca harian di Kota Semarang berdasarkan data iklim. Model yang dibangun mencapai akurasi 93.33 persen dengan performa terbaik pada kelas Hujan Sedang dan Hujan Lebat. Evaluasi dilakukan menggunakan Weka versi 3.8 dengan metode *percentage split* 70:30 dan metrik seperti *precision*, *recall*, dan *f-measure*.

Hasil tersebut membuktikan bahwa tujuan penelitian tercapai, yaitu membangun model klasifikasi cuaca yang andal dan akurat. Metode ini dapat diterapkan sebagai dasar sistem prediksi cuaca sederhana dan memiliki potensi untuk dikembangkan dalam aplikasi yang lebih luas, seperti sistem peringatan dini dan pengambilan keputusan berbasis data iklim.

## REFERENSI

- [1] Umi Zuhriyah, "Dampak Perubahan Iklim Terhadap Sektor Pertanian di Indonesia," *Tirto.Id*, pp. 1–8, 2022, [Online]. Available: <https://tirto.id/dampak-perubahan-iklim-terhadap-sektor-pertanian-di-indonesia-gyHf>
- [2] A. R. Damanik, S. Sumijan, and G. W. Nurcahyo, "Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma *Naive Bayes*," *J. Sistik Inf. dan Teknol.*, vol. 3, pp. 88–94, 2021, doi: 10.37034/jstik.v3i3.49.
- [3] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma *Naive Bayes* Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [4] N. A. Prakoso Indaryono, "Analisa Perbandingan Algoritma *Random Forest* Dan *Naive Bayes* Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 1, pp. 158–167, 2024, doi: 10.29100/jipi.v9i1.4421.
- [5] A. Kukuh Wahyudi, N. Azizah, and H. Saputro, "Data Mining Klasifikasi Kepribadian Siswa Smp Negeri 5 Jepara Menggunakan Metode *Decision Tree* Algoritma C4.5," *J. Inf. Syst. Comput.*, vol. 2, no. 2, pp. 8–13, 2022, doi: 10.34001/jister.v2i2.392.
- [6] M. W. Martadiansyah, A. Ghufron, R. A. Hidayah, D. Salzabila, and L. Amanda, "Perbandingan Algoritma C4.5 dengan *Naive Bayes* Untuk Klasifikasi Curah Hujan," *J. Komput. Antart.*, vol. 3, no. 1, pp.

- 8–17, 2025, doi: 10.70052/jka.v3i1.648.
- [7] A. Nur Kirana, B. Nurhakim, S. Eka Permana, W. Prihartono, and G. Dwilestari, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1637–1642, 2024, doi: 10.36040/jati.v8i2.8967.
- [8] I. Risnawati *et al.*, "Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan Berdampak Banjir Daerah Menggunakan Algoritma C4.5," *J. Insa. J. Inf. Syst. Manag. Innov.*, vol. 3, no. 2, pp. 78–84, 2023, doi: 10.31294/jinsan.v3i2.3050.
- [9] R. Martha and D. E. Herwindiati, "Prediksi Hujan Menggunakan Metode Artificial Neural Network, K-Nearest Neighbors, dan Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 4, pp. 859–865, 2024, doi: 10.47233/jteksis.v6i4.1650.
- [10] F. F. Nugraha, I. Sunandar, and C. Julian, "Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan Algoritma C4.5," *Teknologi*, vol. 7, no. March, pp. 10–20, 2022.
- [11] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, no. June, pp. 86716–86727, 2024, doi: 10.1109/ACCESS.2024.3416838.