



Analisis Klasifikasi Curah Hujan Harian Menggunakan Metode *Decision Tree* (Studi Kasus: Stasiun Pengamatan Radin Inten II Lampung)

Muklas Rivai¹, Rosni^{✉2}, Nanda Azzanina³, Aldila Nur Indah Berliana Ratam⁴
Bening Zalzabillah Zarrouby⁵, Felicia Perangin Angin⁶, Ikhwan Halim Aqil⁷,
Latifah⁸, Robertus Karlius Laia⁹

¹²³⁴⁵⁶⁷⁸⁹Program Studi Sains Aktuaria, Institut Teknologi Sumatera, Lampung

Email: rosni@at.itera.ac.id

Received 14 July 2025,

Accepted 28 September 2025,

Published 30 September 2025

Abstrak

Peranan hujan bagi kehidupan mempunyai pengaruh yang penting terutama dalam memenuhi kebutuhan air. Namun, perubahan cuaca yang terjadi akhir-akhir ini membuat masyarakat kesulitan dalam memprediksi hujan. Maka dari hal tersebut, prediksi terjadinya hujan dapat membantu dalam perencanaan tata kelola sumber daya air, salah satunya di wilayah Bandar Lampung. Analisis klasifikasi menggunakan metode *Decision Tree* dapat memberikan informasi mengenai faktor-faktor penting yang mempengaruhi peluang terjadinya hujan. Sehingga, prediksi terjadinya hujan dalam beberapa waktu kedepan dapat dilakukan. Proses data mining menghasilkan dataset yang terdiri atas kelembapan relatif, titik embun, curah hujan, total tutupan awan, dan tutupan awan rendah. Tahap analisis klasifikasi dengan menggunakan metode *Decision Tree* mencakup pengumpulan data, pemrosesan data, pembentukan pohon keputusan, dan evaluasi model. Hasil penelitian menunjukkan bahwa dalam mengklasifikasikan variabel-variabel yang mempengaruhi terjadinya hujan di Bandar Lampung, kelembapan relatif memiliki pengaruh yang cukup besar dalam menentukan peluang terjadinya hujan. Selain itu, titik embun dan curah hujan juga memberikan kontribusi dalam klasifikasi tersebut. Nilai akurasi terhadap model pohon keputusan yang dibentuk diperoleh sebesar 84,54%. Penelitian ini memberikan kontribusi dalam pengelolaan sumber daya air dan mengantisipasi terjadinya kekurangan air.

Kata Kunci: Analisis Klasifikasi; Curah Hujan; Data Mining; Metode *Decision Tree*; Prediksi

Abstract

The role of rainfall in human life is crucial, particularly in fulfilling water needs. However, recent climate changes have made it increasingly difficult for communities to predict rainfall. Therefore, rainfall prediction can support water resource management planning, particularly in the Bandar Lampung area. Classification analysis using the *Decision Tree* method can provide information on key factors influencing the probability of rainfall occurrence. Hence, rainfall prediction for the upcoming periods can be conducted. The data mining process generated a dataset consisting of relative humidity, dew point, rainfall, total cloud cover, and low cloud cover. The classification analysis stage using the *Decision Tree* method included data collection, data processing, decision tree construction, and model evaluation. The results of the study show that, in classifying

variables influencing rainfall in Bandar Lampung, relative humidity has a significant impact on determining rainfall probability. In addition, dew point and rainfall also contribute to the classification. The decision tree model achieved an accuracy rate of 84.54%. This study contributes to water resource management and helps anticipate potential water shortages.

Keywords: *Classification Analysis; Rainfall; Data Mining; Decision Tree Method; Prediction*

✉ Corresponding author

PENDAHULUAN

Hujan adalah sumber air utama. Agar semua makhluk hidup, terutama manusia, dapat bertahan hidup, air harus tersedia. Di wilayah Bandar Lampung, variabilitas musiman curah hujan menyebabkan tantangan pengelolaan air yang signifikan, terutama selama musim hujan dan kemarau. Musim hujan sering mengakibatkan banjir karena limpasan permukaan yang berlebihan. Sebaliknya, pada musim kemarau dapat menyebabkan terjadinya krisis air yang parah. Hal ini dapat diperparah oleh pengambilan air tanah secara berlebihan untuk memenuhi kebutuhan rumah tangga [1].

Penurunan keseimbangan antara curah hujan dan tingkat penguapan selama kekeringan dapat menyebabkan peningkatan polusi air dan penurunan kualitas air, sehingga memperburuk krisis yang terjadi [2]. Oleh karena itu, memahami faktor cuaca yang mempengaruhi curah hujan sangat penting untuk mengembangkan strategi untuk mengurangi efek buruk kekeringan dan memastikan pengelolaan air yang berkelanjutan [3] [4]. Curah hujan dipengaruhi oleh berbagai faktor meteorologis, antara lain suhu, kelembapan relatif, kecepatan angin, tekanan udara, tutupan awan, dan radiasi matahari [5]. Curah hujan merupakan jumlah presipitasi yang jatuh di suatu lokasi, yang dipengaruhi oleh sejumlah faktor atmosferik. Pengukuran curah hujan umumnya dinyatakan dalam milimeter pada interval waktu harian, bulanan, atau tahunan. Fluktuasi curah hujan serta faktor-faktor alam lainnya di Indonesia berkontribusi secara signifikan terhadap dinamika kehidupan masyarakat.

Unsur-unsur cuaca menunjukkan pola keberagaman yang dapat dijadikan dasar untuk melakukan klasifikasi hujan. Klasifikasi sering digunakan dalam proses data mining. Data dipetakan ke dalam kelompok atau kelas yang telah ditentukan menggunakan algoritma klasifikasi dalam data mining [6]. Data mining disebut pula dengan *pattern recognition*, yaitu suatu teknik dalam pengolahan data guna menemukan pola tertentu yang belum diketahui dari data yang diolah [7]. Permasalahan dalam mengklasifikasikan peristiwa curah hujan secara akurat diperparah oleh adanya variabilitas yang tinggi serta ketidakpastian dalam pola-pola cuaca yang dinamis dan kompleks. Metode statistik tradisional seperti regresi linear sering kali tidak mampu menangkap hubungan non-linier dan interaksi antar variabel dalam data meteorologi. Oleh karena itu, dibutuhkan pendekatan yang lebih adaptif dan fleksibel dalam menganalisis data cuaca, khususnya dalam mengenali pola yang berhubungan dengan kejadian hujan. Salah satu metode yang dinilai mampu menjawab tantangan ini adalah *decision tree* atau pohon keputusan, karena dapat memetakan hubungan antar variabel secara hierarkis dan mudah diinterpretasikan.

Keunggulan utama metode ini terletak pada kemampuannya dalam menyederhanakan proses pengambilan keputusan yang kompleks melalui visualisasi struktur pohon yang jelas dan sistematis. Selain itu, *decision tree* dapat mengatasi persoalan hubungan non-linier dalam data dan memberikan wawasan mengenai tingkat pengaruh masing-masing variabel terhadap target klasifikasi [8][9]. Pendekatan data mining yang disebut *decision tree* digunakan untuk mengklasifikasikan data dengan memecah kumpulan data yang sangat besar menjadi catatan yang lebih kecil dan mempertimbangkan kriteria objektif. Visualisasi

data dalam bentuk pohon keputusan tidak hanya memudahkan pemahaman terhadap proses klasifikasi, tetapi juga dapat digunakan untuk mengidentifikasi variabel cuaca yang memiliki kontribusi dominan terhadap kejadian hujan [10].

Metode *decision tree* merupakan teknik klasifikasi yang efisien dan informatif, serta banyak digunakan dalam penelitian berbasis data mining di berbagai bidang. Pratiwi dan Ibad [11] mengklasifikasikan faktor-faktor yang memengaruhi kehamilan tidak diinginkan dengan *Decision Tree*, sedangkan Ibrahim & khikmah [12] menggunakan *Decision Tree* dan regresi logistik untuk penentuan kelayakan penerima kredit yang menunjukkan bahwa *decision tree* memiliki akurasi yang lebih tinggi. Moslehi *et. al.* [13] menggunakan *decision tree* untuk mengidentifikasi faktor-faktor yang memengaruhi mortalitas pasien COVID-19, sementara Duan *et al.* [14] memanfaatkannya untuk identifikasi litologi batuan vulkanik.

Penelitian-penelitian sebelumnya menunjukkan fleksibilitas dan keandalan *Decision Tree* dalam berbagai konteks. Namun, penelitian yang secara khusus menganalisis klasifikasi curah hujan harian di Indonesia, khususnya di Bandar Lampung, masih sangat terbatas. Kondisi ini menimbulkan celah penelitian yang dapat diisi dengan menganalisis data meteorologis lokal menggunakan metode *Decision Tree* untuk meningkatkan akurasi prediksi hujan. Berdasarkan hal tersebut, penelitian ini bertujuan mengklasifikasikan kejadian hujan berdasarkan sejumlah unsur cuaca yang berperan dengan memanfaatkan data curah hujan dan variabel meteorologis dari Stasiun Pengamatan Radin Inten II Lampung. Selain itu, dapat mengidentifikasi unsur cuaca yang memberikan kontribusi paling besar dalam klasifikasi kejadian hujan sehingga dapat dimanfaatkan sebagai dasar pengambilan keputusan di bidang meteorologi serta mitigasi risiko bencana berbasis cuaca.

METODOLOGI

Analisis klasifikasi hujan ini menggunakan data sekunder yang diperoleh dari BMKG. Data yang diambil dari BMKG mencakup faktor-faktor penyebab hujan di Bandar Lampung tahun 2024. Variabel (faktor) yang digunakan berjumlah sebanyak 11 variabel, yakni: Temperatur (X_1), Titik Embun (X_2), Kelembapan Relatif (X_3), Angin (X_4), Tekanan Udara (X_5), Curah hujan (X_6), Total Tutupan Awan (X_7), Tutupan Awan Rendah (X_8) Durasi Penyinaran Matahari (X_9), Jarak Pandang (X_{10}), dan Kondisi (X_{11}). Variabel yang menjadi variabel respons adalah Kondisi dan variabel lainnya akan menjadi data inputan yang menghasilkan keputusan hujan, pengolahan data dilakukan dengan menggunakan software R Studio.

Adapun langkah-langkah yang dilakukan dalam penelitian adalah sebagai berikut:

1. Mengumpulkan data dari BMKG yaitu data klasifikasi hujan pada stasiun pengamatan Radin Inten II Lampung bulan Oktober tahun 2023 sampai September tahun 2024
2. Mendeskripsikan data berdasarkan struktur data tersebut seperti numerik dan faktor.
3. Membagi data menjadi dua kategori yaitu data pelatihan (*training*) dan data pengujian (*testing*) dengan rasio 70% dengan banyak data pelatihan 269 data dan 30% data pengujian sebanyak 97 data. Pembagian data pelatihan dan pengujian dengan rasio 70% dan 30% dipilih karena memberikan keseimbangan antara kemampuan model dalam belajar dan validitas evaluasi hasil analisis. Selain itu, juga dapat mengurangi risiko *overfitting*, dan mengikuti praktik terbaik yang lazim digunakan dalam penelitian berbasis klasifikasi dan prediksi [15][16][17].
4. Data pelatihan tersebut digunakan untuk menentukan variabel yang mempunyai pengaruh besar hingga terkecil dengan menggunakan metode *decision tree* yang akan menghasilkan pohon keputusan.
5. Data pengujian digunakan untuk memprediksi seberapa besar nilai akurasi peluang data dengan metode *decision tree*.
6. Tahapan metode *decision tree* [18]
 - a) Menyiapkan data yang sudah dikelompokkan ke dalam kelas-kelas tertentu.
 - b) Menentukan akar dari pohon keputusan.

- c) Menghitung nilai *gain* dan *entropy* dari masing-masing atribut, nilai *gain* tertinggi akan menjadi akar pertama. *Entropy* digunakan untuk mengukur ketidakpastian pada *node* tertentu dan membantu menentukan seberapa baik suatu fitur membagi data. Akar pohon diidentifikasi menggunakan nilai *gain* tertinggi dari masing-masing atribut atau nilai indeks *entropy* terendah.

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

Keterangan :

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi S_i terhadap S

Gain adalah konsep yang digunakan dalam *decision tree* untuk mengukur seberapa banyak informasi yang diperoleh dari membagi dataset berdasarkan atribut tertentu. Konsep ini berasal dari teori informasi dan digunakan untuk menentukan fitur mana yang paling relevan dalam memisahkan data ke dalam kelas-kelas yang berbeda. Semakin besar *information gain*, semakin baik atribut tersebut dalam mengurangi ketidakpastian dan membantu dalam klasifikasi data.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan :

S = Himpunan kasus

A = Fitur

n = Jumlah partisi atribut A

$|S_i|$ = Proporsi S_i terhadap S

$|S|$ = Jumlah kasus dalam S

- d) Ulangi langkah b hingga semua variabel terpartisi.
e) Proses partisi pohon keputusan akan berhenti saat :
1) Menyelesaikan tahapan untuk membuat Decision Tree
2) Semua variabel dalam simpul ke-n mendapat kelas yang sama
3) Tidak ada atribut di dalam variabel yang dipartisi lagi
4) Tidak ada variabel di dalam cabang yang kosong

7. Penarikan Kesimpulan

HASIL DAN PEMBAHASAN

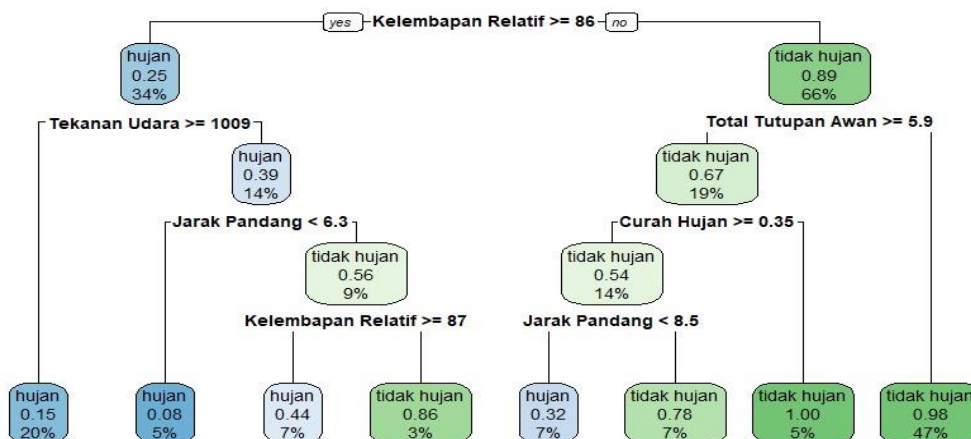
Klasifikasi Menggunakan Metode *Decision Tree*

Data yang dianalisis meliputi kondisi, temperatur, titik embun, kelembapan relatif, angin, tekanan udara, curah hujan, total tutupan awan, tutupan awan rendah, durasi penyinaran matahari, dan jarak pandang. Data tersebut diambil pada periode Oktober 2023 hingga September 2024 di Bandar Lampung. Penelitian ini menggunakan 366 data klasifikasi hujan di Bandar Lampung tahun 2024, yang dibagi menjadi 269 data pelatihan dan data pengujian. Data pelatihan (70%) dan data pengujian (30%) untuk evaluasi model. Pembagian data menjadi data pelatihan dan pengujian bertujuan meningkatkan validitas hasil analisis.

Plot *Decision Tree*

Model klasifikasi yang telah dibentuk dari data pelatihan (*training data*) divisualisasikan dalam bentuk pohon keputusan (*decision tree*) seperti ditampilkan pada Gambar 1. Visualisasi

ini bertujuan untuk memudahkan interpretasi hasil klasifikasi dan melihat struktur pengambilan keputusan berdasarkan variabel-variabel prediktor.



Gambar 1. Plot Decision Tree

Gambar 1 menunjukkan bahwa atribut yang paling berpengaruh dalam proses klasifikasi adalah kelembapan relatif, yang menjadi simpul utama (*root node*) dengan batas nilai 86%. Artinya, kelembapan relatif $\geq 86\%$ merupakan indikator awal yang digunakan model untuk membedakan kemungkinan terjadinya hujan dan tidak hujan. Nilai ambang 86% tersebut diperoleh melalui perhitungan algoritma *Decision Tree* menggunakan konsep *entropy* dan *information gain*. Algoritma mengurutkan data kelembapan relatif, kemudian menguji beberapa nilai kandidat sebagai titik potong. Untuk setiap kandidat, dihitung nilai *entropy* dan *information gain*, nilai ambang yang memberikan *information gain* tertinggi dipilih sebagai pemisah data terbaik. Dalam penelitian ini, ambang 86% terbukti menghasilkan pemisahan paling optimal antara kelas hujan dan tidak hujan, sehingga dipilih secara otomatis oleh algoritma sebagai simpul utama.

Jika kelembapan relatif tinggi ($\geq 86\%$), maka model selanjutnya mengevaluasi tekanan udara. Tekanan udara ≥ 1009 hPa diikuti oleh nilai jarak pandang dan kelembapan relatif tambahan untuk menentukan probabilitas hujan secara lebih spesifik. Misalnya, pada kondisi kelembapan tinggi dan tekanan udara ≥ 1009 hPa, peluang hujan meningkat hingga 39%. Namun, jika jarak pandang $< 6,3$ km dan kelembapan relatif $\geq 87\%$, maka model justru memprediksi tidak terjadi hujan, mengindikasikan bahwa hubungan antar variabel bersifat kompleks dan tidak selalu linier.

Sebaliknya, jika kelembapan relatif $< 86\%$, maka model cenderung memprediksi tidak terjadi hujan, dengan probabilitas sangat tinggi, yaitu 89%. Cabang selanjutnya menunjukkan bahwa total tutupan awan dengan ambang 5,9 menjadi faktor penting berikutnya. Total tutupan awan $\geq 5,9$ justru mendukung prediksi tidak hujan, yang secara ilmiah menunjukkan bahwa kondisi mendung belum tentu disertai hujan. Selain itu, variabel curah hujan dan jarak pandang digunakan sebagai pembeda tambahan untuk mendeteksi kemungkinan hujan dalam kondisi kelembapan rendah. Plot yang dihasilkan pada Gambar 1. memiliki hasil bahwa pada data pelatihan, variabel yang paling berpengaruh adalah kelembapan relatif, tekanan udara, total tutupan awan, jarak pandang, dan curah hujan terhadap klasifikasi hujan.

Evaluasi Kinerja Model Klasifikasi Menggunakan Data Pengujian

Model klasifikasi yang telah dibangun sebelumnya menggunakan data pelatihan (*training data*) kemudian diuji performanya menggunakan 30% bagian data klasifikasi tahun 2024 sebagai data pengujian (*testing data*). Proses pengujian ini bertujuan untuk mengetahui

kemampuan generalisasi model terhadap data baru yang belum pernah digunakan dalam pelatihan. Berikut hasil data prediksi hujan dan tidak hujan disajikan pada Tabel 1.

Tabel 1. Hasil Prediksi		
Prediksi	Hujan	Tidak Hujan
Hujan	21	12
Tidak Hujan	3	61

Berdasarkan hasil prediksi yang tercantum pada Tabel 1, diketahui bahwa model berhasil mengklasifikasikan 21 kejadian hujan dengan benar (*true positives*) dan 61 kejadian tidak hujan dengan benar (*true negatives*). Namun demikian, masih terdapat 12 kejadian tidak hujan yang salah diklasifikasikan sebagai hujan (*false positives*), serta 3 kejadian diprediksi tidak hujan tetapi sebenarnya hujan (*false negatives*). Berdasarkan data tersebut diperoleh nilai Akurasi, Kappa, P-Value, dan Sensitivitas disajikan pada Tabel 2.

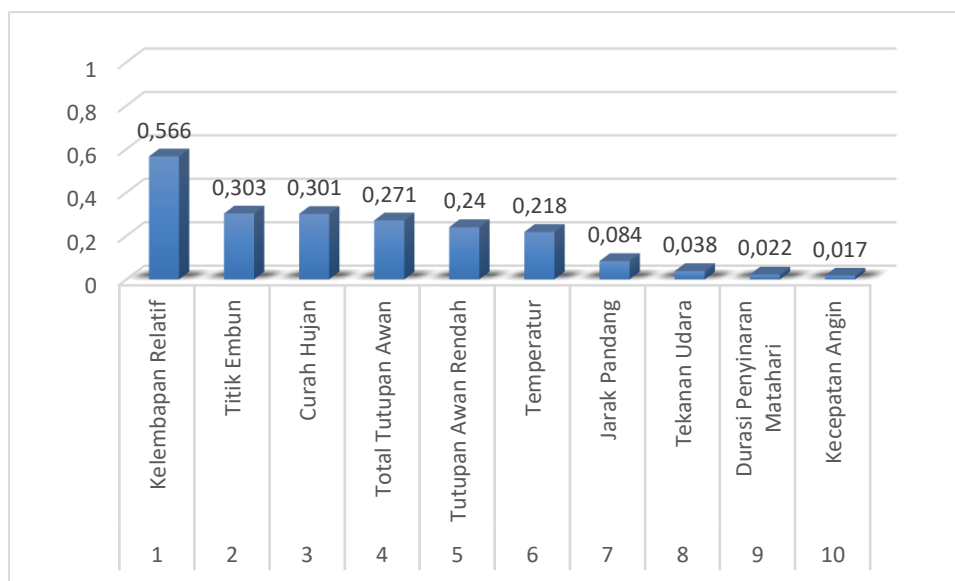
Tabel 2. Ringkasan Hasil Evaluasi Kinerja Model Klasifikasi

Metrik Evaluasi	Nilai
Akurasi	0,8454
Kappa	0,6312
P-Value	0,03887
Sensitivitas	0,8750

Berdasarkan hasil evaluasi kinerja model pada Tabel 2, diperoleh bahwa nilai akurasi sebesar 84,54% menunjukkan klasifikasi yang dilakukan pada data sudah cukup akurat, nilai kappa sebesar 63,12% menunjukkan bahwa konsistensi pengukuran yang dilakukan sudah baik, nilai sensitivitas sebesar 87,50% yang menunjukkan pengaruh setiap variabel X (*independent*) terhadap Y (*dependent*), dan nilai P-Value (alpha) sebesar 0,03887%. Pengaruh unsur-unsur cuaca terhadap hujan disajikan pada Tabel 3.

Tabel 3. Nilai Korelasi X dan Y		
No	Variabel X	Nilai
1	Kelembapan Relatif	0,566
2	Titik Embun	0,303
3	Curah Hujan	0,301
4	Total Tutupan Awan	0,271
5	Tutupan Awan Rendah	0,240
6	Temperatur	0,218
7	Jarak Pandang	0,084
8	Tekanan Udara	0,038
9	Durasi Penyinaran Matahari	0,022
10	Kecepatan Angin	0,017

Pada Tabel 3 menunjukkan pengaruh variabel X (unsur-unsur cuaca) terhadap variabel Y (hujan), diperoleh bahwa kelembapan relatif mempunyai pengaruh sebesar 0,566, titik embun sebesar 0,303, curah hujan sebesar 30,09%, total tutup awan sebesar 0,271, tutupan awan rendah sebesar 0,240, temperatur sebesar 0,218, jarak pandang sebesar 0,084, tekanan udara sebesar 0,038, durasi penyinaran matahari sebesar 0,022 dan kecepatan angin sebesar 0,017. Nilai korelasi ini juga disajikan pada Gambar 2.



Gambar 2. Diagram Batang Besaran Pengaruh Variabel X dan variabel Y

Berdasarkan Gambar 2, masing-masing variabel X menunjukkan kontribusi yang berbeda terhadap variabel Y. Kelembapan relatif memberikan pengaruh paling dominan dengan kontribusi sebesar 0,566, sedangkan variabel angin menunjukkan pengaruh paling rendah, yaitu sebesar 0,017 terhadap variabel Y. Secara ilmiah, dominasi pengaruh kelembapan relatif ini dapat dijelaskan melalui karakteristik fisis atmosfer. Kelembapan udara yang tinggi cenderung berkorelasi kuat dengan peningkatan potensi pembentukan awan dan curah hujan. Sebaliknya, rendahnya kontribusi variabel angin dapat mengindikasikan bahwa dalam periode dan lokasi pengamatan yang dikaji, variasi kecepatan atau arah angin tidak memberikan pengaruh signifikan terhadap dinamika variabel Y. Meskipun angin memiliki peran penting dalam transportasi uap air secara horizontal, namun pengaruhnya terhadap pembentukan hujan atau variabel meteorologis lainnya dapat bervariasi tergantung skala spasial dan temporal kajian.

SIMPULAN

Dari hasil penelitian yang dilakukan terhadap data hujan harian di wilayah Bandar Lampung, beberapa kesimpulan yang dapat diambil yaitu bahwa dengan menggunakan metode *decision tree* dalam mengklasifikasikan variabel-variabel yang mempengaruhi terjadinya hujan di Bandar Lampung maka kelembapan relatif mempunyai pengaruh sebesar 0,566. Kemudian diikuti oleh titik embun sebesar 0,303 dan curah hujan sebesar 0,301. Hal ini menandakan bahwa ketiga faktor ini perlu dipertimbangkan ketika melakukan prediksi terjadinya hujan. Selain itu, nilai akurasi terhadap model pohon keputusan yang dibentuk diperoleh sebesar 84,54%. Persentase yang cukup besar ini menandakan bahwa model pohon keputusan yang telah dibangun sudah cukup baik. Meskipun akurasi model pohon keputusan mencapai 84,54%, model tersebut masih memiliki tingkat kesalahan sebesar 15,46%.

DAFTAR PUSTAKA

- [1] R. Sulistyiorini, "Alternatif penanganan permasalahan infrastruktur kebutuhan air bersih di Kota Bandar Lampung melalui rain water harvesting," *Journal of Systems Integration*, Dec. 2020, doi: 10.23960/JSI.V1I1.3.
- [2] N. Dercas, "Editorial to the special issue 'Drought and water scarcity: Monitoring, modelling and mitigation'," *Hydrology*, Jun. 2023, doi: 10.3390/hydrology10060134.

- [3] M. Safdar et al., "Drought monitoring with multiple indices and management through various techniques: A review," *Proceedings of the ASECS 2023*, Dec. 2023, doi: 10.3390/asec2023-16602.
- [4] A. E. Kenawy, S. M. Vicente-Serrano, C. Murphy, and L. Gimeno, "Editorial: Advances in drought analytical tools for better understanding of current and future climate change," *Frontiers in Earth Science*, Feb. 2023, doi: 10.3389/feart.2023.1140658.
- [5] B. Susilo, "Mengenal Iklim dan Cuaca di Indonesia". Yogyakarta: DIVA Press, 2021.
- [6] P. B. Setio, D. R. Saputro, and B. Winarno, "Klasifikasi dengan pohon keputusan berbasis algoritme C4.5," *Prisma, Prosiding Seminar Nasional Matematika*, vol. 3, pp. 65–68, 2020.
- [7] R. Oktavia, J. T. Hardinata, and I. Irawan, "Penerapan metode algoritma K-means dalam pengelompokan angka harapan hidup saat lahir menurut provinsi," *Kesatria: Jurnal Penerapan Ilmu dan Teknologi*, vol. 1, no. 4, pp. 154–161, Okt. 2020, doi: 10.30645/kesatria.v1i4.41
- [8] A. Rodionov and M. Mukhitdinova, "The role of a decision tree as a method of artificial intelligence for the analysis of big data problems," *Economics and Innovative Technologies*, vol. 11, no. 2, pp. 400–407, Apr. 2023, doi: 10.55439/eit/vol11_iss2/i39.
- [9] K. G. Komarasamy and T. N. Ravishankar, "The application of decision tree method for data mining," *Technoarete Transactions on Intelligent Data Mining and Knowledge Discovery (TTIDMKD)*, vol. 2, no. 3, Oct. 2022, doi: 10.36647/ttidmkd/02.03.a005.
- [10] N. Nurliana, N. Suarna, and W. Prihatono, "Analisis keputusan pembelian produk terhadap promosi dan viral marketing di TikTok Shop menggunakan algoritma C4.5," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8 no.3, pp. 2871–2872, 2024. 10.36040/jati.v8i3.9583.
- [11] U. M. Pratiwi and M. Ibad, "Klasifikasi faktor yang berpengaruh dalam kehamilan tidak diinginkan menggunakan metode algoritma decision tree," *Lebesque*, vol. 3, no. 2, pp. 406–416, Agus. 2022, doi: 10.46306/lb.v3i2.129
- [12] N. H. Ibrahim dan L. Khikmah, "Perbandingan metode algoritma C4.5, Naïve bayes dan logistic regression untuk penentuan kelayakan penerimaan kredit". Unipdu, vol. 4. no. 2 pp. 89-93. 2024.
- [13] S. Moslehi, N. Rabiei, A. R. Soltanian, and M. Mamani, "Application of machine learning models based on decision trees in classifying the factors affecting mortality of COVID-19 patients in Hamadan, Iran," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, Jul. 2022, doi: 10.1186/s12911-022-01939-x.
- [14] Y. Duan et al., "Application of the decision tree method to lithology identification of volcanic rocks-taking the Mesozoic in the Laizhouwan Sag as an example," *Scientific Reports*, vol. 10, no. 1, p. 19209, Nov. 2020, doi: 10.1038/s41598-020-76303-y.
- [15] S. Samadianfard, F. Mikaeilia and R. Prasad, "Evaluation of classification and decision trees in predicting daily precipitation occurrences," *Water Supply*, vol. 22, no. 4, pp. 3879–3895, 2022, doi: 10.2166/ws.2022.017.
- [16] A. Nugroho, "Analisa Splitting Criteria pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan," *JSITIK*, vol. 1, no. 1, pp. 41–49, 2022. <https://doi.org/10.53624/jsitik.v1i1.154>.
- [17] E. Muningsih et. al., "Optimized Decision Tree Classification Method for Diabetes Prediction," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 22, no. 4, pp. 941–948, 2024. doi: 10.12928/TELKOMNIKA.v22i4.25656.
- [18] D.T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley and Sons, Inc., New Jersey, 2005.