

## Penerapan Algoritma Random Forest Dalam Klasifikasi Prakiraan Cuaca

Deny Saputra<sup>1</sup>, Menur Wahyu Pangestika<sup>2</sup>, Barry Ceasar Octariadi<sup>3</sup>

Email: [1ddeny.s161203@gmail.com](mailto:1ddeny.s161203@gmail.com), [2menur.wahyu@unmuhpnk.ac.id](mailto:2menur.wahyu@unmuhpnk.ac.id), [3barry.ceasaro@unmuhpnk.ac.id](mailto:3barry.ceasaro@unmuhpnk.ac.id)

<sup>123</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Pontianak

Diterima: 26 Desember 2025 | Direvisi: - | Disetujui: 31 Desember 2025

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,  
Universitas Muhammadiyah Riau, Indonesia

### Abstrak

Cuaca memiliki peranan penting dalam berbagai aspek kehidupan seperti pertanian dan transportasi. Namun prediksi cuaca masih menjadi tantangan karena dipengaruhi banyak faktor yang kompleks. Cuaca ekstrem seperti badai dan banjir dapat menimbulkan kerugian besar, sehingga diperlukan sistem klasifikasi prakiraan cuaca yang akurat. Penelitian ini menerapkan algoritma *Random Forest* untuk meningkatkan akurasi prediksi dan dioptimasi menggunakan *Grid Search Cross Validation* dan di evaluasi menggunakan *confusion matrix*. Metode yang digunakan adalah *CRISP-DM* dengan enam tahapan utama. Data diperoleh dari (BMKG), mencakup fitur seperti suhu, kelembapan, kecepatan angin, tutupan awan, jarak pandang, dan arah angin, dengan label Kondisi Cuaca dan Nama Daerah sebagai penanda kategori cuaca dan lokasi wilayah yang diklasifikasikan. Evaluasi akhir menggunakan *confusion matrix* dengan hasil akurasi 98,84% pada data latih dan 95,33% pada data uji, menandakan performa stabil dan kemampuan generalisasi yang baik.

**Kata kunci:** *Random Forest*, Klasifikasi Cuaca, *CRISP-DM*

### *Application of the Random Forest Algorithm in Weather Forecast Classification*

#### *Abstract*

*Weather plays an important role in various aspects of life, such as agriculture and transportation. However, weather prediction remains challenging because it is influenced by many complex factors. Extreme weather events, such as storms and floods, can cause significant losses, making accurate weather forecast classification systems essential. This study applies the Random Forest algorithm to improve prediction accuracy and optimizes it using Grid Search Cross Validation. The method used is CRISP-DM, consisting of six main stages. The data were obtained from the Meteorological, Climatological, and Geophysical Agency (BMKG), containing features such as temperature, humidity, wind speed, cloud cover, visibility, and wind direction, with the labels Weather Condition and Region Name serving as indicators of the classified weather category and location. The final evaluation uses a confusion matrix, yielding an accuracy of 98.84% on the training data and 95.33% on the testing data, demonstrating stable performance and strong generalization capability.*

**Keywords:** *Random Forest, Weather Classification, CRISP-DM*

## 1. PENDAHULUAN

Cuaca memiliki peran penting dalam kehidupan manusia, terutama dalam sektor pertanian, penerbangan, transportasi, mitigasi bencana dan berbagai bidang lainnya. Namun, prediksi cuaca menjadi tantangan karena dipengaruhi oleh banyak faktor yang kompleks dan saling berinteraksi[1]. Salah satu tantangan utama dalam prediksi cuaca adalah cuaca ekstrem, yaitu peristiwa cuaca yang tidak biasa dan tidak lazim yang dapat menyebabkan dampak signifikan, termasuk kerugian ekonomi dan korban jiwa[2]. Pusat Informasi Lingkungan AS melaporkan bahwa kerusakan cuaca mencapai \$7,9 miliar pada tahun 2015. Di sisi lain, kemampuan prakiraan surya menawarkan peluang untuk menghemat \$455 juta pada tahun 2040 dalam penggunaan sumber

energi terbarukan[3]. Bahkan, dalam 20 tahun terakhir, bencana alam terbesar kedua di Asia Pasifik dengan 191.164 korban jiwa, terjadi di Indonesia. Laporan tersebut menyatakan bahwa Indonesia mengalami kerugian ekonomi sebesar US\$22,5 miliar[4]. Oleh karena itu, diperlukan Metode klasifikasi prakiraan cuaca yang tepat untuk mengurangi risiko dan memaksimalkan peluang, seperti pengelolaan sumber daya, dan mengurangi dampak negatif di berbagai bidang.

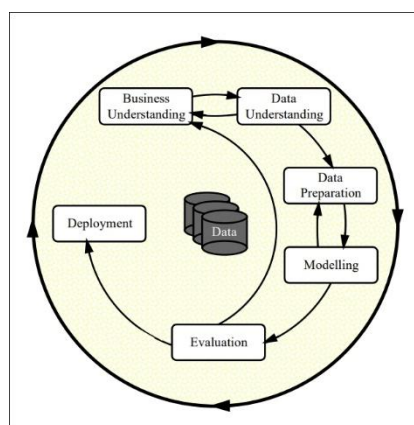
Penelitian terdahulu menunjukkan bahwa *Machine Learning* (ML), khususnya algoritma *Random Forest* terbukti signifikan dalam meningkatkan akurasi prakiraan cuaca. Sebagai contoh, penelitian oleh Risanti dkk, yang melakukan analisis model prediksi cuaca menggunakan beberapa algoritma termasuk *Random Forest*, *Support Vector Machine*, *Gradient Boosting*, dan *Decision Tree*, dengan menggunakan parameter seperti suhu, kelembapan, kecepatan angin, dan kode cuaca. Hasil akhir menunjukkan bahwa algoritma *Random Forest* mendapatkan akurasi tertinggi dengan akurasi 85% dan presisi 86%[5]. Tidak hanya analisis prediksi cuaca saja algoritma *Random forest* juga terbukti baik dalam melakukan klasifikasi deteksi penyakit jantung, dari penelitian Tengku dkk, dengan menggunakan 13 fitur seperti tekanan darah, kolesterol, detak jantung maksimum, dan lain-lain, serta satu atribut target bertingkat, mendapatkan nilai akuarasi 0,89%[6]. Dapat disimpulkan bahwa *Random Forest* dapat menyelesaikan permasalahan yang ada pada masalah lingkungan seperti prakiraan cuaca.

Konsep dasar *Random Forest* adalah membangun sekumpulan pohon keputusan, yang masing-masing dilatih menggunakan data dan atribut yang dipilih secara acak. Setiap pohon berfungsi sebagai model unik yang menghasilkan hasil prediksi yang unik dalam bentuk label kelas dan nilai numerik[7]. Selain itu, analisis *Random Forest* telah diakui sebagai alat yang efektif untuk peramalan cuaca[8]. Oleh karena itu, metode ini dianggap sesuai untuk digunakan dalam penelitian ini guna meningkatkan akurasi dalam klasifikasi dan prediksi prakiraan cuaca.

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi prakiraan cuaca berbasis parameter meteorologis serta menganalisis kinerja algoritma *Random Forest* dalam menghasilkan tingkat akurasi klasifikasi yang optimal. Berbeda dengan penelitian sebelumnya, penelitian ini menambahkan fitur identitas wilayah berupa nama daerah guna meningkatkan ketepatan identifikasi lokasi pada hasil klasifikasi. Penambahan fitur tersebut didasarkan pada pertimbangan bahwa setiap daerah memiliki karakteristik geografis dan dinamika cuaca yang berbeda, sehingga informasi lokasi menjadi elemen penting dalam menghasilkan klasifikasi yang lebih akurat dan kontekstual. Proses pengembangan sistem dilakukan dengan mengacu pada metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) agar setiap tahapan penelitian dilakukan secara sistematis dan terstruktur, sehingga hasil klasifikasi yang dihasilkan memiliki validitas dan keandalan yang tinggi.

## 2. METODE PENELITIAN

Penelitian ini menggunakan metode *CRISP-DM*, Metode *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) adalah kerangka kerja atau metodologi untuk data mining dan analisis data yang membantu profesional menjalankan proyek data mining secara sistematis dari tahap perencanaan hingga pelaksanaan. Metodologi ini sangat terstruktur dan banyak digunakan[9]. Penerapan metode *CRISP-DM* bertujuan untuk menyediakan kerangka analisis yang sistematis dalam memecahkan permasalahan penelitian maupun kebutuhan bisnis[10]. Oleh karena itu, metode ini dinilai sesuai untuk digunakan dalam penelitian ini, khususnya dalam menganalisis permasalahan klasifikasi cuaca serta melakukan pengolahan data secara terstruktur. Untuk mempermudah pemahaman mengenai tahapan dalam metodologi *CRISP-DM*, alur proses tersebut dapat dilihat pada Gambar 1.



Gambar 1 Tahapan Metode CRISP-DM

### 2.1. Business Understanding

Pada tahap ini dilakukan memahami permasalahan yang ada dan menetapkan tujuan apa yang harus diterapkan pada penelitian. Adapun permasalahan utama dalam penelitian ini ialah bagaimana membangun model klasifikasi prakiraan cuaca dengan memanfaatkan parameter meteorologis, dan untuk mengurangi resiko terjadinya termasuk kerugian ekonomi dan korban jiwa.

## 2.2. Data Understanding

Pada tahap ini dilakukan proses pemahaman data yang mencakup pengumpulan data serta eksplorasi awal, antara lain pemeriksaan nilai hilang (*missing values*), analisis korelasi antar fitur, deteksi outlier, identifikasi data duplikat, dan evaluasi ketidakseimbangan kelas (*data imbalance*).

## 2.3. Data Preparation

Pada tahap ini dilakukan proses persiapan atau pra-pemrosesan data, meliputi seleksi fitur, penanganan nilai hilang, penghapusan data duplikat, penghapusan data *outliers*, penyeimbangan distribusi kelas, serta transformasi data.

## 2.4. Modelling

Pemodelan dilakukan menggunakan algoritma *Random Forest* yang dioptimasi melalui *Grid Search CV* untuk mengklasifikasikan prakiraan cuaca. Dataset dibagi kembali menggunakan metode *train-test* split menjadi 80% data latih dan 20% data uji. Selanjutnya, model dilatih menggunakan data latih dan dievaluasi dengan data uji untuk menghasilkan prediksi yang kemudian dinilai menggunakan *confusion matrix*.

## 2.5. Evaluasi

Tahap evaluasi dilakukan untuk menilai sejauh mana model mampu mengklasifikasikan kondisi cuaca berdasarkan parameter cuaca. Evaluasi dilakukan menggunakan *confusion matrix* yang mencakup metrik akurasi, presisi, recall, dan F1-score sehingga memberikan gambaran mengenai performa model.

## 2.6. Deployment

Pada tahap akhir, model klasifikasi diintegrasikan ke dalam aplikasi berbasis web agar mudah diakses oleh pengguna. Model yang telah dilatih disimpan dalam format *pickle* dan dihubungkan ke sistem aplikasi untuk menghasilkan prediksi berdasarkan parameter cuaca yang dimasukkan. Dengan demikian, model tidak hanya divalidasi secara eksperimental, tetapi juga siap digunakan sebagai alat pendukung keputusan dalam prakiraan cuaca.

# 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian yang diperoleh melalui penerapan metode *CRISP-DM* dari tahap awal hingga tahap akhir. Model yang dihasilkan kemudian diimplementasikan ke dalam aplikasi berbasis web. Penyajian hasil ini bertujuan untuk menggambarkan bagaimana data yang telah diproses digunakan dalam pengembangan model klasifikasi prakiraan cuaca serta bagaimana model tersebut dapat dimanfaatkan secara langsung melalui aplikasi interaktif.

## 3.1. Business Understanding

*Business understanding* pada penelitian ini menekankan kebutuhan untuk mengembangkan model klasifikasi prakiraan cuaca yang memanfaatkan parameter meteorologis secara optimal. Tujuan utamanya adalah menghasilkan prediksi cuaca yang akurat guna meminimalkan berbagai dampak negatif, termasuk kerugian ekonomi dan risiko keselamatan. Pemahaman ini menjadi landasan dalam penentuan fokus penelitian, pemilihan metode, serta strategi pengolahan data dan pengembangan model.

## 3.2. Data Understanding

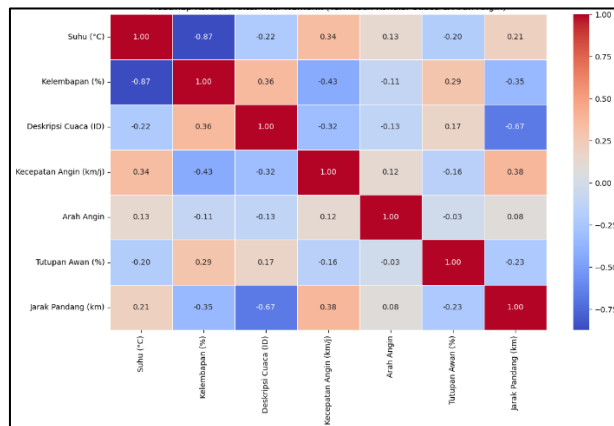
Penelitian ini menggunakan data yang diperoleh melalui dataset terbuka yang tersedia di situs resminya <https://data.bmkg.go.id/prakiraan-cuaca/>. Dataset tersebut mengandung 31.784 entri yang dicatat setiap tiga jam dari 16 Juli 2025 hingga 3 Agustus 2025, dan mencakup kondisi meteorologi di Provinsi Kalimantan Barat. Waktu (UTC dan lokal), suhu, kelembapan, jarak pandang, kecepatan angin, arah angin, tutupan awan, deskripsi cuaca dan wilayah. Dapat dilihat pada tabel 1.

Tabel 1 Sample data

Waktu Utc	Waktu Lokal	Suhu (^C)	Kelembapan (%)	Jarak Pandang (km)	Kecepatan Angin (km/j)	Arah Angin	Tutupan Awan (%)	Deskripsi Cuaca	Wilayah
7/16/2025 17:00	7/17/2025 0:00	25	96	> 10 km	6.6	SE	100	Berawan	Desa Segarau Parit
7/16/2025 20:00	7/17/2025 3:00	25	96	> 10 km	5.7	S	100	Berawan	Desa Sungai Raya
7/16/2025 23:00	7/17/2025 6:00	26	93	> 10 km	8.1	SE	50	Cerah	Desa Kapur
7/17/2025 2:00	7/17/2025 9:00	30	70	> 10 km	13.4	SE	15	Cerah	Desa Mekar Sari

7/17/2025 5:00	7/17/2025 12:00	32	64	> 10 km	7.6	S	33	Cerah	Desa Parit Baru
-------------------	--------------------	----	----	---------	-----	---	----	-------	-----------------

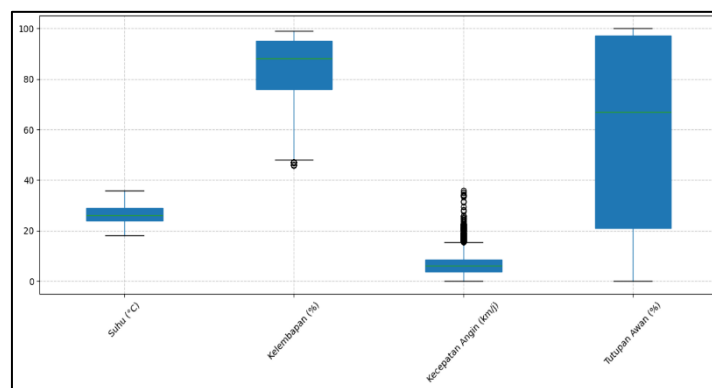
Selanjutnya dilakukan pemeriksaan *missing values* dan tidak ditemukan adanya data yang hilang. Proses dilanjutkan dengan pengecekan data duplikat, di mana teridentifikasi 97 data duplikat yang harus dihapus. Tahap berikutnya adalah analisis korelasi antar fitur numerik, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2 Korelasi antar feature numerik

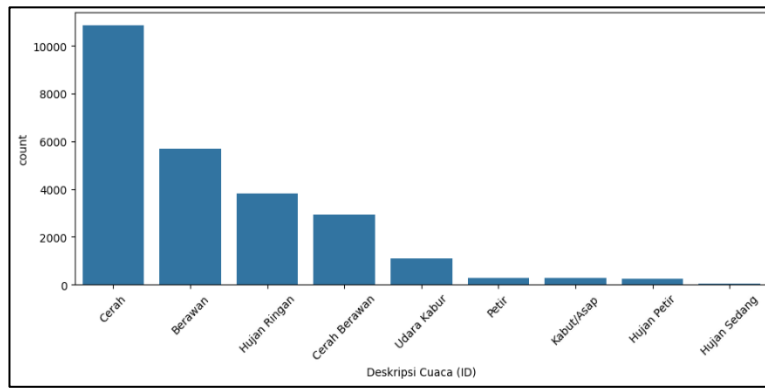
Analisis korelasi pada gambar 2 menunjukkan bahwa Deskripsi Cuaca (ID) berhubungan negatif dengan Suhu (-0.22) dan Jarak Pandang (-0.67), serta positif dengan Kelembapan (0.36), menandakan bahwa cuaca buruk cenderung terjadi saat suhu rendah, kelembapan tinggi, dan jarak pandang pendek. Suhu memiliki hubungan negatif kuat dengan Kelembapan (-0.87). Kecepatan Angin berhubungan positif dengan Suhu (0.34) dan Jarak Pandang (0.38), serta negatif dengan Kelembapan (-0.43). Arah Angin dan Tutupan Awan hanya menunjukkan hubungan lemah terhadap variabel lain. Secara keseluruhan, suhu, kelembapan, dan jarak pandang merupakan fitur yang paling berpengaruh.

Tahap berikutnya dilanjutkan dengan deteksi *outliers* dapat dilihat pada gambar 3 dibawah ini :



Gambar 3 boxplot keempat fitur

Pada Gambar 3 teridentifikasi keberadaan *outliers* pada fitur kelembapan dan kecepatan angin, yang berpotensi menurunkan kinerja model. Oleh karena itu, nilai-nilai tersebut perlu dihapus agar data yang digunakan dalam pelatihan model menjadi lebih bersih dan menghasilkan prediksi yang lebih akurat. Tahap selanjutnya adalah analisis ketidakseimbangan data (*data imbalance*), yang ditampilkan pada Gambar 4.



Gambar 4 Distribusi Data Imbalance

Gambar 4 menunjukkan distribusi label pada fitur Cuaca sebelum pembagian data dan penerapan *oversampling*. Terlihat adanya ketidakseimbangan kelas, di mana kelas “Cerah Berawan” dan “Berawan” mendominasi, sedangkan kelas seperti “Hujan Petir” dan “Hujan Sedang” memiliki jumlah yang jauh lebih sedikit. Ketimpangan ini berpotensi menimbulkan bias terhadap kelas mayoritas, sehingga diperlukan teknik *oversampling* untuk menyeimbangkan distribusi kelas dan meningkatkan kualitas prediksi model.

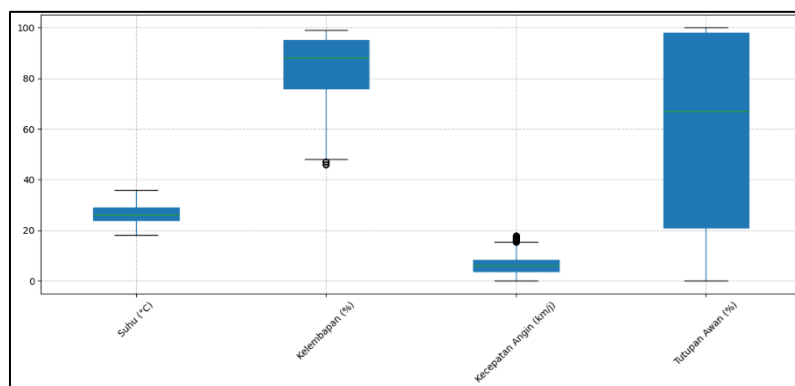
### 3.3. Data Preparation

Tahap awal dalam proses persiapan data adalah melakukan seleksi fitur untuk menentukan variabel yang relevan digunakan dalam analisis. Fitur yang dipertahankan meliputi Suhu, Kelembapan, Kecepatan Angin, Arah Angin, Tutupan Awan, Jarak Pandang, Deskripsi Cuaca (ID), serta Wilayah yang selanjutnya dijadikan sebagai label atau kelas pada proses klasifikasi. Perbandingan antara data asli dengan data setelah melalui tahap seleksi fitur ditampilkan pada Tabel 1 di atas dan Tabel 2 setelah dilakukan seleksi fitur, berikut.

Tabel 2 Dataset Setelah Seleksi Fitur

Suhu (^C)	Kelembapan (%)	Jarak Pandang (km)	Kecepatan Angin (km/j)	Arah Angin	Tutupan Awan (%)	Deskripsi Cuaca	Wilayah
25	96	> 10 km	6.6	SE	100	Berawan	Desa Segarau Parit
25	96	> 10 km	5.7	S	100	Berawan	Desa Sungai Raya
26	93	> 10 km	8.1	SE	50	Cerah	Desa Kapur
30	70	> 10 km	13.4	SE	15	Cerah	Desa Mekar Sari
32	64	> 10 km	7.6	S	33	Cerah	Desa Parit Baru

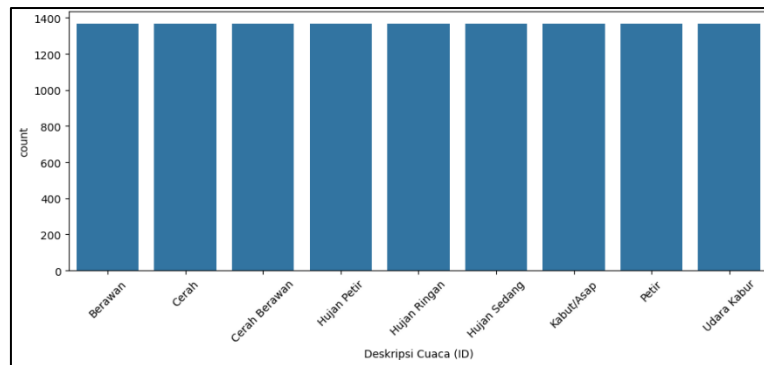
Tahap selanjutnya adalah pemeriksaan dan penanganan data yang memiliki nilai hilang (*missing values*). Berdasarkan hasil pengecekan, seluruh data tidak mengandung nilai yang hilang, sehingga proses dapat dilanjutkan ke tahap berikutnya yaitu Penghapusan nilai *outlier* dilakukan berdasarkan hasil identifikasi yang telah ditunjukkan pada Gambar 3. Setelah proses tersebut, data yang telah bebas dari outlier dapat dilihat pada Gambar 5 berikut.



Gambar 5 Setelah penghapusan nilai outliers

Gambar 5 menampilkan kondisi data setelah proses penghapusan outlier. Pada Gambar 3 masih terlihat banyak titik di luar rentang distribusi normal, maka setelah pembersihan jumlah outlier berkurang signifikan, sehingga data menjadi lebih rapi dan layak digunakan pada tahap pemodelan.

Tahap selanjutnya adalah penanganan ketidakseimbangan label pada dataset, sebagaimana terlihat pada Gambar 4 sebelumnya. Proses penyeimbangan dilakukan menggunakan metode *SMOTE* (*Synthetic Minority Over-sampling Technique*) hingga seluruh kelas memiliki proporsi yang seimbang. Hasil penyeimbangan tersebut ditampilkan pada Gambar 6.



Gambar 6 Distribusi Label Setelah Penyeimbangan Kelas

Gambar 6 menampilkan distribusi label fitur Cuaca setelah penyeimbangan dengan *SMOTE*. Seluruh kelas memiliki jumlah data yang setara, sehingga proses *oversampling* ini membantu mengurangi bias model dan meningkatkan akurasi pada tahap pelatihan. Setelah proses penyeimbangan data selesai, Tahap berikutnya adalah transformasi data, yaitu mengubah tipe data kategorikal menjadi numerik. Pada fitur Deskripsi Cuaca dan Wilayah digunakan metode *label encoding*, sedangkan fitur Arah Angin diterapkan *one-hot encoding*. Hasil transformasi ditampilkan pada tabel berikut.

Tabel 3 Hasil transformasi fitur Deskripsi Cuaca

Deskripsi Cuaca	Deskripsi Cuaca_Encoded
Berawan	0
Hujan Petir	3
Hujan Ringan	4
Udara Kabur	6
Cerah Berawan	9

Tabel 3 menunjukkan hasil penerapan metode *label encoding* pada fitur Deskripsi Cuaca. Setiap kategori cuaca diubah menjadi representasi numerik untuk memudahkan pemrosesan oleh algoritma pembelajaran mesin

Tabel 4 Hasil transformasi fitur Wilayah

Wilayah	Wilayah_Encoded
Desa Batu Makjage	95
Desa Muara Semayok	699
Desa Dalam Kaum	1662
Desa Sibau Hulu	1296
Desa Pemangkat Kota	940

Tabel 4 menyajikan hasil *label encoding* pada fitur Wilayah. Setiap nama wilayah dikonversi menjadi nilai numerik unik untuk memfasilitasi pengolahan data oleh algoritma pembelajaran mesin Cara mudah membuat layout adalah dengan menggunakan panduan ini secara langsung. Selanjutnya fitur arah angin yang menerapkan *One-Hot-Encoding*. Teknik ini mengonversi setiap kategori pada fitur Arah Angin menjadi variabel biner yang menunjukkan keberadaan kategori tersebut pada setiap observasi. Sebagai contoh, dapat diperhatikan pada Tabel 5 berikut.

Tabel 5 Hasil transformasi fitur Arah Angin

Arah Angin Sebelum Encoding	Angin_E	Angin_N	Angin_NE	Angin_NW	Angin_S
E	1	0	0	0	0
N	0	1	0	0	0
NE	0	0	1	0	0
NW	0	0	0	1	0
S	0	0	0	0	1

*Label encoding* dan *one-hot encoding* diterapkan untuk mengubah data kategorikal menjadi bentuk numerik agar dapat diolah oleh algoritma pembelajaran mesin. Transformasi ini memastikan setiap kategori direpresentasikan dengan tepat sehingga mendukung proses pemodelan secara lebih akurat dan efisien.

### 3.4. Modelling

Pada tahap ini, algoritma *Random Forest* dioptimalkan menggunakan *Grid Search CV* untuk memperoleh kombinasi parameter terbaik. Parameter optimal yang diperoleh kemudian digunakan dalam pembangunan model *Random Forest*. Hasil optimasi parameter dapat dilihat pada Tabel 6 berikut.

Tabel 6 Hasil pencarian parameter terbaik

Parameter	Nilai Terbaik	Keterangan
<i>n_estimators</i>	150	Jumlah pohon dalam ensemble
<i>max_depth</i>	15	Kedalaman maksimum tiap pohon
<i>max_features</i>	'sqrt'	Fitur yang dipilih pada setiap split
<i>min_samples_split</i>	2	Minimum sampel untuk split internal
<i>min_samples_leaf</i>	1	Minimum sampel di daun pohon

Berdasarkan Tabel 6, parameter terbaik yang diperoleh selanjutnya digunakan dalam penerapan algoritma *Random Forest*. Setelah itu, dataset kembali dibagi menggunakan metode *train-test split* dengan proporsi 80% data latih dan 20% data uji.

### 3.5. Evaluasi

Tahap evaluasi dilakukan untuk menilai kinerja model *Random Forest* menggunakan *confusion matrix* yang menghasilkan metrik akurasi, presisi, recall, dan F1-score. Evaluasi ini dilakukan setelah dataset dibagi menjadi 80% data latih dan 20% data uji. Hasil evaluasi model disajikan pada Tabel 7 berikut.

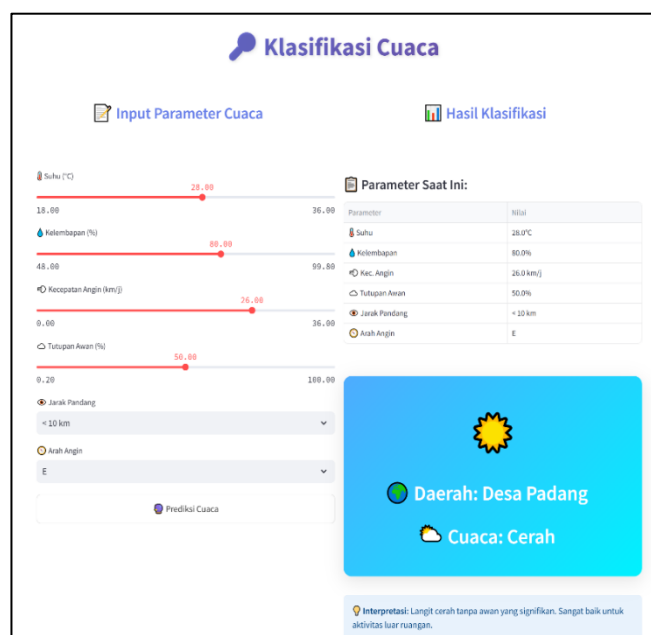
Tabel 7 Hasil evaluasi model *Random Forest*

Metrik	Data Latih	Data Uji
Akurasi	98.84 %	95.33 %
Presisi	98.87 %	95.31 %
Recall	98.84 %	95.33 %
F1 – Score	98.84 %	95.27 %

Tabel 7 menunjukkan bahwa model *Random Forest* memperoleh akurasi tinggi, yaitu 98,84% pada data latih dan 95,33% pada data uji, dengan metrik evaluasi lain yang konsisten. Hal ini menegaskan bahwa model memiliki kemampuan generalisasi yang baik dan efektif untuk klasifikasi data prakiraan cuaca.

### 3.6. Deployment

Pada tahap akhir, model yang telah dilatih diintegrasikan ke dalam aplikasi berbasis web. Model tersebut disimpan dalam format pickle dan dihubungkan dengan sistem aplikasi untuk menghasilkan prediksi berdasarkan parameter cuaca yang diinput oleh pengguna. Hasil implementasi ditampilkan pada gambar 7 berikut.



Gambar 7 Hasil integrasi model *Random Forest* pada aplikasi web

Gambar 7 menunjukkan antarmuka yang memungkinkan pengguna memasukkan parameter cuaca untuk memperoleh hasil klasifikasi berupa prediksi daerah dan kondisi cuaca yang sesuai.

#### **4. KESIMPULAN**

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* menunjukkan kinerja yang baik. Dalam data latih, model mencapai akurasi 98,84%, presisi 98,87%, recall 98,84%, dan skor F1-nya 98,84%. Dalam data uji, kinerjanya tetap tinggi dengan akurasi 95,33%, presisi 95,31%, recall 95,33%, dan skor F1-nya 95,27%. Hasil ini menunjukkan bahwa *Random Forest* memiliki kemampuan untuk memberikan klasifikasi prakiraan cuaca yang akurat dan stabil. Selain itu, aplikasi web ini dapat membantu pengguna mendapatkan hasil prediksi dengan antarmuka yang mudah digunakan.

**DAFTAR PUSTAKA**

- [1] A. M. Siregar, “Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning,” *PETIR*, vol. 13, no. 2, pp. 138–147, Sep. 2020, doi: 10.33322/petir.v13i2.998.
- [2] A. Zulfiani and C. Fauzi, “Penerapan Algoritma Backpropagation Untuk Prakiraan Cuaca Harian Dibandingkan Dengan Support Vector Machine dan Logistic Regression,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, pp. 1–9, Jul. 2023, doi: 10.30865/mib.v7i3.6173.
- [3] M. Ciszynski and K. Chrominski, “Applying machine learning techniques in weather forecast modeling,” in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 4133–4141. doi: 10.1016/j.procs.2024.09.252.
- [4] Y. Rawie and M. S. Rofii, “Analisis kebijakan bantuan internasional bencana alam dalam dinamika ketahanan nasional dengan pendekatan Delphi,” *Journal of National Paradigm-Based Resilience Strategy*, vol. 1, no. 1, Feb. 2024, doi: 10.61511/napbres.v1i1.2024.653.
- [5] R. Risanti, “ANALISIS MODEL PREDIKSI CUACA MENGGUNAKAN SUPPORT VECTOR MACHINE, GRADIENT BOOSTING, RANDOM FOREST, DAN DECISION TREE,” Jakarta Timur, Jan. 2024, pp. 1–10. doi: 10.21009/03.1201.FA18.
- [6] T. Fawwaz, F. Ihsan, I. Ramadhan, D. R. Akbar, and E. Ismanto, “Perbandingan Model Machine Learning Untuk Klasifikasi Deteksi Penyakit Jantung,” vol. 6, no. 2, p. 203, 2025, [Online]. Available: <http://ejournal.umri.ac.id/index.php/coscitech/index>
- [7] Munaldi. S.kom., M.kom, *Algoritma Random Forest dan XGBoost Menggunakan Python*. Banyumas: PT. GANESHA KREASI SEMESTA, 2025.
- [8] A. S. A. Yuda, M. D. A. Rosady, N. I. Faisal, and E. Ismanto, “Analisis Kinerja Algoritma K-Nearest Neighbors (KNN) dan Random Forest untuk Klasifikasi Kondisi Cuaca,” *Jurnal Computer Science and Information Technology (CoSciTech)*, vol. 6, no. 2, p. 338, 2024, [Online]. Available: <http://ejournal.umri.ac.id/index.php/coscitech/index>
- [9] F. Sulianta, *Buku Dasar Data Mining from AtoZ-Feri SLN*, no. January. Bandung: ResearchGate, 2024. [Online]. Available: <https://www.researchgate.net/publication/377018853%0ABuku>
- [10] Nur Cholifah Sastya dan and Isna Nugraha, “Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution,” Kota Surabaya, Jawa Timur, Oct. 2023. [Online]. Available: <http://ejournal.unis.ac.id/index.php/UNISTEK>