

EVALUASI KINERJA ALGORITMA *RANDOM FOREST* DAN KNN UNTUK PREDIKSI CUACA DI JAKARTA

Nurdin wibowo¹, Wahyu Hardyanto², Alpon Sepriando³,
Djuniadi⁴, Tri Nurmayati⁵, Fakhrol Alam⁷
Universitas Negeri Semarang^{1,2,4}
Badan Meteorologi Klimatologi dan Geofisika^{3,5,7}
nurdinwibowo@students.unnes.ac.id¹

ABSTRAK

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* dan *K-Nearest Neighbors* (KNN) dalam klasifikasi kategori cuaca harian di Stasiun Meteorologi Kemayoran, Jakarta. Data yang digunakan merupakan data observasi BMKG periode 2017–2023 dengan target klasifikasi *No Significant Weather*, RA, TS, dan TS,RA. Variabel yang digunakan meliputi suhu, curah hujan, tekanan udara, kelembaban, kecepatan angin, durasi penyinaran matahari, arah angin, dan bulan. Data melalui tahap prapengolahan, kemudian dibagi menjadi data pelatihan dan pengujian dengan rasio 80:20. Hasil penelitian menunjukkan bahwa *Random Forest* memperoleh akurasi 78% dengan F1-score rata-rata tertimbang 0,75, sedangkan KNN memperoleh akurasi 65% dengan F1-score rata-rata tertimbang 0,59. *Random Forest* menunjukkan kinerja lebih baik dalam mengklasifikasikan kategori cuaca dominan, meskipun kedua algoritma masih mengalami keterbatasan pada kategori minoritas, khususnya TS. Hasil ini menunjukkan bahwa *Random Forest* lebih sesuai digunakan untuk klasifikasi cuaca harian berbasis data observasi BMKG di wilayah perkotaan.

Kata Kunci: *K-Nearest Neighbors*; Prediksi Cuaca; *Random Forest*

ABSTRACT

This study aims to compare the performance of Random Forest and K-Nearest Neighbors (KNN) algorithms in classifying daily weather categories at Kemayoran Meteorological Station, Jakarta. The data used were BMKG observation data from 2017 to 2023, with classification targets consisting of No Significant Weather, RA, TS, and TS,RA. The variables included temperature, rainfall, air pressure, humidity, wind speed, sunshine duration, wind direction, and month. The data were preprocessed and divided into training and testing sets using an 80:20 ratio. The results showed that Random Forest achieved an accuracy of 78% with a weighted average F1-score of 0.75, while KNN achieved an accuracy of 65% with a weighted average F1-score of 0.59. Random Forest performed better in classifying dominant weather categories, although both algorithms still had limitations in identifying minority categories, particularly TS. These findings indicate that Random Forest is more suitable for daily weather classification based on BMKG observation data in urban areas.

Keywords: *K-Nearest Neighbors; Random Forest; Weather Classification*

PENDAHULUAN

Cuaca merupakan salah satu unsur atmosfer yang berpengaruh langsung terhadap aktivitas masyarakat, terutama di wilayah perkotaan yang memiliki kepadatan penduduk, mobilitas tinggi, serta aktivitas ekonomi yang kompleks. DKI Jakarta sebagai pusat pemerintahan dan perekonomian nasional memiliki tingkat kebutuhan yang tinggi terhadap informasi cuaca harian yang akurat. Perubahan suhu, kelembaban, curah hujan, tekanan udara, kecepatan angin, arah angin, dan durasi penyinaran matahari dapat memengaruhi aktivitas transportasi, kesehatan, kegiatan ekonomi, hingga kesiapsiagaan masyarakat terhadap bencana hidrometeorologi.

Jakarta juga termasuk wilayah perkotaan yang memiliki kerentanan terhadap kejadian cuaca ekstrem, khususnya hujan lebat yang dapat memicu genangan dan banjir. Priyambodoho et al. (2022) menjelaskan bahwa perkembangan wilayah perkotaan dan perubahan iklim dapat meningkatkan risiko banjir di Jakarta melalui perubahan intensitas hujan, debit puncak, dan luasan genangan. Nasution et al. (2022) juga menunjukkan bahwa kerentanan banjir di Jakarta berkaitan dengan kondisi wilayah perkotaan, kepadatan penduduk, dan riwayat kejadian banjir. Oleh karena itu, peningkatan kualitas prediksi dan klasifikasi cuaca harian menjadi penting untuk mendukung mitigasi risiko serta pengambilan keputusan di wilayah perkotaan.

Perkembangan teknologi informasi mendorong pemanfaatan metode

machine learning dalam analisis dan prediksi cuaca. Bochenek dan Ustrnul (2022) menyatakan bahwa *machine learning* semakin banyak digunakan dalam kajian prediksi cuaca dan iklim karena mampu memproses data meteorologi yang besar dan kompleks. Zhang et al. (2025) juga menjelaskan bahwa metode *machine learning* memiliki kemampuan dalam mengidentifikasi pola hubungan antarvariabel cuaca yang sulit ditangkap oleh pendekatan konvensional. Dalam konteks prediksi cuaca, pendekatan ini dapat digunakan untuk mengenali pola historis dari parameter meteorologi dan menghasilkan klasifikasi kondisi cuaca berdasarkan data pengamatan.

Beberapa penelitian menunjukkan bahwa metode *machine learning* dapat diterapkan untuk memprediksi maupun mengklasifikasikan kondisi cuaca. AbdulRaheem et al. (2022) mengevaluasi beberapa algoritma *machine learning* untuk prediksi cuaca dan menunjukkan bahwa pendekatan klasifikasi dapat digunakan untuk menilai kemampuan model dalam mengenali kondisi atmosfer. Fowdur dan Nazir (2022) mengembangkan sistem prediksi cuaca berbasis *machine learning* secara kolaboratif dengan beberapa lokasi prediktor, sedangkan Safia et al. (2023) menggunakan pendekatan *supervised learning* untuk klasifikasi kondisi cuaca pada beberapa kota. Penelitian-penelitian tersebut menunjukkan bahwa data historis meteorologi dapat dimanfaatkan sebagai dasar pengembangan model klasifikasi cuaca.

Random Forest merupakan salah satu algoritma *machine learning* yang

banyak digunakan dalam prediksi dan klasifikasi cuaca karena mampu menangani hubungan antarvariabel yang kompleks. Meenal et al. (2021) menunjukkan bahwa *Random Forest* dapat digunakan dalam prediksi cuaca dan memberikan performa yang baik pada data meteorologi. Kim et al. (2021) juga menerapkan *Random Forest* untuk prediksi visibilitas berbasis data observasi meteorologi dan memperoleh hasil yang lebih baik dibandingkan keluaran model prediksi konvensional. Selain itu, Meenal et al. (2022) menunjukkan bahwa metode *machine learning* dapat meningkatkan hasil prediksi parameter cuaca dibandingkan pendekatan empiris konvensional.

Dalam konteks Indonesia, penerapan *Random Forest* untuk prediksi cuaca juga telah banyak dilakukan. Hamami dan Dahlan (2022) menerapkan *Random Forest* untuk klasifikasi cuaca di Provinsi DKI Jakarta dengan teknik *oversampling* untuk menangani ketidakseimbangan kelas. Dwiyantri dan Prianto (2023) menggunakan *Random Forest* untuk prediksi cuaca Kota Jakarta dan menunjukkan bahwa algoritma tersebut mampu memberikan hasil klasifikasi yang cukup baik. Putra et al. (2024) juga menerapkan *Random Forest* dan *Grid Search Cross-Validation* untuk prediksi hujan pada sektor pertanian, sedangkan Rakhmat dan Mutohar (2023) menggunakan *Random Forest* dan *cross validation* dalam prakiraan hujan.

Penelitian-penelitian tersebut menunjukkan bahwa *Random Forest* relevan digunakan pada data cuaca historis di Indonesia.

Selain *Random Forest*, *K-Nearest Neighbors* (KNN) juga digunakan dalam

klasifikasi cuaca karena memiliki konsep sederhana dan mudah diterapkan. KNN bekerja dengan menentukan kelas suatu data berdasarkan kedekatan data tersebut terhadap sejumlah tetangga terdekat. Putri et al. (2025) menerapkan metode KNN untuk prediksi data cuaca dan memperoleh akurasi terbaik pada nilai k tertentu. Safia et al. (2023) menjelaskan bahwa KNN dapat memberikan performa yang baik pada beberapa variabel cuaca, tetapi hasilnya sangat dipengaruhi oleh karakteristik data dan skala fitur. Shaiba et al. (2022) juga memasukkan KNN sebagai salah satu algoritma dalam sistem prediksi cuaca berbasis *ensemble machine learning*.

Penelitian yang membandingkan beberapa algoritma klasifikasi cuaca menunjukkan bahwa performa model dapat berbeda sesuai karakteristik data, jumlah kelas, serta distribusi data. Kareem et al. (2021) membandingkan beberapa algoritma klasifikasi untuk prediksi kondisi cuaca dan menunjukkan bahwa *Random Forest* memperoleh performa terbaik dibandingkan beberapa algoritma lain pada dataset yang digunakan. Yuda et al. (2025) secara khusus membandingkan KNN dan *Random Forest* dalam klasifikasi kondisi cuaca, dengan hasil bahwa *Random Forest* memiliki akurasi lebih tinggi dan performa yang lebih stabil pada kelas mayoritas. Darmawan et al. (2025) juga menunjukkan bahwa pemilihan model dan pengaturan parameter berpengaruh terhadap kinerja klasifikasi cuaca.

Meskipun demikian, klasifikasi cuaca masih menghadapi tantangan, terutama pada kategori cuaca yang jumlah datanya sedikit. Zhang et al. (2025) menegaskan bahwa kejadian cuaca yang jarang muncul masih menjadi salah satu

tantangan dalam penerapan *machine learning* untuk prediksi cuaca. Yuda et al. (2025) juga menunjukkan bahwa model *Random Forest* dan KNN dapat mengalami penurunan performa pada kelas minoritas akibat ketidakseimbangan data. Kondisi ini relevan dengan klasifikasi cuaca harian karena beberapa kategori cuaca, seperti kejadian hujan disertai petir, umumnya memiliki jumlah data lebih sedikit dibandingkan kategori cuaca dominan.

Berdasarkan penelitian terdahulu, penggunaan *Random Forest* dan KNN dalam prediksi maupun klasifikasi cuaca telah banyak dilakukan. Namun, penelitian yang secara khusus membandingkan kinerja *Random Forest* dan KNN menggunakan data observasi BMKG Stasiun Meteorologi Kemayoran periode 2017–2023 dengan target klasifikasi *No Significant Weather*, RA, TS, dan TS,RA masih perlu dikaji. Kebaruan penelitian ini terletak pada evaluasi perbandingan dua algoritma tersebut dalam mengklasifikasikan kategori cuaca harian berbasis data observasi meteorologi pada wilayah perkotaan tropis Jakarta.

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja algoritma *Random Forest* dan *K-Nearest Neighbors* dalam klasifikasi kategori cuaca harian di Stasiun Meteorologi Kemayoran, Jakarta. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui kemampuan masing-masing algoritma dalam mengenali kategori cuaca. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai algoritma yang lebih sesuai untuk klasifikasi cuaca berbasis data

observasi BMKG serta menjadi masukan awal dalam pengembangan model prediksi cuaca berbasis *machine learning* di wilayah perkotaan.

METODE PENELITIAN

Penelitian ini menggunakan data observasi cuaca harian dari Stasiun Meteorologi Kemayoran, Jakarta, periode 2017–2023. Data yang digunakan terdiri atas 1.343 baris dan 10 variabel, yaitu kategori cuaca, suhu minimum, suhu maksimum, curah hujan, tekanan udara, kelembaban relatif, kecepatan angin maksimum, durasi penyinaran matahari, arah angin, dan bulan. Kategori cuaca yang menjadi target klasifikasi terdiri atas *No Significant Weather*, RA, TS, dan TS,RA. Sebelum digunakan dalam pemodelan, data melalui tahap prapengolahan yang mencakup pemeriksaan nilai hilang, penyesuaian tipe data, pengkodean variabel kategorikal, serta normalisasi variabel numerik agar data memiliki skala yang seragam.

Tahapan penelitian dilakukan dengan membagi dataset menjadi data pelatihan dan data pengujian menggunakan rasio 80:20. Dua algoritma yang digunakan dalam penelitian ini adalah *Random Forest* dan *K-Nearest Neighbors*. Kedua model dilatih menggunakan data pelatihan, kemudian diuji menggunakan data pengujian untuk mengetahui kemampuan masing-masing algoritma dalam mengklasifikasikan kategori cuaca. Evaluasi kinerja model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Hasil evaluasi kedua algoritma kemudian dibandingkan untuk menentukan model yang memiliki performa lebih baik dalam klasifikasi

kategori cuaca harian di wilayah Kemayoran, Jakarta.

HASIL PENELITIAN

Dataset

Dataset yang digunakan dalam penelitian ini terdiri atas 1.343 baris dan

10 variabel. Variabel tersebut meliputi kategori cuaca, suhu minimum, suhu maksimum, curah hujan, tekanan udara, kelembaban relatif, kecepatan angin maksimum, durasi penyinaran matahari, arah angin, dan bulan. Contoh struktur dataset ditampilkan pada Tabel 1.

Tabel 1.
Dataset Output

	Weather Specific	T Min	T Max	Rr	Qff	Rh	W Max	Sunshine 24h	W Dir	Month
0	Ra	27.2	34.0	0.2	1007.7	69.825	4.1	5.0	W	1
1	No Significant Weather	27.8	33.8	0.0	1008.3	67.775	4.1	7.5	N	1
2	No Significant Weather	25.8	33.0	0.0	1007.4	70.700	5.7	8.2	N	1
3	Ra	26.4	31.2	3.0	1006.9	76.375	4.1	5.3	W	1
4	No Significant Weather	26.8	31.4	0.0	1006.8	72.025	4.1	4.0	Nw	1
...
1338	No Significant Weather	27.0	33.6	0.0	1009.5	70.250	7.7	7.8	C	12
1339	No Significant Weather	26.2	34.2	0.0	1010.3	63.500	8.2	8.0	C	12
1340	Ts,Ra	26.4	30.4	36.0	1012.4	79.750	3.6	2.0	C	12
1341	Ra	27.0	32.5	2.5	1010.7	78.750	3.8	3.8	C	12
1342	Ra	26.4	32.0	0.0	1009.9	80.500	5.1	2.4	C	12

Pemeriksaan Kualitas Dataset

Pemeriksaan kualitas dataset dilakukan untuk mengetahui tipe data pada setiap variabel. Hasil pemeriksaan menunjukkan bahwa variabel *WEATHER SPECIFIC* dan *Wdir* termasuk data kategorikal, sedangkan

variabel *Tmin*, *Tmax*, *RR*, *QFF*, *RH*, *Wmax*, *SUNSHINE 24H*, dan *MONTH* termasuk data numerik. Pemeriksaan ini dilakukan untuk memastikan bahwa setiap variabel telah sesuai dengan kebutuhan proses prapengolahan dan pemodelan.

Hasil pemeriksaan kualitas dataset karakteristik data cuaca harian di Stasiun ditampilkan pada Gambar 1.

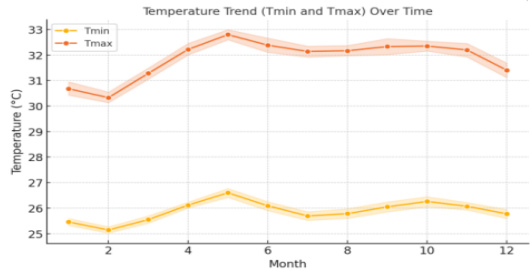
WEATHER SPECIFIC	object
Tmin	float64
Tmax	float64
RR	float64
QFF	float64
RH	float64
Wmax	float64
SUNSHINE 24H H	float64
Wdir	object
MONTH	int64

dtype: object

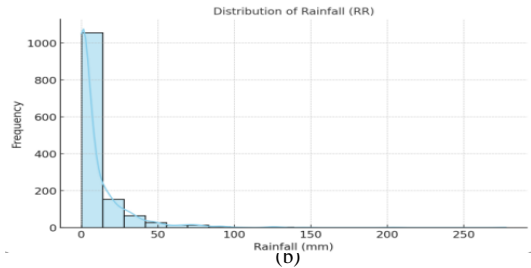
Gambar 1.
Pemeriksaan Dataset

Eksplorasi Dataset

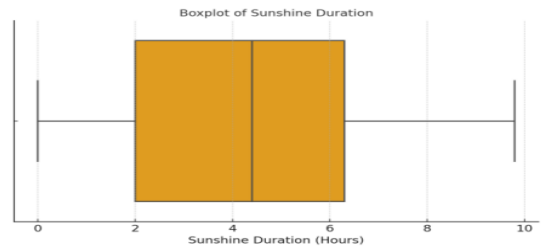
Eksplorasi dataset dilakukan untuk melihat pola awal dari variabel meteorologi yang digunakan dalam penelitian. Hasil eksplorasi menunjukkan bahwa curah hujan memiliki sebaran yang tidak merata dan terdapat beberapa nilai ekstrem. Variabel suhu minimum, suhu maksimum, tekanan udara, kelembaban relatif, kecepatan angin maksimum, dan durasi penyinaran matahari menunjukkan pola distribusi yang lebih stabil. Pola ini memberikan gambaran awal mengenai



Meteorologi Kemayoran. Hasil eksplorasi dataset ditampilkan pada Gambar 2.



(c)



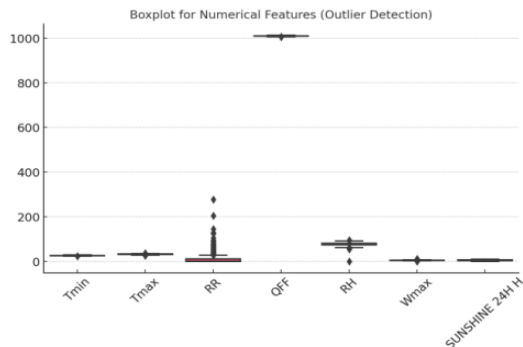
(d)

Gambar 2.

Visualisasi Eksplorasi Dataset Untuk Gambar (a), (b), (c), (d)

Sanitasi Dataset

Proses membersihkan dan memvalidasi data untuk memastikan kualitas, konsistensi, dan kesiapan analisis. Proses ini mencakup penanganan data hilang, deteksi outlier, penghapusan

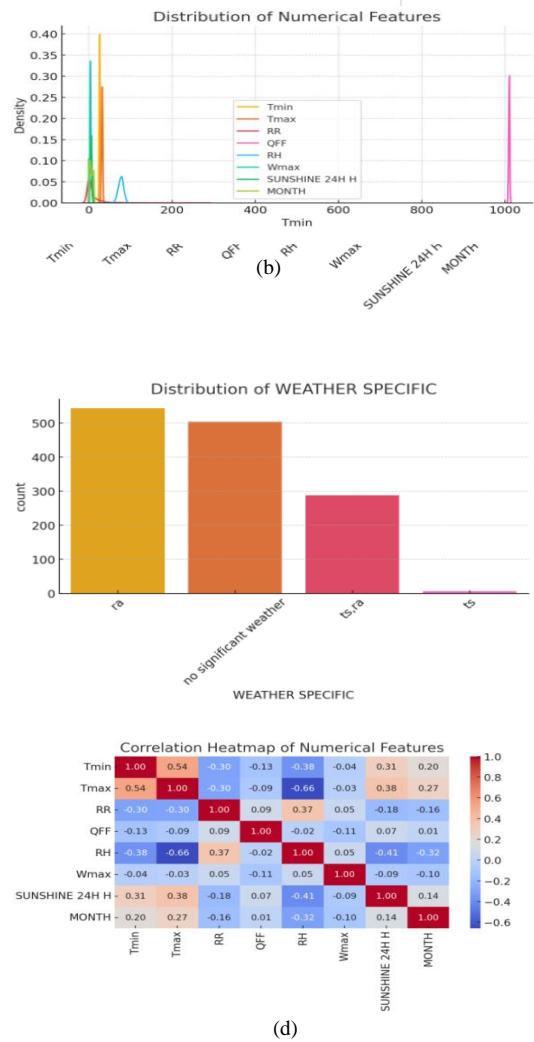


(a)

duplikasi, normalisasi, dan validasi nilai agar sesuai dengan standar.

Visualisasi hasil sanitasi dan analisis dataset menunjukkan beberapa temuan penting. Pada boxplot, terlihat bahwa variabel numerik seperti RR (curah hujan) masih memiliki outlier signifikan, sementara variabel lainnya memiliki distribusi yang lebih terkonsentrasi dan stabil ditampilkan pada Gambar 3(a). Histogram distribusi numerik memperlihatkan persebaran variabel seperti suhu minimum (Tmin), suhu maksimum (Tmax), dan curah hujan (RR), dengan sebagian besar nilai terkonsentrasi pada rentang tertentu namun memiliki distribusi yang berbeda antar variabel ditampilkan pada Gambar 3(b).

Sebaran kategori cuaca yang divisualisasikan dalam WEATHER SPECIFIC menunjukkan bahwa kategori RA (hujan) dan no significant weather adalah yang paling dominan, sedangkan kategori lain seperti TS (thunderstorm) muncul dengan frekuensi yang jauh lebih rendah ditampilkan pada Gambar 3(c). Heatmap korelasi menunjukkan hubungan moderat hingga rendah antar variabel numerik, di mana terdapat korelasi positif antara Tmin dan Tmax, serta hubungan negatif antara Tmax dan RH. Visualisasi ini memberikan gambaran mendalam yang dapat digunakan untuk analisis lebih lanjut dan pengembangan model prediksi yang lebih akurat ditampilkan pada Gambar 3(d).



Gambar 3.

Visualisasi Sanitasi Dataset Untuk Gambar (a), (b), (c), (d)

Pembagian Dataset

Dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Data pelatihan terdiri atas 1.074 data dengan 9 variabel fitur, sedangkan data pengujian terdiri atas 269 data dengan 9 variabel fitur. Pembagian ini digunakan untuk melatih dan menguji performa algoritma *Random Forest* dan *K-Nearest*

Neighbors dalam mengklasifikasikan kategori cuaca harian.

Hasil Klasifikasi Random Forest

Hasil pengujian model *Random Forest* menunjukkan bahwa model mampu mengklasifikasikan kategori cuaca dengan akurasi sebesar 0,78. Kategori No Significant Weather memperoleh nilai precision, recall, dan F1-score yang tinggi. Kategori RA juga menunjukkan performa yang cukup baik, sedangkan kategori TS dan TS,RA masih memiliki nilai recall dan F1-score yang lebih rendah. Hasil evaluasi model *Random Forest* ditampilkan pada Tabel 2.

Tabel 2.

Metrik akurasi dan F1 - Score *Random Forest*

Kategori Cuaca	Precision	Recall	F1-Score	Support
No significant weather	0.91	0.97	0.94	105
RA	0.68	0.85	0.75	104
TS	1.00	0.00	0.00	1
TS,RA	0.70	0.32	0.44	59
Accuracy		0.78		269
Macro Avg	0.82	0.53	0.53	269
Weighted Avg	0.78	0.78	0.75	269

Hasil Klasifikasi K-Nearest Neighbors

Hasil pengujian model *K-Nearest Neighbors* menunjukkan bahwa model memperoleh akurasi sebesar 0,65. Kategori No Significant Weather memiliki performa yang cukup baik, sedangkan kategori RA memiliki recall yang relatif tinggi. Namun, kategori TS dan TS,RA masih sulit diklasifikasikan dengan baik oleh model. Hasil evaluasi

model *K-Nearest Neighbors* ditampilkan pada Tabel 3.

Tabel 3.

Metrik akurasi dan F1 - Score KNN

Kategori Cuaca	Precision	Recall	F1-Score	Support
No significant weather	0.81	0.78	0.80	105
RA	0.55	0.88	0.68	104
TS	1.00	0.00	0.00	1
TS,RA	0.50	0.03	0.06	59
Accuracy	-	0.650	-	269
Macro Avg	0.715	0.422	0.385	269
Weighted Avg	0.6448	0.650	0.587	269

Perbandingan Hasil Klasifikasi

Berdasarkan hasil evaluasi, *Random Forest* memiliki performa yang lebih baik dibandingkan *K-Nearest Neighbors* dalam klasifikasi kategori cuaca harian di Stasiun Meteorologi Kemayoran. *Random Forest* memperoleh akurasi dan F1-score rata-rata tertimbang yang lebih tinggi dibandingkan KNN. Meskipun demikian, kedua model masih mengalami kesulitan dalam mengenali kategori cuaca minoritas, terutama TS, karena jumlah data pada kategori tersebut sangat terbatas. Temuan ini menunjukkan bahwa distribusi data antar kategori cuaca menjadi salah satu faktor yang memengaruhi performa model klasifikasi.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memiliki kinerja yang lebih baik dibandingkan *K-Nearest Neighbors* dalam klasifikasi kategori cuaca harian di Stasiun

Meteorologi Kemayoran. Keunggulan *Random Forest* dapat dikaitkan dengan cara kerja algoritma ini yang menggabungkan banyak pohon keputusan sehingga model lebih stabil dalam mengenali pola hubungan antarvariabel meteorologi. Pada data cuaca yang memiliki variasi nilai suhu, curah hujan, kelembaban, tekanan udara, kecepatan angin, dan penyinaran matahari, kemampuan *Random Forest* dalam menangani pola data yang kompleks menjadi salah satu faktor yang mendukung performanya. Temuan ini sejalan dengan Meenal et al. (2021) yang menunjukkan bahwa *Random Forest* dapat digunakan untuk prediksi cuaca berbasis data meteorologi. Hasil ini juga selaras dengan Yuda et al. (2025) yang menemukan bahwa *Random Forest* memiliki performa lebih tinggi dibandingkan KNN dalam klasifikasi kondisi cuaca.

Kinerja *Random Forest* yang lebih baik juga relevan dengan penelitian Hamami dan Dahlan (2022) yang menerapkan *Random Forest* untuk klasifikasi cuaca di DKI Jakarta. Penelitian tersebut menunjukkan bahwa *Random Forest* dapat digunakan untuk mengolah data cuaca di wilayah perkotaan, meskipun ketidakseimbangan kelas tetap menjadi tantangan dalam proses klasifikasi. Dalam penelitian ini, *Random Forest* mampu mengenali kategori mayoritas seperti *No Significant Weather* dan RA dengan lebih baik dibandingkan kategori minoritas. Hal ini menunjukkan bahwa algoritma *Random Forest* cukup efektif dalam mengenali pola cuaca dominan, tetapi performanya masih dipengaruhi oleh distribusi data pada setiap kelas.

Sementara itu, algoritma *K-Nearest Neighbors* memperoleh performa yang lebih rendah dibandingkan *Random Forest*. Kondisi ini dapat dijelaskan karena KNN bekerja berdasarkan kedekatan jarak antar data, sehingga hasil klasifikasi sangat dipengaruhi oleh skala variabel, pemilihan nilai k , dan distribusi data latih. Pada data meteorologi yang memiliki karakteristik variabel beragam, seperti curah hujan yang memiliki rentang nilai lebar dan variabel cuaca lain yang cenderung lebih stabil, proses klasifikasi berbasis jarak dapat menjadi kurang optimal. Putri et al. (2025) menjelaskan bahwa KNN dapat digunakan untuk prediksi data cuaca, tetapi kinerjanya sangat bergantung pada karakteristik data dan parameter yang digunakan. Oleh karena itu, hasil penelitian ini menunjukkan bahwa KNN masih dapat mengenali kategori cuaca tertentu, tetapi belum sebaik *Random Forest* dalam mengklasifikasikan keseluruhan kategori cuaca.

Kedua algoritma masih mengalami kesulitan dalam mengenali kategori cuaca minoritas, terutama TS. Hal ini disebabkan oleh jumlah data pada kategori tersebut yang sangat sedikit dibandingkan kategori lainnya. Ketidakseimbangan data menyebabkan model lebih banyak mempelajari pola dari kelas mayoritas, sehingga kemampuan model dalam mengenali kelas minoritas menjadi terbatas. Yuda et al. (2025) juga menemukan bahwa *Random Forest* dan KNN mengalami penurunan performa pada kelas minoritas akibat distribusi data yang tidak seimbang. Kondisi ini menunjukkan bahwa kualitas dan keseimbangan data menjadi faktor penting dalam membangun model klasifikasi cuaca yang lebih andal.

Hasil penelitian ini memperkuat pandangan bahwa metode *machine learning* dapat digunakan sebagai pendekatan pendukung dalam analisis dan klasifikasi cuaca. Bochenek dan Ustrnul (2022) menjelaskan bahwa *machine learning* memiliki peluang besar dalam prediksi cuaca dan analisis iklim karena mampu memproses data meteorologi yang kompleks. Namun, penerapan metode ini tetap membutuhkan perhatian terhadap kualitas data, pemilihan fitur, distribusi kelas, dan evaluasi model. Dengan demikian, penggunaan *Random Forest* dan KNN dalam penelitian ini dapat memberikan gambaran awal mengenai potensi pemanfaatan *machine learning* untuk klasifikasi kategori cuaca berbasis data observasi BMKG.

Berdasarkan hasil perbandingan, *Random Forest* lebih sesuai digunakan pada dataset penelitian ini karena mampu memberikan performa klasifikasi yang lebih stabil dibandingkan KNN. Meskipun demikian, keterbatasan pada kategori minoritas menunjukkan bahwa model masih memerlukan pengembangan lebih lanjut, terutama dari sisi keseimbangan data dan jumlah sampel pada kategori cuaca tertentu. Temuan ini penting karena prediksi kategori cuaca harian tidak hanya memerlukan akurasi yang tinggi pada kelas mayoritas, tetapi juga kemampuan mengenali kejadian cuaca yang lebih jarang muncul. Oleh sebab itu, hasil penelitian ini dapat menjadi dasar awal untuk pengembangan model klasifikasi cuaca yang lebih baik pada wilayah perkotaan tropis seperti Jakarta.

SIMPULAN

Penelitian ini menunjukkan bahwa algoritma *Random Forest* memiliki kemampuan klasifikasi yang lebih baik dibandingkan *K-Nearest Neighbors* dalam memprediksi kategori cuaca harian di Stasiun Meteorologi Kemayoran, Jakarta. *Random Forest* lebih mampu mengenali pola pada kategori cuaca dominan, sedangkan KNN masih memiliki keterbatasan dalam membedakan beberapa kategori cuaca. Meskipun demikian, kedua algoritma masih mengalami kendala dalam mengklasifikasikan kategori cuaca dengan jumlah data yang sangat sedikit. Hasil penelitian ini menunjukkan bahwa metode *machine learning*, khususnya *Random Forest*, berpotensi digunakan sebagai pendekatan pendukung dalam klasifikasi cuaca harian berbasis data observasi BMKG, dengan tetap memperhatikan kualitas dan keseimbangan data yang digunakan.

DAFTAR PUSTAKA

- Abdul Raheem, M., Awotunde, J. B., Adeniyi, A. E., Oladipo, I. D., & Adekola, S. O. (2022). Weather Prediction Performance Evaluation on Selected Machine Learning Algorithms. *IAES International Journal of Artificial Intelligence*, *11*(4), 1535–1544. <https://doi.org/10.11591/ijai.v11.i4.pp1535-1544>
- Bochenek, B., & Ustrnul, Z. (2022). Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. *Atmosphere*, *13*(2), 180. <https://doi.org/10.3390/atmos13020180>

- Darmawan, Z. M. E., Dianta, A. F., Fathoni, K., Rachmawati, O. C. R., & Apriandy, K. I. (2025). Comparison of Machine Learning Classification Methods for Weather Prediction: A Performance Analysis. *G-Tech: Jurnal Teknologi Terapan*, 9(2), 715–727. <https://doi.org/10.70609/gtech.v9i2.6649>
- Dwiyanti, Z. A., & Prianto, C. (2023). Prediksi Cuaca Kota Jakarta Menggunakan Metode *Random Forest*. *Jurnal Tekno Insentif*, 17(2), 127–137. <https://doi.org/10.36787/jti.v17i2.1136>
- Fowdur, T. P., & Nazir, R. M. N. U. D. I. (2022). A Real-Time Collaborative Machine Learning Based Weather Forecasting System with Multiple Predictor Locations. *Array*, 14, 100153. <https://doi.org/10.1016/j.array.2022.100153>
- Hamami, F., & Dahlan, I. A. (2022). Klasifikasi Cuaca Provinsi DKI Jakarta Menggunakan Algoritma *Random Forest* dengan Teknik Oversampling. *Jurnal Teknoinfo*, 16(1), 87–92. <https://doi.org/10.33365/jti.v16i1.1533>
- Kareem, F. Q., Abdulazeez, A. M., & Hasan, D. A. (2021). Predicting Weather Forecasting State Based on Data Mining Classification Algorithms. *Asian Journal of Research in Computer Science*, 9(3), 13–24. <https://doi.org/10.9734/ajrcos/2021/v9i330222>
- Kim, B. Y., Cha, J. W., Chang, K. H., & Lee, C. (2021). Visibility Prediction Over South Korea Based on *Random Forest*. *Atmosphere*, 12(5), 552. <https://doi.org/10.3390/atmos12050552>
- Meenal, R., Michael, P. A., Pamela, D., & Rajasekaran, E. (2021). Weather Prediction Using *Random Forest* Machine Learning Model. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 1208–1215. <https://doi.org/10.11591/ijeecs.v22.i2.pp1208-1215>
- Meenal, R., Kailash, K., Michael, P. A., Joseph, J. J., Josh, F. T., & Rajasekaran, E. (2022). Machine Learning Based Smart Weather Prediction. *Indonesian Journal of Electrical Engineering and Computer Science*, 28(1), 508–515. <https://doi.org/10.11591/ijeecs.v28.i1.pp508-515>
- Nasution, B. I., Saputra, F. M., Kurniawan, R., Ridwan, A. N., Fudholi, A., & Sumargo, B. (2022). Urban Vulnerability to Floods Investigation in Jakarta, Indonesia: A Hybrid Optimized Fuzzy Spatial Clustering and News Media Analysis Approach. *International Journal of Disaster Risk Reduction*, 83, 103407. <https://doi.org/10.1016/j.ijdrr.2022.103407>
- Priyambodoho, B. A., Kure, S., Januriyadi, N. F., Farid, M., Varquez, A. C. G., Kanda, M., & Kazama, S. (2022). Effects of Urban Development on Regional Climate Change and Flood Inundation in Jakarta, *Indonesia. Journal of Disaster Research*, 17(4),

- 516–525.
<https://doi.org/10.20965/jdr.2022.p0516>
- Putra, A. F. D., Azmi, M. N., Wijayanto, H., Utama, S., & Wirawan, I. G. P. W. W. (2024). Optimizing Rain Prediction Model Using *Random Forest* and Grid Search Cross-Validation for Agriculture Sector. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 23(3), 519–530. <https://doi.org/10.30812/matrik.v23i3.3891>
- Putri, A. D. P., Al Haris, M., Fauzi, F., & Amri, S. (2025). K-Nearest Neighbor (KNN) Method For Weather Data Prediction: Penerapan metode K-Nearest Neighbour (KNN) untuk prediksi data cuaca. *Journal of Data Insights*, 3(1), 56–64. <https://doi.org/10.26714/jodi.v3i1.214>
- Rakhmat, G. A., & Mutohar, W. (2023). Prakiraan hujan Menggunakan Metode *Random Forest* dan Cross Validation. *MIND Journal*, 8(2), 173–187. <https://doi.org/10.26760/mindjournal.v8i2.173-187>
- Safia, M., Shahid, M., & Qureshi, M. A. (2023). Classification of Weather Conditions Based on Supervised Learning for Swedish Cities. *Atmosphere*, 14(7), 1174. <https://doi.org/10.3390/atmos14071174>
- Shaiba, H., Marzouk, R., Nour, M. K., Negm, N., Hilal, A. M., Mohamed, A., Motwakel, A., Yaseen, I., Zamani, A. S., & Rizwanullah, M. (2022). Weather Forecasting Prediction Using Ensemble Machine Learning for Big Data Applications. *Computers, Materials & Continua*, 73(2), 3367–3382. <https://doi.org/10.32604/cmc.2022.030067>
- Yuda, A. S. A., Rosady, M. D. A., Faisal, N. I., & Ismanto, E. (2025). Performance analysis of *K-Nearest Neighbors* (KNN) and *Random Forest* Algorithms for Classification of Weather Conditions. *Jurnal CoSciTech (Computer Science and Information Technology)*, 6(2), 337–343. <https://doi.org/10.37859/coscitech.v6i2.9827>
- Zhang, H., Liu, Y., Zhang, C., & Li, N. (2025). Machine Learning Methods for Weather Forecasting: A survey. *Atmosphere*, 16(1), 82. <https://doi.org/10.3390/atmos16010082>