

Penerapan Algoritma *Naïve Bayes* untuk Prediksi Potensi Hujan di Bandar Udara Tunggul Wulung Cilacap Berbasis Data *Radiosonde*

Suharti ^{1*}, Abdul Haq ², dan Noviansyah Dasaprawira ³

¹ Universitas Nahdlatul Ulama Al Ghazali, Cilacap; asmawirejasuharti@gmail.com

² Universitas Nahdlatul Ulama Al Ghazali, Cilacap; abdulhaq@unughad.id

³ Universitas Nahdlatul Ulama Al Ghazali, Cilacap; viarprawira93@unughad.id

*Penulis Korespondensi: Suharti

Abstract: Weather prediction, particularly rainfall potential, plays a crucial role in supporting airport operations to minimize the risk of flight disruptions, particularly in coastal areas prone to extreme weather. This study aims to provide practical solutions for airport managers in anticipating operational disruptions due to rainfall by utilizing historical data and probabilistic models for weather condition classification. Rain potential prediction is performed using radiosonde index data representing atmospheric lability conditions, namely the Lifted Index (LI), K-Index (KI), Showalter Index (SI), and Total Totals Index (TT) as predictor variables. Data were obtained from the Tunggul Wulung Cilacap Meteorological Station database for the 2020–2024 period, then used in the model training and evaluation process by handling missing values and measuring performance using a confusion matrix. The results show that the Naïve Bayes machine learning method is capable of producing a rainfall potential prediction model with an accuracy of 71.56%. These findings are expected to support more timely and efficient decision-making and have the potential to be applied in an early warning system based on actual observations for disaster mitigation.

Keywords: Naïve Bayes; Radiosonde; Metar; Dataset; Rainfall

Abstrak: Prediksi cuaca, khususnya potensi hujan, memiliki peranan penting dalam mendukung operasional bandar udara guna meminimalkan risiko gangguan penerbangan, terutama di wilayah pesisir yang rentan terhadap cuaca ekstrem. Penelitian ini bertujuan memberikan solusi praktis bagi pengelola bandara dalam mengantisipasi gangguan operasional akibat hujan melalui pemanfaatan data historis dan model probabilistik untuk klasifikasi kondisi cuaca. Prediksi potensi hujan dilakukan menggunakan data indeks radiosonde yang merepresentasikan kondisi labilitas atmosfer, yaitu Lifted Index (LI), K-Index (KI), Showalter Index (SI), dan Total Totals Index (TT) sebagai variabel prediktor. Data diperoleh dari database Stasiun Meteorologi Tunggul Wulung Cilacap selama periode 2020–2024, kemudian digunakan dalam proses pelatihan dan evaluasi model dengan penanganan missing values serta pengukuran kinerja menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa metode machine learning Naïve Bayes mampu menghasilkan model prediksi potensi hujan dengan nilai akurasi sebesar 71,56%. Temuan ini diharapkan dapat mendukung pengambilan keputusan yang lebih tepat waktu dan efisien serta berpotensi diterapkan dalam sistem peringatan dini berbasis observasi aktual untuk mitigasi bencana.

Kata kunci: Naïve Bayes; Radiosonde; Metar; Dataset; Hujan

Direvisi: July 4, 2026
Diterima: July 21, 2026
Diterbitkan: July 27, 2026
Versi sekarang: July 30, 2026



Hak cipta: © 2026 oleh penulis.
Diserahkan untuk kemungkinan publikasi akses terbuka berdasarkan syarat dan ketentuan lisensi Creative Commons Attribution (CC BY SA) (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Pendahuluan

Bandar udara adalah suatu area tertentu yang berada di daratan atau di laut yang berfungsi sebagai tempat untuk pesawat udara melakukan lepas landas dan pendaratan yang dilengkapi dengan sarana keselamatan, keamanan, dan penunjang sesuai standar penerbangan sipil [1]. Unit Penyelenggara Bandar Udara Tunggul Wulung Cilacap berperan dalam mengelola operasional penerbangan dan memberikan pelayanan kepada masyarakat [2]. Sementara itu, Stasiun Meteorologi Tunggul Wulung (BMKG Cilacap) bertugas menyediakan informasi cuaca secara akurat dan tepat waktu untuk mendukung keselamatan penerbangan [3].

Kabupaten Cilacap berada di pesisir selatan Pulau Jawa memiliki pola cuaca yang berubah-ubah dengan curah hujan tinggi. Hujan lebat dapat menurunkan jarak pandang, menyebabkan genangan di landasan pacu, serta memicu turbulensi akibat angin kencang sehingga meningkatkan risiko keselamatan, khususnya saat lepas landas dan pendaratan [4]. Oleh karena itu, diperlukan koordinasi yang baik antar pihak terkait dengan dukungan informasi cuaca yang akurat untuk mewujudkan operasional penerbangan yang aman, lancar, dan efisien [5].

Informasi cuaca diperoleh melalui laporan METAR (Meteorological Aerodrome Report) yang dihasilkan oleh AWOS (Automated Weather Observing System) dalam format standar internasional. Data tersebut memuat informasi arah dan kecepatan angin, jarak pandang, serta kondisi awan di sekitar bandara [6]. Namun, sistem ini masih memiliki keterbatasan dalam mendeteksi fenomena cuaca lokal yang berubah cepat, seperti wind shear dan microburst, serta memiliki jangkauan pengamatan yang terbatas sehingga belum mampu merepresentasikan kondisi atmosfer secara menyeluruh [7].

Untuk mengatasi keterbatasan tersebut, diperlukan pemanfaatan data observasi udara atas (radiosonde) yang mampu memberikan informasi profil vertikal atmosfer berupa suhu, kelembapan, dan tekanan udara pada berbagai lapisan atmosfer. Integrasi data radiosonde terbukti efektif dalam menganalisis dan memprediksi potensi hujan serta mendukung mitigasi bencana hidrometeorologis. Pengembangan sistem berbasis data indeks radiosonde diharapkan dapat meningkatkan akurasi prediksi potensi hujan dan penyebaran informasi cuaca secara lebih efektif [8].

Salah satu metode yang dapat digunakan adalah Naive Bayes, yaitu metode klasifikasi berbasis probabilitas yang memanfaatkan Teorema Bayes dengan asumsi setiap atribut saling independen [9]. Metode ini dipilih karena memiliki kompleksitas rendah, proses komputasi yang efisien, serta mampu memberikan kinerja yang kompetitif pada dataset berukuran besar [10]. Dibandingkan metode seperti Random Forest, Naive Bayes memiliki keunggulan dalam kecepatan pelatihan dan kemudahan interpretasi model, meskipun Random Forest umumnya menghasilkan akurasi yang lebih tinggi dengan kompleksitas komputasi yang lebih besar [11]. Berdasarkan pertimbangan tersebut, penelitian ini bertujuan menerapkan algoritma Naive Bayes untuk prediksi potensi hujan yang mendukung keselamatan serta kelancaran operasional penerbangan di Bandar Udara Tunggul Wulung Cilacap.

2. Tinjauan Literatur

2.1. Data Mining dan Machine Learning

Data mining merupakan proses menganalisis data dengan jumlah banyak atau rumit secara otomatis untuk menemukan pola, tren, dan informasi yang tersembunyi yang dapat membantu dalam pengambilan keputusan [12]. Proses ini didukung oleh metode statistika, matematika, dan teknologi Artificial Intelligence. Tahapan data mining secara umum meliputi eksplorasi (pembersihan, normalisasi, transformasi, dan penanganan data tidak lengkap), pembangunan serta validasi model, dan penerapan model pada data baru. Machine learning merupakan bagian penting dari ilmu komputer yang berfokus pada pembuatan algoritma dan model matematika sehingga sistem dapat belajar dari data tanpa pemrograman langsung [16]. Metode ini memanfaatkan statistika dan analisis data untuk menemukan pola kompleks, meningkatkan akurasi prediksi, serta mendukung pengambilan keputusan berdasarkan data historis [17].

2.2 Radiosonde

Radiosonde adalah alat yang dilengkapi sensor dan dibawa ke atmosfer menggunakan balon berbahan karet yang diisi gas helium [13]. Radiosonde dapat mengukur kondisi atmosfer seperti tekanan udara, suhu, kelembapan udara, serta arah dan kecepatan angin, kemudian

mengirimkan data secara real-time ke ground station melalui frekuensi radio 433 MHz [14]. Data yang diperoleh diolah menggunakan aplikasi RAOB untuk menghasilkan indeks atmosfer, yaitu Lifted Index (LI), K-Index (KI), Showalter Index (SI), dan Total Totals Index (TT). Indeks-indeks tersebut digunakan sebagai dasar untuk memprediksi potensi cuaca, khususnya potensi hujan di masa mendatang [15].

2.3. Hujan

Hujan adalah salah satu bentuk uap air yang jatuh dari atmosfer ke permukaan bumi dalam bentuk tetesan air [22]. Peristiwa ini terjadi ketika udara hangat naik ke daratan sehingga menyebabkan pertumbuhan awan konvektif [23]. Curah hujan dipengaruhi oleh beberapa faktor, seperti garis lintang, ketinggian tempat, jarak dari laut, lokasi dan luas daratan, arah angin terhadap sumber air, bentuk permukaan wilayah, serta suhu tanah. Curah hujan menunjukkan jumlah air yang dihasilkan dari hujan pada suatu wilayah tertentu. Pengukuran curah hujan menggunakan alat yang disebut penakar hujan (rain gauge), dengan satuan milimeter (mm) [24].

2.4. Naïve Bayes

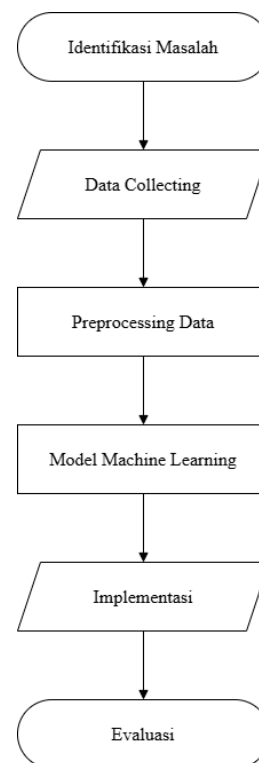
Algoritma Naïve Bayes merupakan metode klasifikasi yang didasarkan pada konsep probabilitas dan statistika. Metode tersebut bekerja dengan menghitung kemungkinan setiap fitur terhadap suatu kelas tertentu dengan asumsi bahwa setiap fitur bersifat independen dan tidak dipengaruhi oleh fitur lainnya [18]. Naïve Bayes dikenal memiliki tingkat akurasi yang tinggi serta proses komputasi yang relatif cepat, sehingga sangat efektif digunakan untuk mengklasifikasikan dan memprediksi data. Namun, salah satu kekurangan metode ini adalah jika probabilitas kondisional bernilai nol maka probabilitas prediksi juga akan bernilai nol, oleh karena itu pengecekan kualitas dataset dan penanganan missing value pada tahap pre-processing menjadi sangat penting untuk menjaga akurasi prediksi [19].

2.5. Confusion matrix

Confusion matrix merupakan alat yang berfungsi mengevaluasi kinerja model klasifikasi dalam machine learning, termasuk algoritma Naïve Bayes. Metode ini menampilkan hasil prediksi model dalam bentuk tabel yang membandingkan nilai aktual dengan nilai prediksi sehingga dapat diketahui jumlah prediksi yang benar maupun yang salah secara lebih mudah [20]. Melalui confusion matrix, berbagai metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* dapat dihitung untuk memahami performa model secara lebih baik. Terdapat empat istilah utama dalam confusion matrix, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Keempat nilai tersebut digunakan untuk menghitung metrik evaluasi.[21].

3. Metode

Penelitian ini menggunakan pendekatan machine learning dengan algoritma Naïve Bayes untuk mengidentifikasi potensi hujan berdasarkan data indeks labilitas atmosfer. Desain penelitian bersifat kuantitatif prediktif, di mana data historis digunakan untuk membangun model klasifikasi yang selanjutnya diuji akurasinya menggunakan data independen.



Gambar 1. Diagram Alur Penelitian

3.1. Pengumpulan Data (Data Collecting)

Pengumpulan data dilakukan terhadap data cuaca lapisan atas berupa empat parameter indeks labilitas atmosfer, yaitu Showalter Index, Lifted Index, Konvergence Index, dan Total Total Index, serta data pengamatan cuaca penerbangan per jam yang diperoleh dari tahun 2020 dan 2024. Pemilihan periode tersebut didasarkan pada ketersediaan data pengamatan udara atas yang konsisten dilaksanakan setiap hari pada jam 00Z dan 12Z.

3.2. Pra-Pemrosesan Data (Preprocessing)

Tahap preprocessing merupakan proses penting untuk mengubah data mentah menjadi format yang sesuai dan efisien untuk analisis [25]. Tahap ini meliputi empat langkah, yaitu:

1. Persiapan data, di mana data diolah menggunakan aplikasi RAOB untuk memperoleh nilai indeks labilitas, kemudian digabungkan dengan data METAR dan disimpan dalam format .xlsx;
2. Penanganan missing value, yaitu pengecekan dan penghapusan data yang hilang akibat kegagalan sensor atau balon yang pecah sebelum mencapai ketinggian yang cukup;
3. Seleksi fitur, di mana dari seluruh keluaran RAOB hanya dipilih empat indeks labilitas atmosfer (Showalter Index, Konvergence Index, Lifted Index, dan Total Total Index) beserta laporan sandi METAR per setengah jam; serta
4. Pembagian training set dan testing set, yang dilakukan berdasarkan persentase jumlah data menggunakan fitur dari framework Python untuk mengoptimalkan akurasi model.

3.3. Pembangunan Model Machine Learning

Model machine learning dikembangkan dengan memanfaatkan algoritma Naïve Bayes menggunakan bahasa pemrograman Python dan beberapa pustaka pendukung di platform Google Colab. Naïve Bayes adalah teknik klasifikasi berbasis statistik yang berlandaskan pada Teorema Bayes yang digunakan untuk memprediksi kemungkinan suatu kategori berdasarkan data yang sudah ada. Metode ini memiliki kemampuan klasifikasi yang sebanding dengan pohon keputusan dan jaringan syaraf buatan, serta telah terbukti memberikan tingkat akurasi yang tinggi dan kecepatan pemrosesan yang cukup cepat pada basis data skala besar.

3.4. Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan confusion matrix, yang membandingkan nilai prediksi model dengan nilai aktual. Dari confusion matrix diperoleh empat komponen, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Keempat komponen tersebut digunakan untuk menghitung metrik evaluasi berupa Accuracy, Precision, Recall, dan F1-Score. Semakin tinggi nilai metrik tersebut, semakin baik pula kinerja model yang dihasilkan [21].

4. Hasil dan Pembahasan

Pada penelitian ini menggunakan data parameter cuaca lapisan atas yang diperoleh dari pengamatan balon cuaca atau radiosonde, serta data cuaca penerbangan dalam bentuk sandi yang diamati setiap setengah jam atau dikenal sebagai data METAR. Rentang waktu penelitian mencakup periode tahun 2020 sampai dengan tahun 2024. Data yang digunakan berupa kumpulan data labilitas atmosfer berjumlah 1.057 data yang terdiri atas sembilan variabel atribut, yaitu tahun, bulan, tanggal, jam, SI, KI, LI, TT, dan RA. Selain itu, penelitian ini juga menggunakan 1.228 data cuaca penerbangan yang tersusun atas lima variabel atribut, yaitu tahun, bulan, tanggal, jam, dan kondisi cuaca hujan pada saat pengamatan dilakukan.

Berdasarkan data tersebut, kedua jenis data kemudian digabungkan untuk membentuk dataset yang sesuai dengan kebutuhan penelitian. Data yang digunakan dari hasil pengamatan radiosonde meliputi variabel SI (Showalter Index), KI (K Index), LI (Lifted Index), dan TT (Total Totals Index). Keempat atribut tersebut kemudian digabungkan dengan data cuaca penerbangan yang telah dipilih sebelumnya. Adapun contoh kondisi hujan yang diperoleh dari data METAR disajikan pada Tabel 1 berikut.

Tabel 1. Contoh data kondisi hujan dari pengamatan METAR

Tanggal	Waktu Pengamatan	Hujan (mm)	RA
02/03/2026	15.00 WIB	13	HUJAN
03/03/2026	13.00 WIB	0	TIDAK

Data kondisi hujan selama sembilan jam dilihat dari pengamatan mulai pukul 07.00 WIB sampai dengan 16.00 WIB. Kondisi hujan yang diperoleh dari pelaporan cuaca penerbangan tersebut disederhanakan agar dapat dengan mudah dibaca dan memudahkan dalam tahap pengolahan data selanjutnya.

Setelah data hujan diperoleh, tahap selanjutnya adalah pengambilan data dari aplikasi RAOB, yaitu data indeks labilitas atmosfer yang merupakan output dari aplikasi tersebut. Data yang digunakan dipilih sesuai kebutuhan penelitian, kemudian digabungkan ke dalam satu dataset. Adapun hasil indeks labilitas atmosfer dari pengolahan data radiosonde menggunakan aplikasi RAOB disajikan pada Tabel 2.

Tabel 2. Data labilitas atmosfer hasil pengoperasian Radiosonde

Tanggal	SI	KI	LI	TT
01/01/2020	-2,65	39,6	-3,66	46,6
02/01/2020	0,64	34,6	-1,46	42
03/01/2020	0,43	34,6	-0,68	41
...	...	...	...	...
30/12/2024	-1,59	39,8	-4,58	44,2
31/12/2024	-1,38	36,2	-3,23	43,4

Dari dua data yang telah dikumpulkan tersebut, selanjutnya dilakukan penggabungan menjadi satu dataset baru yang akan digunakan pada tahap berikutnya. Dataset hasil penggabungan disajikan pada Tabel 3.

Tabel 3. Dataset gabungan yang digunakan dalam penelitian

Tanggal	SI	KI	LI	TT	HUJAN
01/01/2020	-2,65	39,6	-3,66	46,6	YA
02/01/2020	0,64	34,6	-1,46	42	YA
03/01/2020	0,43	34,6	-0,68	41	YA
...	...	...	...	...	...
31/12/2024	-1,38	36,2	-3,23	43,4	TIDAK

Dataset di atas terdiri dari enam kolom, yaitu Tanggal, SI, KI, LI, TT, dan Hujan. Kolom Hujan berisi informasi mengenai kejadian hujan atau tidak hujan dalam rentang waktu sembilan jam, yaitu mulai pukul 07.00 hingga 16.00 WIB.

4.1 . Preprocessing Data

Setelah dataset terbentuk dari penggabungan kedua jenis data tersebut, langkah selanjutnya adalah melakukan pengecekan terhadap kelengkapan data. Proses ini bertujuan untuk memastikan bahwa tidak terdapat data yang kosong atau hilang, karena kelengkapan data sangat berpengaruh terhadap kualitas dataset yang akan digunakan pada tahap berikutnya [30]. Data yang kosong atau tidak tersedia perlu melalui proses pembersihan (cleaning) agar data yang digunakan dalam pemodelan bersifat lengkap dan valid. Oleh karena itu, tahap preprocessing dilakukan untuk menghasilkan kumpulan data yang lebih optimal sehingga dapat dimanfaatkan secara maksimal dalam proses penelitian selanjutnya.

4.1.1 Persiapan Data

Data yang disiapkan adalah data pengamatan METAR dan data pengamatan cuaca udara atas menggunakan radiosonde selama tahun 2020 sampai dengan 2024. Dataset ini memiliki delapan kolom yang terdiri dari "Tahun", "Bulan", "Tanggal", "Jam", "SI", "LI", "KI", "TT", dan "RA". Kolom "RA" merupakan singkatan dari Rain yang menunjukkan ada tidaknya kejadian hujan yang teramati pada 07.00 WIB sampai dengan 16.00 WIB pada tanggal tersebut.

4.1.2 Penanganan Missing Values

Dalam penelitian ini, untuk mengatasi missing values dilakukan dengan menghapus data yang kosong melalui perintah "df.dropna". Berdasarkan hasil pengecekan, ditemukan missing values pada variabel SI sebanyak 5 data dan variabel TT sebanyak 1 data, sementara variabel KI dan LI tidak memiliki data yang hilang. Setelah proses penghapusan data kosong dilakukan, dari total 1.057 data awal, dataset yang digunakan dalam membangun model menjadi 1.053 data.

4.1.3 Seleksi Fitur

Seleksi fitur tahap preprocessing yang digunakan untuk menghilangkan data yang tidak memiliki korelasi[26]. Dari sembilan kolom variabel yang tersedia pada dataset, variabel yang dipilih untuk membangun pemodelan dalam penelitian ini adalah SI, KI, LI, TT, dan HUJAN. Dari kelima variabel tersebut, empat variabel berisi data numerik bertipe float dan satu variabel, yaitu variabel HUJAN, berisi data kategori bertipe object.

4.1.4 Pembagian Data Latih dan Data Uji

Langkah selanjutnya adalah membagi dataset menjadi dua bagian, yaitu variabel x dan variabel y. Variabel x merupakan atribut independent yang berisi nilai labilitas atmosfer pada kolom SI, KI, LI, dan TT. Variabel y (target) merupakan data kejadian hujan dan tidak hujan, yang dikodekan dengan nilai "1" untuk ada hujan dan nilai "0" untuk tidak ada hujan.

4.2 . Implementasi Model

Model yang telah dibangun dapat digunakan untuk memprediksi kejadian hujan dengan memasukkan data indeks labilitas atmosfer yang diperoleh dari hasil pengamatan radiosonde. Sebagai contoh implementasi, digunakan data hasil pengolahan RAOB untuk pengamatan tanggal 02 Maret 2026. Berdasarkan hasil pengolahan, didapatkan nilai indeks labilitas atmosfer sebagai berikut: SI sebesar -1,1; KI sebesar 35,7; LI sebesar -1,2; dan TT sebesar 41,7. Nilai-nilai tersebut kemudian dimasukkan ke dalam model prediksi yang telah dibangun, sebagaimana ditunjukkan pada Gambar 2.

```
import pandas as pd
#COBA TES HUJAN ATAU TIDAK

# Masukkan nilai variabel LI, SI, KI, TT untuk prediksi
# Contoh: SI = -1.1, LI = 1.2, KI = 35.7, TT = 41.7
# Anda bisa mengubah nilai-nilai ini
data_input_new = pd.DataFrame([
    {
        'SI': -1.1,
        'LI': 1.2,
        'KI': 35.7,
        'TT': 41.7
    }
], columns=['SI', 'LI', 'KI', 'TT'])

# Melakukan prediksi menggunakan model nbtrain_smote
prediksi_smote = nbtrain_smote.predict(data_input_new)

print(f"Input variabel: SI={data_input_new['SI'].iloc[0]}, LI={data_input_new['LI'].iloc[0]}, KI={
    if prediksi_smote[0] == 1:
        print("Hasil prediksi: (1) Akan terjadi hujan.")
    else:
        print("Hasil prediksi: (0) Tidak akan terjadi hujan.")

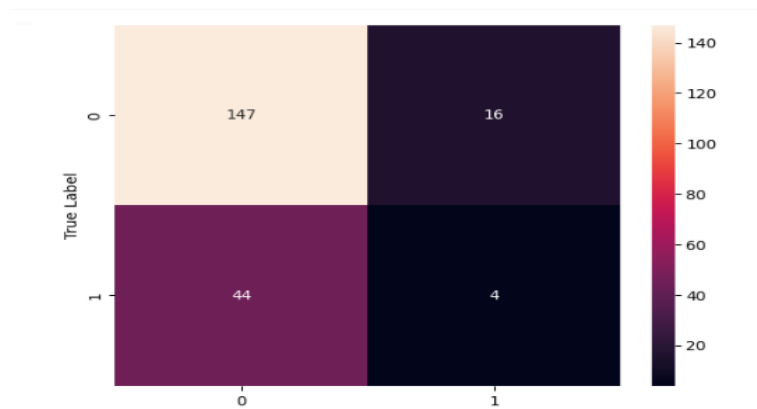
--- Input variabel: SI=-1.1, LI=1.2, KI=35.7, TT=41.7
Hasil prediksi: (1) Akan terjadi hujan.
```

Gambar 2. Alur prediksi dengan algoritma Naïve Bayes

Dari hasil prediksi, diperoleh nilai prediksi hujan sebesar 1, yang berarti dalam rentang waktu tersebut berpotensi terjadi hujan. Hasil prediksi ini kemudian diverifikasi dengan data pengamatan cuaca penerbangan (METAR) aktual yang menginformasikan bahwa pada kurun waktu yang dimaksud memang terjadi hujan, khususnya pada tanggal 02 Maret 2026 pukul 15.00 WIB hingga 16.00 WIB. Hal ini menunjukkan bahwa model yang dibangun mampu memprediksi kejadian hujan dengan tepat untuk kasus tersebut.

4.3 . Evaluasi Model.

Evaluasi model dilakukan dengan mengukur performa hasil prediksi menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score. Semakin tinggi nilai precision, recall, dan F1-score, maka semakin baik performa model machine learning yang dihasilkan. Perlu dicatat bahwa pada kasus data tidak seimbang (imbalanced data), nilai accuracy yang tinggi belum tentu menunjukkan performa model yang baik, karena model cenderung memprediksi kelas mayoritas. Oleh karena itu, precision, recall, dan F1-score perlu diperhatikan agar hasil evaluasi model menjadi lebih akurat dan komprehensif. Perhitungan metrik evaluasi tersebut dilakukan menggunakan confusion matrix, yang ditunjukkan pada Gambar 3.



Gambar 3. Confusion matrix algoritma Naïve Bayes

Berdasarkan *confusion matrix* tersebut, diperoleh nilai TP sebesar 42, FP sebesar 88, FN sebesar 6, dan TN sebesar 75. Dari nilai-nilai tersebut, perhitungan metrik evaluasi dilakukan menggunakan rumus berikut:

$$\text{precision} = \frac{TP}{(TP + FP)} \quad (1)$$

$$\text{recall} = \frac{TP}{(TP + FN)} \quad (2)$$

$$f1score = \frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (3)$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (4)$$

Hasil perhitungan metrik evaluasi menggunakan algoritma Naïve Bayes menghasilkan nilai accuracy sebesar 71% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Namun demikian, nilai precision dan F1-score yang relatif rendah mengindikasikan adanya ketidakseimbangan data (*imbalanced data*) antara kelas "hujan" dan "tidak hujan" pada dataset yang digunakan.

5. Perbandingan

Penelitian ini, algoritma Naïve Bayes digunakan sehingga menghasilkan nilai accuracy sebesar 71%, precision sebesar 20%, dan F1-score sebesar 11% pada dataset yang digunakan. Hasil ini sejalan dengan penelitian Rizqi dan Kusumaningsih [9], yang menunjukkan bahwa algoritma Naïve Bayes mampu memberikan performa klasifikasi yang baik pada data curah hujan, tetapi sangat dipengaruhi oleh karakteristik dan distribusi data. Selain itu, tinjauan sistematis oleh Zulfikri dkk. [10] juga menyatakan bahwa performa Naïve Bayes pada dataset cuaca bergantung pada kualitas data dan proses preprocessing. Dibandingkan dengan penelitian Hermansyah dkk. [8] yang menggunakan metode Stacking Ensemble, penelitian ini menerapkan algoritma yang lebih sederhana sehingga proses komputasi menjadi lebih cepat dan mudah diimplementasikan, meskipun akurasi yang diperoleh masih dapat ditingkatkan. Oleh karena itu, untuk meningkatkan hasil model, dapat dilakukan pengembangan lebih lanjut dengan menerapkan teknik penyeimbangan data, seperti oversampling atau undersampling, agar nilai precision dan recall meningkat.

6. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa penelitian ini berhasil membangun sistem prediksi potensi hujan berbasis data radiosonde menggunakan algoritma Naïve Bayes sebagai metode pemodelan. Model yang dihasilkan mampu memprediksi kejadian hujan dengan tingkat akurasi sebesar 71,56% berdasarkan dataset yang digunakan dalam penelitian. Dalam proses pengolahan data, ditemukan adanya missing value pada dataset hasil pengolahan radiosonde. Oleh karena itu, dilakukan penanganan missing value dengan metode pengabaian data kosong, yaitu menghapus baris data yang memiliki nilai kosong. Metode tersebut dinilai cukup efektif karena mampu menghasilkan dataset yang lebih baik dan lebih siap digunakan dalam proses pemodelan serta prediksi potensi hujan. Untuk meningkatkan performa model, jumlah dataset perlu ditambah agar data latih dan data testing menjadi lebih banyak sehingga variasi pembagian data dapat dilakukan dengan lebih optimal dan berpotensi menghasilkan akurasi yang lebih tinggi. Selain itu, atribut prediktor yang digunakan juga dapat dikembangkan dengan menambahkan variabel lain selain SI, KI, LI, dan TT sehingga hubungan antara tribut dapat dianalisis secara lebih mendalam. Dalam penanganan missing value, penelitian selanjutnya juga dapat mempertimbangkan metode lain, seperti imputasi, yaitu memberikan nilai pengganti pada data yang kosong dengan menggunakan perhitungan mean, median, atau mode agar kualitas dataset menjadi lebih baik untuk proses pemodelan.

Referensi

- [1] M. Agil Septiyana Putra and Hodi, "Upaya Unit PKP-PK Dalam Meningkatkan Kualitas Sumber Daya Manusia Di Bandar Udara Tunggul Wulung Cilacap," *J. Kaji. dan Penelit. Umum*, vol. 1, no. 4, pp. 26–36, 2023, doi: 10.47861/jkpu-nalanda.v1i4.352.
- [2] S. H. Umar and Diah Anggraeni, "Pengaruh Safety Culture Terhadap Keselamatan Penerbangan Di Bandar Udara Tunggul Wulung Cilacap," *Optimal*, vol. 17, no. 1, pp. 105–127, 2020.
- [3] A. Q. Islam, D. Anggraini, F. A. Ningrum, S. S. Prayuda, and M. K. Milad, "Peran Komunikasi Interpersonal dalam Penyampaian Informasi Cuaca dan Peringatan Dini oleh BMKG Juanda: Studi Kasus pada Layanan Penerbangan dan Publik," *J. Penelit. Multidisiplin Bangsa*, vol. 1, no. 12, pp. 2150–2164, 2025, doi: 10.59837/jpnmb.v1i12.426.
- [4] W. A. Iskandar, "Sistem Monitoring Automated Weather Observing Sistem (AWOS) Berbasis Android Studi Kasus BMKG Sultan Mahmud Badauruddin II Palembang," *SMATIKA*, vol. 11, pp. 1–6, 2020.
- [5] F. Sabur, A. Bahrawi, and M. Agung Rahardjo, "Analisis Pengaruh Instrument Landing System (ILS) untuk Peningkatan Pelayanan Keselamatan di Bandar Udara Haluleo Kendari," *Airman J. Tek. dan Keselam. Transp.*, vol. 3, no. 1, pp. 78–87, 2020, doi: 10.46509/ajtk.v3i1.159.
- [6] W. A. Iskandar, L. Lindawati, and M. Fadhli, "Sistem Monitoring Automated Weather Observing System (AWOS) Berbasis Android Studi Kasus BMKG Sultan Mahmud Badaruddin II Palembang," *Smatika J.*, vol. 11, no. 02, pp. 101–112, 2021, doi: 10.32664/smatika.v11i02.595.
- [7] A. Y. Pratama et al., "Optimalisasi Awos (Automated Weather Observing System) Dalam Pemberian Informasi Penerbangan Di Bandar Udara Gt. Sjamsir Alam Kotabaru," pp. 1–7, 2020.
- [8] M. Hermansyah, A. Saikhu, and B. Amaliah, "Pemodelan data radiosonde menggunakan stacking ensemble untuk klasifikasi hujan," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 2, pp. 1678–1687, 2025, doi: 10.29100/jupi.v10i2.7806.
- [9] A. A. Rizqi and D. Kusumaningsih, "Klasifikasi Curah Hujan di Kota Bogor Provinsi Jawa Barat dengan Menggunakan Metode Naive Bayes," *J. Semin. Nas. Mhs. Fak. Teknol. Inf.*, no. September, pp. 542–550, 2022.
- [10] A. Zulfikri, G. Gunawan, and W. Andriani, "Tinjauan Pustaka Sistematis: Penerapan Metode Naives Bayes untuk Klasifikasi dalam Dataset Cuaca," *J. Pract. Comput. Sci.*, vol. 2, no. 2, pp. 90–95, 2023, doi: 10.37366/jpcs.v2i2.2101.
- [11] N. A. Prakoso Indaryono, "Analisa Perbandingan Algoritma Random Forest Dan Naive Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 1, pp. 158–167, 2024, doi: 10.29100/jupi.v9i1.4421.
- [12] R. Amalia, "Penerapan Data Mining Untuk Memprediksi Hasil Kelulusan Siswa," *J. Inform. dan Sist. Inf.*, vol. 6, no. 1, pp. 33–42, 2020.
- [13] R. A. Pratama, S. T. M. T. M. E. P. D. Zainul Abidin, and S. T. M. T. M. S. Dr-Ing. Onny Setyawa, "Implementasi Internet of Things (IoT) Pada Rancang Bangun Payload Radiosonde," 2021. [Online]. Available: <http://repository.ub.ac.id/id/eprint/186903/>
- [14] Apriadi and H. Suroyo, "Analisa Data Radiosonde Dengan Menggunakan Bahasa Pemrograman Phytion," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 2, pp. 1002–1010, 2023, doi: 10.30865/klik.v4i2.1278.
- [15] E. Fibriantika and D. Mayangwulan, "Analisis Spasial Indeks Stabilitas Udara di Indonesia Spatial Analysis of Air Stability Index in Indonesia," vol. 21, no. 1, pp. 1–12, 2020.
- [16] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [17] I. Hapsari and S. Pandya Wisesa, "Evaluasi Model Prediksi Curah Hujan Berbasis Machine Learning di Kota Bandung," *J. Nas. Teknol. dan Sist. Inf.*, vol. 11, no. 2, pp. 136–143, 2025, doi: 10.25077/teknosi.v11i2.2025.136-143.
- [18] Y. I. Kurniawan, F. Razi, Nofiyati, B. Wijayanto, and M. L. Hidayat, "Naive bayes modification for intrusion detection system classification with zero probability," *Bull. Electr. Eng. Informatics*, vol. 10, no. 5, pp. 2751–2758, 2021, doi: 10.11591/eei.v10i5.2833.
- [19] B. K. Khotimah, M. Miswanto, and H. Suprajitno, "Optimization of feature selection using genetic algorithm in naïve Bayes classification for incomplete data," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 1, pp. 334–343, 2020, doi: 10.22266/ijies2020.0229.31.
- [20] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [21] L. Firdaus and T. Setiadi, "Perbandingan Algoritma Naive Bayes, Decision Tree, dan KNN untuk Klasifikasi Produk Populer Adidas US dengan Confusion Matrix," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 185, 2023, doi: 10.30865/json.v5i2.6124.
- [22] S. Agustiarini et al., "Analisis Anomali Curah Hujan Periode Musim Hujan Saat Terjadi La-Nina di NTB," vol. 2, no. 2, pp. 11–17, 2022. [Online]. Available: <https://www.chc.ucsb.edu/data/chirps>
- [23] A. A. Azani and N. Kusumawardani, "Kajian Indeks Stabilitas Atmosfer Terhadap Kejadian Hujan Lebat Di Kota Bitung (Studi Kasus Tahun 2020-2021)," *J. Widya Climago*, vol. 4, no. 1, pp. 29–36, 2022.
- [24] I. Mahmuda, W. Ariesta, P. C. Sinteks, and D. I. K. Malang, "Prediksi Curah Hujan Bulanan Menggunakan Metodedi Kota Malang," vol. 13, no. 1, pp. 35–42, 2024.
- [25] D. D. N. Cahyo and A. Sunyoto, "Analisis Perbandingan Klasifikasi dalam Data Mining pada Prediksi Hujan dengan menggunakan Algoritma LSTM dan GRU," *J. Sains dan Inform.*, vol. 11, no. 1, pp. 40–49, 2025, doi: 10.34128/jsi.v11i1.1212.
- [26] S. Clara, D. Laksmi Prianto, R. Al Habsi, E. Friscila Lumbantobing, and N. Chamidah, "Feature Selection Implementation in Machine Learning Classification Algorithms to Predict Income on Adult Income Dataset," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, vol. 2, no. 1, pp. 741–747, 2021.