

# Penerapan Algoritma Decision Tree Untuk Kategorisasi Curah Hujan Di Kecamatan Gurah

<sup>1\*</sup>Mayo Alvarosy Chrisnatae, <sup>2</sup>Umi Mahdiyah, <sup>3</sup>Resty Wulanningrum

<sup>1</sup> Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: <sup>1</sup>[mayorosy150@gmail.com](mailto:mayorosy150@gmail.com)

*Penulis Korespondens : Mayo Alvarosy Chrisnatae*

**Abstrak** Prediksi curah hujan memainkan peran penting dalam perencanaan pertanian, terutama di wilayah yang memiliki pola iklim yang tidak menentu. Penelitian ini bertujuan untuk mengelompokkan intensitas curah hujan di Kecamatan Gurah menggunakan algoritma klasifikasi pohon keputusan. *Dataset* yang digunakan berasal dari tahun 2015 hingga 2024, yang mencakup fitur-fitur seperti suhu rata-rata, kelembapan, dan data curah hujan bulanan. Langkah praproses mencakup penanganan data yang hilang serta transformasi data waktu kategorikal menjadi format numerik. Model *Decision Tree* dilatih menggunakan data tahun 2015 hingga 2023, dan diuji dengan data tahun 2024. Model klasifikasi ini mencapai akurasi sebesar 73,3%, presisi 75,1%, *recall* 73,3%, dan skor F1 sebesar 73,4%. Hasil ini menunjukkan bahwa klasifikasi menggunakan *Decision Tree* merupakan metode yang layak untuk mengelompokkan tingkat curah hujan ke dalam tiga kategori: rendah, sedang, dan tinggi. Temuan ini memberikan dasar untuk penerapan lebih lanjut dalam alat bantu pengambilan keputusan pertanian guna mengurangi risiko yang disebabkan oleh variabilitas cuaca.

**Kata Kunci**— Klasifikasi Curah Hujan, Pohon Keputusan, *Data Mining*, Prediksi Cuaca

**Abstract** Rainfall prediction plays a critical role in agricultural planning, especially in regions with unpredictable climate patterns. This study aims to categorize rainfall intensity in Gurah District using the Decision Tree classification algorithm. The dataset, sourced from 2015 to 2024, includes features such as average temperature, humidity, and monthly rainfall records. Preprocessing steps involved handling missing values and transforming categorical time data into numerical format. The Decision Tree model was trained using data from 2015 to 2023 and tested on data from 2024. The classification model achieved an accuracy of 73.3%, a precision of 75.1%, a recall of 73.3%, and an F1-score of 73.4%. These results indicate that Decision Tree classification is a viable method for categorizing rainfall levels into three distinct classes: low, medium, and high. The findings provide a basis for future implementation in agricultural decision-making tools to mitigate risks caused by weather variability..

**Keywords**— Rainfall Classification, Decision Tree, Data Mining, Weather Prediction

This is an open access article under the CC BY-SA License.



## I. PENDAHULUAN

Indonesia, sebagai negara tropis yang terletak di garis khatulistiwa, memiliki dua musim utama, yaitu musim kemarau dan musim penghujan. Curah hujan, yang diukur dalam satuan milimeter (mm), merupakan salah satu komponen penting bagi keberlangsungan ekosistem dan sektor pertanian. Beberapa tahun belakangan ini, teknologi *data mining* telah menjadi alat yang diperlukan dalam menganalisis data besar untuk mendapatkan pola dan tren yang tersembunyi. Tantangan utama tetap pada pemilihan metode yang tepat untuk meminimalkan kesalahan

prediksi dan meningkatkan akurasi, khususnya dalam mengelola data cuaca yang bersifat temporal dan kompleks[1]. Curah hujan adalah salah satu jenis cuaca yang diprediksi oleh Badan Meteorologi Klimatologi dan Geofisika (BMKG). Dari data curah hujan yang terdapat di *database* tentu saja ada sejumlah variabel-variabel prediktor yang bisa dipakai untuk memperkirakan variabel target terdiri dari suhu rata-rata dan kelembaban. [2]

Proses prediksi curah hujan dibuat menggunakan bahasa pemrograman yaitu *python* [3]. Karena cuaca ini banyak parameter yang mempengaruhi sehingga para petani sulit meramalkannya. Sering petani di pedesaan mengalami gagal panen sehingga mengalami kerugian yang besar, [4] Proses cuaca penentuan menggunakan metode *Decision Tree* untuk menghasilkan cuaca penentuan keluaran.

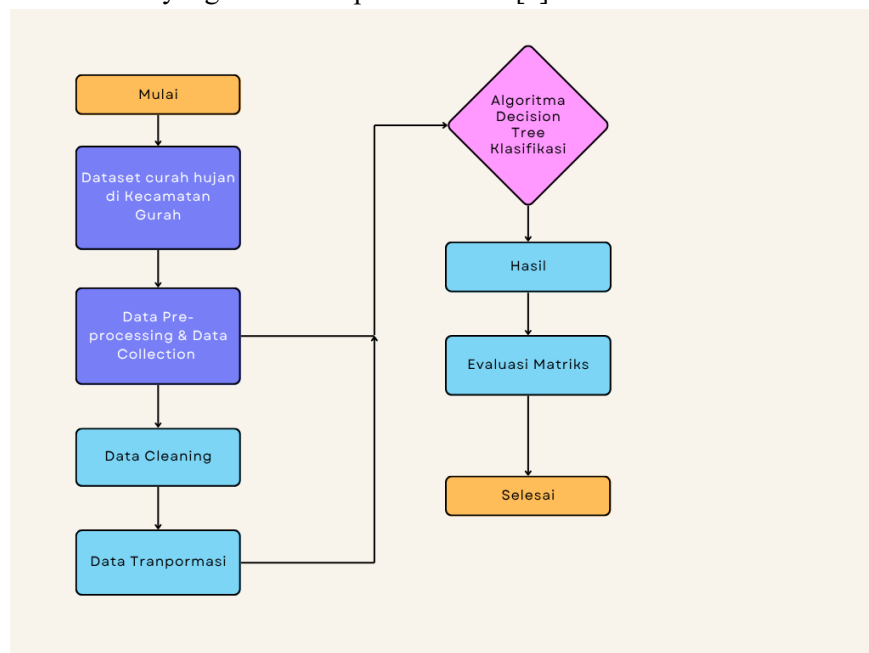
Curah hujan merupakan jumlah air yang jatuh di permukaan tanah datar yang diukur dengan satuan tinggi milimeter (mm). Curah hujan juga dapat diartikan sebagai ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap dan tidak mengalir. Perubahan iklim atau *climate change* akan mengikuti pemanasan *global*, seperti peningkatan curah hujan di beberapa belahan dunia yang menyebabkan banjir dan erosi.

*Decision Tree* adalah metode klasifikasi berstruktur pohon yang digunakan untuk memprediksi nilai dari variabel target berdasarkan fitur-fitur yang ada. [5] Di antara berbagai algoritma klasifikasi, peneliti memilih algoritma C4.5 atau *Decision Tree*, karena yang paling populer dan efektif.

## II. METODE

### 2.1 Tahap Penelitian

Penelitian ini menggunakan metode *Decision Tree* Klasifikasi yang bertujuan untuk memecahkan masalah yang ada dalam penelitian ini [4].



Gambar 1. Metode Penelitian

Pada Gambar 1. Menunjukkan alur diagram untuk penelitian kategori curah hujan menggunakan metode *Decision Tree* klasifikasi. Dimulai dari mengumpulkan *dataset* curah hujan, mempersiapkan *dataset*, mengecek dan mengubah data dengan nominal (angka) bila perlu, pemrosesan data menggunakan algoritma *Decision Tree*, mendapatkan hasil dari algoritma tersebut, dan yang terakhir mengevaluasi hasil *dataset*.

## 2.2 Pengumpulan *Dataset*

Tahap penelitian ini dimulai dengan pengambilan *dataset* curah hujan di Kecamatan Gurah mulai dari tahun 2015 sampai dengan 2024 lewat situs web resmi Badan Pusat Statistik (<https://kedirikab.bps.go.id/id/publication/2024/09/26/779e0a577ecf4bb7441ad961/kecamatan-gurah-dalam-angka-2024.html>) . *Dataset* curah hujan di Kecamatan Gurah dari tahun 2015 sampai 2024 yang akan digunakan untuk memprediksi curah hujan.

Table 1. *Dataset* curah hujan

Nama Atribut	Keterangan
Tahun	Periode waktu selama 12 bulan
Bulan	Bagian dari tahun, digunakan untuk melihat variasi data cuaca dalam skala lebih kecil
Curah Hujan	Jumlah air hujan yang jatuh di suatu wilayah dalam satuan milimeter (mm)
Kelembapan	Tingkat kandungan uap air di udara
Suhu rata - rata	Nilai tengah suhu udara
Kategori Curah Hujan	Pengelompokkan intensitas hujan berdasarkan jumlahnya

Pada Tabel 1, merupakan contoh dataset yang akan kami lakukan untuk penelitian menentukan kategori curah hujan dengan menggunakan pohon keputusan klasifikasi.

## 2.3 Data *Cleaning*

Untuk penelitian ini hanya dilakukan *preprocessing* secara umum yang tahapannya dengan melakukan pengecekan terhadap ada atau tidaknya *missing value*, dari data yang akan digunakan untuk penelitian [6].

C	D	E	F	G
bulan	curah hujan	kelembapan	suhu rata-rata	kategori
Januari	237	82	26	
Februari	252	82	26	
Maret	434	82	26	
April	211	83	26	
Mei	168	82	26	
Juni		82	26	
Juli		82	26	
Agustus		90	27	
September		69	28	
Oktober	2	68	29	
November	6	70	29	
Desember	12	80	29	
Januari	289	79	26	
Februari	401	80	26	
Maret	271	81	26	

Gambar 2. *Missing value*

Pada Gambar 2, terdapat data yang kosong atau *missing value*. Sebelum melakukan data *train* dan data *test*, kita harus mengecek datanya terlebih dahulu apakah ada data kosong (*missing value*) atau tidak. Hal ini akan berdampak pada hasil evaluasinya termasuk nilai akurasi dan lain-lain.

#### 2.4 Data Transformasi

Dataset yang telah dilakukan cleaning, akan melakukan perubahan pada atribut yang bersifat angka dengan jenis data yaitu Integer [7]. Sebagai contoh pada gambar dibawah ini.

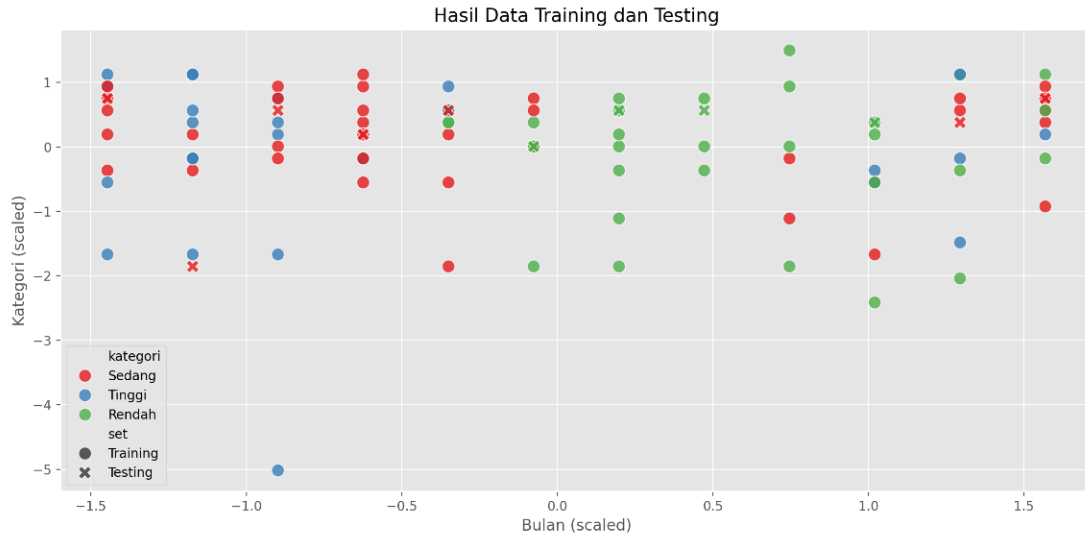
Table 2. Hasil transformasi

Sebelum Transformasi	Sesudah Transformasi
Januari	1
Februari	2
Maret	3
April	4

Pada Tabel 2, merupakan contoh mentranferkan data ke dalam numerik (angka), mengubah nama bulan menjadi angka 1 sampai 12. Hal ini sangat umum di kalangan algoritma klasifikasi untuk digunakan sebagai *input* numerik (angka) model.

## 2.5 Proses *Decision Tree* Klasifikasi

Proses klasifikasi dengan menggunakan algoritma *Decision Tree* diawali dengan pembagian data menjadi data latih (*training*) dan data uji (*testing*) [8]. Sebagai contoh pada gambar dibawah ini.



Gambar 3. Hasil data *train* dan *test*

Pada Gambar 3, menunjukkan sebaran data hasil tranformasi dan pembagian *dataset* menjadi data *training* dan *testing*, berdasarkan fitur bulan dan kategori yang telah melalui proses standardisasi atau transformasi. Setiap titik pada grafik dibedakan berdasarkan kategori curah hujan: Rendah (hijau), Sedang (merah), dan Tinggi (biru). Selain warna, bentuk titik juga mempresentasikan pembagian data: simbol bulat menandakan data *training* (tahun 2015 sampai 2023) dan simbol silang menandakan data *testing* (tahun 2024).

Melalui visualisasi Gambar 3, terlihat bahwa data *testing* (simbol silang) mencerminkan pola yang serupa dengan data *testing* (simbol bulat). Hal ini menandakan bahwa model berjalan dengan baik serta keberhasilan terhadap *preprocessing* dan pembagian data yang dilakukan sebelum pelatihan model klasifikasi *Decision Tree*.

## 2.6 Evaluasi Model

Dalam *confusion matrix*, terdapat nilai-nilai acuan yang digunakan untuk membandingkan metode, yaitu nilai akurasi, presisi, *recall*, dan *F-1 score* yang juga dihasilkan dari masing-masing metode. Pengertian dan perhitungan dari masing-masing acuan adalah sebagai berikut [8]:

Akurasi adalah ukuran umum yang digunakan untuk mengevaluasi kinerja model klasifikasi. Ini mengukur seberapa sering model benar-benar memprediksi kelas dengan benar, baik positif maupun negatif. Semakin tinggi nilai akurasi, semakin baik model dalam membuat prediksi yang benar [9].

$$\text{Akurasi} = \frac{TP + TN}{TP + FP + FN + TN} \dots \dots \dots (1)$$

*Recall* dapat didefinisikan sebagai rasio jumlah total positif diklasifikasikan dengan benar membagi jumlah total positif [4].

$$\text{Recall} = \frac{TP}{TP + FP} \dots \dots \dots (2)$$

*Precision* memberikan proporsi prediksi positif yang benar terhadap semua prediksi positif. *Precision* yang tinggi berarti sebagian besar prediksi positif model adalah benar [4].

$$Precision = TP / TP + FP \dots\dots\dots (3)$$

*F1 Score* merupakan rata-rata harmonis antara *precision* dan *recall*. *F1 Score* mencapai nilai terbaik di 1 dan nilai terburuk di 0 [9].

$$F1\ Score = 2 \times Precision \cdot Recall / Precision + Recall \dots\dots\dots (4)$$

### III. HASIL DAN PEMBAHASAN

*Decision Tree* adalah struktur *flowchart* yang menyerupai *tree* (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes [5]. Metode *Decision Tree* klasifikasi dapat diterapkan untuk melakukan prediksi curah hujan yang dapat digunakan untuk memperoleh informasi dalam perencanaan pengelolaan resiko banjir.

#### 3.1 Hasil Evaluasi Model

Tahapan ini dilakukan dalam menguji hasil estimasi dan memilih metode untuk melakukan klasifikasi data awal [7]. Kebanyakan data yang dibutuhkan untuk evaluasi model ini adalah akurasi, presisi, *recall*, dan *F1-Score*.

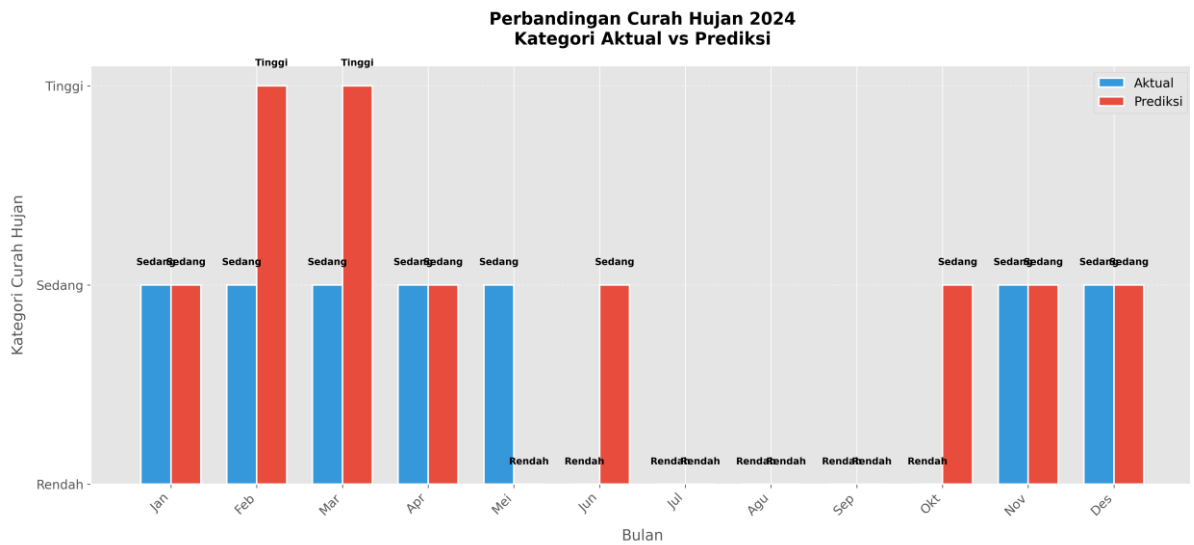
Tabel 3. Hasil evaluasi model

<i>Metric</i>	<i>Value</i>
<i>Accuracy</i>	0.733
<i>Precision</i>	0.751
<i>Recall</i>	0.733
<i>F1-Score</i>	0.734

Pada Tabel 3. Memperllihatkan hasil dari evaluasi model *Decision Tree* klasifikasi, yang dimana terdapat nilai *accuracy* 0.733, *precision* 0.751, *recall* 0.733, dan *F1-Score* 0.734 hasil tersebut terbilang cukup baik. Karena nilai – nilai tersebut berada di kisaran 0,70 (70%) ke atas. Jika nilai akurasi 0,90 (90%), maka nilai tersebut dianggap sangat baik, karena model tersebut dapat memprediksi 9 dari 10 kejadian dengan benar. Jika nilai presisi, *recall*, dan *F1-Score* mencapai 0,80 (80%), maka nilai tersebut menunjukkan kinerja yang kuat.

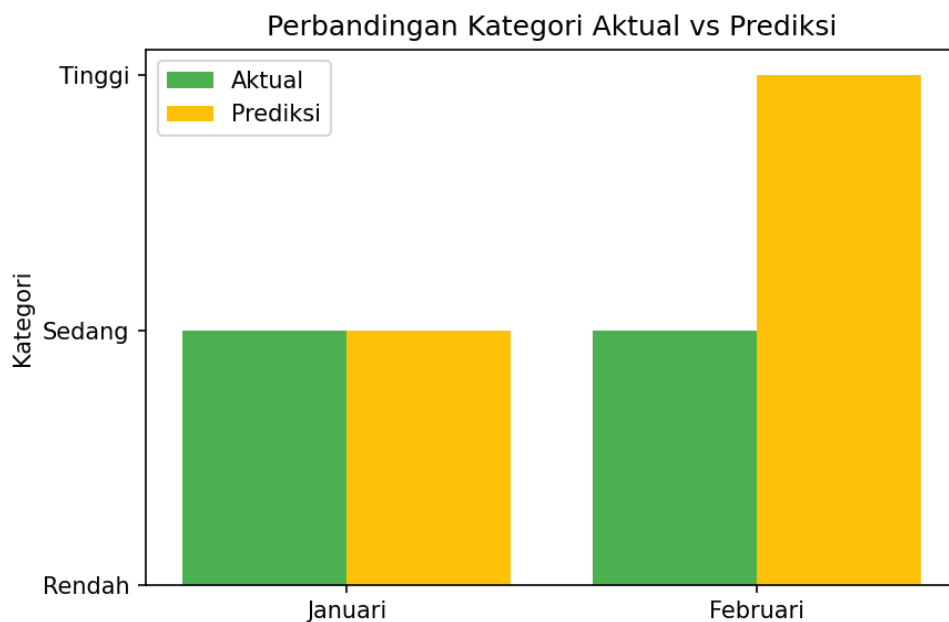
#### 3.2 Hasil Prediksi

Gambar dibawah ini adalah perbandingan antara satu tahun dan per bulan dari metode *Decision Tree* Klasifikasi.



Gambar 4. Hasil perbandingan satu tahun

Pada Gambar 4. Memperlihatkan hasil perbandingan kategori aktual dan prediksi selama satu tahun, yang dimana pada gambar grafik tersebut nilai aktual berwarna biru dan nilai prediksi berwarna merah. Kita bisa melihat adanya perbedaan di antara kedua nilai tersebut dimulai dari bulan februari, maret, mei, juni, dan oktober.

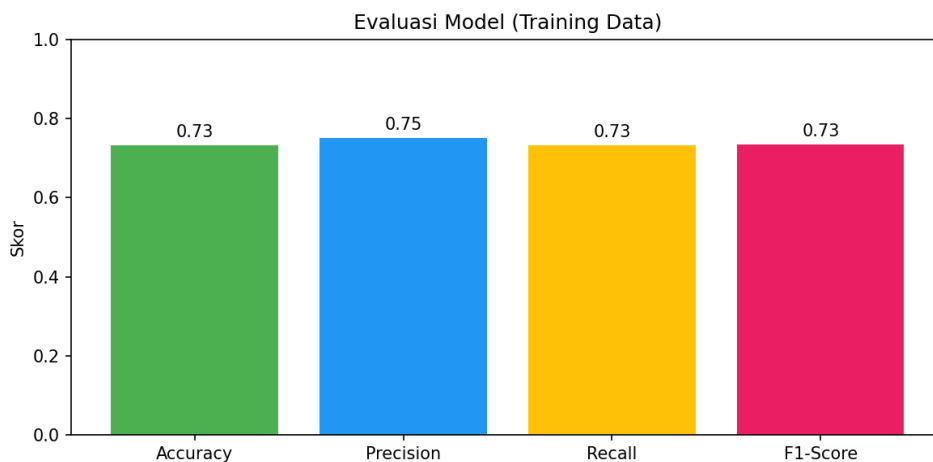


Gambar 5. Hasil perbandingan per bulan

Pada Gambar 5. Memperlihatkan hasil perbandingan kategori aktual dan prediksi diambil per bulan, yang dimana pada gambar grafik tersebut nilai aktual berwarna hijau dan nilai prediksi berwarna kuning.

Terdapat dua gambar yang berbeda, pada Gambar 4 menunjukkan hasil prediksi dari 1 (satu) tahun, sedangkan Gambar 5 menunjukkan hasil prediksi 2 (dua) bulan. Kedua gambar tersebut menunjukkan perbandingan dari hasil nilai aktual (nilai asli) dan nilai prediksi (nilai hasil dari prediksi model klasifikasi pohon keputusan).

### 3.3 Analisis Hasil



Gambar 6. Grafik evaluasi model

Pada Gambar 6. Memperlihatkan grafik hasil evaluasi model, yang terdiri dari grafik *accuracy* (berwarna hijau) nilai 0.73 (73%), grafik *precision* (berwarna biru) nilai 0.75 (75%), grafik *recall* (berwarna kuning) nilai 0.73 (73%), dan grafik *F1-Score* (berwarna merah) nilai 0.73 (73%).

## IV. KESIMPULAN

Berdasarkan hasil penelitian diatas, evaluasi penelitian menggunakan *confusion matrix* dengan melihat beberapa factor. [4] *Accuracy* salah satu yang paling populer digunakan, sebelum membuat klasifikasi kita harus mengukur kinerja algoritme yang akan digunakan. Nilai – nilai tersebut terdiri dari *Accuracy* 0.733, *Precision* 0.751, *Recall* 0.733, *F1-Score* 0.734. Dalam prediksi curah hujan menggunakan metode pohon keputusan klasifikasi, cukup membantu untuk memahami algoritma dan jenis data yang digunakan. Sebelum membangun aplikasi prediksi curah hujan, seharusnya melihat hasil penelitian ini apakah akurasi tersebut tinggi atau rendah. Dengan banyaknya metode – metode tersebut mulai dari *Machine Learning* hingga ke *Deep Learning*. Dalam prediksi curah hujan menggunakan metode *Decision Tree* Klasifikasi.

Hasil akurasi cenderung menurun ketika *prosentase data training* berkurang. Hal ini membuktikan bahwa semakin besar *data training*, maka akurasi akan semakin tinggi. Karena semakin banyak *data training* yang digunakan, model akan semakin banyak mengenali pola dari data [10].



## DAFTAR PUSTAKA

- [1] M. I. Pradipta, “Klasifikasi Curah Hujan Menggunakan Metode Ensemble Subset K-Nearest Neighbor,” p. 67, 2020.
- [2] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [3] N. Nur, F. Wajidi, S. Sulfayanti, and W. Wildayani, “Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga,” *J. Komput. Terap.*, vol. 9, no. 1, pp. 58–64, 2023, doi: 10.35143/jkt.v9i1.5917.
- [4] Gelinas, Ulric, Oram, Alan, Wiggins, and William, “Accounting Information System,” pp. 17–30, 1990.
- [5] S. Shalsabilla, P. Rachmawati, K. Vidya Prakusa, and S. Rihastuti, “Penerapan Data Mining dengan Metode Decision Tree untuk Prediksi Cuaca di Kota Seattle menggunakan Aplikasi Weka,” *Semin. Nas. Amikom Surakarta*, no. November, pp. 93–100, 2023.
- [6] S. R. Aisy, M. K. Ramadhan, and A. S. Salsabila, “Perbandingan Algoritma Klasifikasi Data Mining dalam Memprediksi Curah Hujan di Jawa Barat,” vol. 2024, no. Senada, pp. 180–192, 2024.
- [7] M. SILAT and E. DITA, “Penerapan Algoritma Decision Tree C4. 5 Untuk,” *Simki.Unpkediri.Ac.Id*, pp. 45–52, 2025, [Online]. Available: [http://simki.unpkediri.ac.id/mahasiswa/file\\_artikel/2018/12.1.03.02.0331.pdf](http://simki.unpkediri.ac.id/mahasiswa/file_artikel/2018/12.1.03.02.0331.pdf)
- [8] A. Khoirunnisa, N. Y. Putri, K. Marsell, and V. Raiqhan, “Penerapan Machine Learning dalam Klasifikasi Kejadian Hujan di Kabupaten Tuban Tahun 2019-2024,” vol. 2024, no. April, pp. 101–113, 2024.
- [9] F. Rahmasari, M. Rifany, T. Priharyanto, and R. Kurniawan, “Perbandingan Algoritma Machine Learning dalam Klasifikasi Status Banjir di Sumatera Utara,” vol. 2024, no. Senada, pp. 307–318, 2024.
- [10] A. Nugroho, “Analisa Splitting Criteria Pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan,” *JSITIK J. Sist. Inf. dan Teknol. Inf. Komput.*, vol. 1, no. 1, pp. 41–49, 2022, doi: 10.53624/jsitik.v1i1.154.