KLASIFIKASI CURAH HUJAN MENGGUNAKAN ALGORITMA *DECISION TREE C4.5*

SKRIPSI

Oleh :
ABDURRAHMAN WAHID KHASSOGI
NIM. 18650070

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

KLASIFIKASI CURAH HUJAN MENGGUNAKAN ALGORITMA *DECISION TREE C4.5*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
ABDURRAHMAN WAHID KHASSOGI
NIM. 18650070

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

HALAMAN PERSETUJUAN

KLASIFIKASI CURAH HUJAN MENGGUNAKAN ALGORITMA *DECISION TREE C4.5*

SKRIPSI

Oleh :
ABDURRAHMAN WAHID KHASSOGI
NIM. 18650070

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 5 Juni 2025

Pembimbing I,	Pembimbing II,
Dr. Zainal Abidin, M.Kom	Tri Mukti Lestari, M.Kom
NIP. 19760613 200501 1 004	NIP. 19911108 202012 2 005

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Ir. Fachrul Kurniawan S.T., M.MT, IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI CURAH HUJAN MENGGUNAKAN ALGORITMA *DECISION TREE C4.5*

SKRIPSI

Oleh :
ABDURRAHMAN WAHID KHASSOGI
NIM. 18650070

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 25 Juni 2025

Susunan Dewan Penguji

Ketua Penguji	: Dr. Ir. Fachrul Kurniawan S.T., M.MT., IPU NIP. 19771020 200912 1 001	( )
Anggota Penguji I	: Khadijah F.H. Holle, M.Kom NIP. 19900626 202203 2 002	( )
Anggota Penguji II	: Dr. Zainal Abidin, M.Kom NIP. 19760613 200501 1 004	( )
Anggota Penguji III	: Tri Mukti Lestari, M.Kom NIP. 19911108 202012 2 005	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan S.T., M.MT, IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Abdurrahman Wahid Khassogi
NIM	: 18650070
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Klasifikasi Curah Hujan Menggunakan Algoritma *Decision Tree C4.5.*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 25 Juni 2025
Yang membuat pernyataan,

Abdurrahman Wahid Khassogi
NIM.18650070

MOTTO

لَئِن شَكَرْتُمْ لأَزِيدَنَّكُمْ

... Sesungguhnya jika kamu bersyukur, pasti Kami akan menambah (nikmat) kepadamu ...

(QS. Ibrahim :7)

HALAMAN PERSEMBAHAN

Karya ini dipersembahkan dengan penuh rasa syukur dan hormat kepada:

Allah SWT, Rabb semesta alam,
Yang telah memberikan segala nikmat iman, kesehatan, ilmu, serta kekuatan hingga skripsi ini dapat terselesaikan.

Nabi Muhammad SAW, suri teladan sepanjang zaman,
Yang telah menjadi teladan hidup dan inspirasi bagi umat muslim.

Ayah saya La Ode Amsir, Ibu saya Hamriah dan Seluruh keluarga tercinta,
Yang senantiasa mendoakan dan menyemangati penulis sampai titik ini.

Dr. Zainal Abidin, M.Kom dan Tri Mukti Lestari, M.Kom
Yang dengan penuh kesabaran dan keikhlasan telah membimbing dan mengarahkan penulis dalam menyusun skripsi ini.

Teman-teman seperjuangan,
Yang telah menjadi bagian dari proses panjang ini dengan doa dan semangat yang tak ternilai.

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh.

Puji syukur penulis panjatkan atas kehadirat Allah SWT yang telah melimpahkan rahmat dan kasih sayang-Nya, sehingga penulis mampu menyelesaikan skripsi ini dengan baik dan tepat waktu. Sholawat serta salam semoga tercurahkan kepada Nabi besar Muhammad SAW, atas syafaatnya yang telah menuntun umat manusia dari jalan kebatilan ke jalan yang benar. Semoga kita semua masuk ke dalam golongan yang dituntun Allah SWT serta mendapatkan pertolongan Nabi Muhammad SAW. *Aamiin.*

Selanjutnya penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dan mendukung terselesaikannya skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada :

1. Prof. Dr. M. Zainuddin, MA selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan S.T., M.MT., IPU, selaku Ketua Program Studi Teknik Informatika Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus dosen ketua penguji yang telah berkenan mengoreksi dan memberikan saran untuk kebaikan penulis.

4. Dr. Zainal Abidin, M.Kom dan Tri Mukti Lestari, M.Kom, selaku pembimbing I dan II yang telah meluangkan waktu untuk bimbingan dan memberikan arahan.
5. A'la Syauqi, M.Kom, selaku wali dosen yang telah berkenan mengoreksi dan memberikan saran untuk kebaikan penulis.
6. Khadijah F.H. Holle, M.Kom, selaku dosen penguji satu yang telah berkenan meluangkan waktunya untuk membimbing dan memberikan arahan.
7. Nia Faricha, S.Si sebagai Admin Program Studi Teknik Informatika yang dengan sangat sabar banyak membantu penulis terkait informasi administrasi dan berkas studi.
8. Segenap civitas akademika Jurusan Teknik Informatika, terutama seluruh dosen, terima kasih atas ilmu dan bimbingan yang telah diberikan.
9. Kedua orang tua saya Alm. La Ode Amsir dan Alm. Hamriah, seluruh keluarga tercinta yang senantiasa memberikan do'a dan dukungan baik secara moral dan materi sehingga penulis dapat menyelesaikan skripsi.
10. Saudara Teknik Informatika UFO *(Unity of Informatics Force)* angkatan 2018 yang selalu memberikan semangat dan bantuan selama menyusun skripsi. Dan juga untuk rekan-rekan sesama last batch, atas dukungan dan kebersamaan di sisa perjalanan studi yang begitu berarti bagi penulis.

11. Teman-teman discord yang selalu memberikan dukungan dan bantuan selama menyusun skripsi.

Malang, 25 Juni 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL ..... **i**
HALAMAN PENGAJUAN ..... **ii**
HALAMAN PERSETUJUAN ..... **iii**
HALAMAN PENGESAHAN ..... **iv**
PERNYATAAN KEASLIAN TULISAN ..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
KATA PENGANTAR ..... **viii**
DAFTAR ISI ..... **xi**
DAFTAR GAMBAR ..... **xiii**
DAFTAR TABEL ..... **xiv**
ABSTRAK ..... **xv**
ABSTRACT ..... **xvi**
مستخلص البحث ..... **xvii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 4
BAB II STUDI PUSTAKA **5**
2.1 Penelitian Terkait 5
2.2 *Machine Learning* (ML) 8
2.3 Klasifikasi 9
2.4 Metode *Decision Tree* 10
2.4.1 Algoritma *C4.5* 11
2.5 Curah Hujan 14
2.5.1 Prediksi Curah Hujan 15
2.6 Evaluasi model dengan *Confusion Matrix* 16
BAB III DESAIN DAN IMPLEMENTASI **19**
3.1 Desain Penelitian 19
3.2 Pengumpulan Data 20
3.3 *Preprocessing* Data 21
3.4 Desain Sistem 22
3.5 Perhitungan Manual Algoritma *C4.5* 23
3.6 Skenario Pengujian 28
3.6.1 *Confusion matrix* 29
BAB IV HASIL DAN PEMBAHASAN **31**
4.1 Implementasi Sistem 31
4.2 Hasil Uji Coba 37
4.2.1 Visualisasi *Decision Tree* 44
4.3 Pembahasan 49
4.4 Integrasi Islam 51

4.4.1 Mu'amalah Ma'a Allah 51
4.4.2 Mu'amalah Ma'a An-Nas 52
4.4.3 Mu'amalah Ma'a Alam 53
BAB V KESIMPULAN DAN SARAN 55
5.1 Kesimpulan 55
5.2 Saran 55
DAFTAR PUSTAKA
LAMPIRAN

DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian 19
Gambar 3.2 *Flowchart* Desain Sistem 22
Gambar 4.1 Visualisasi Proses Distribusi dan Proporsi Target 34
Gambar 4.2 Hasil Pengujian Model 1 38
Gambar 4.3 Hasil Pengujian Model 2 39
Gambar 4.4 Hasil Pengujian Model 3 40
Gambar 4.5 Hasil Pengujian Model 4 42
Gambar 4.6 Hasil Pengujian Model 5 43
Gambar 4.7 Visualisasi Pohon Keputusan (Rasio 90:10) 45
Gambar 4.8 Visualisasi Pohon Keputusan (Rasio 80:20) 45
Gambar 4.9 Visualisasi Pohon Keputusan (Rasio 70:30) 46
Gambar 4.10 Visualisasi Pohon Keputusan (Rasio 60:40) 47
Gambar 4.11 Visualisasi Pohon Keputusan (Rasio 50:50) 48

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait 5
Tabel 2.2 Struktur *Confusion Matrix* untuk Evaluasi Model 16
Tabel 3.1 *Dataset* Laporan Curah Hujan Kab.Bogor 21
Tabel 3.2 Contoh *Dataset* 24
Tabel 3.3 *Confusion Matrix* Hasil Klasifikasi Menggunakan Algoritma *C4.5* 29
Tabel 4.1 Hasil Evaluasi Model Klasifiaksi *C4.5* 49

ABSTRAK

Khassogi, Abdurrahman Wahid. 2025. **Klasifikasi Curah Hujan Menggunakan Algoritma *Decision Tree C4.5*.** Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr.Zainal Abidin, M.Kom (II) Tri Mukti Lestari, M.Kom.

.

Kata Kunci: Curah Hujan*, Decision Tree*, Algoritma *C4.5*, Klasifikasi, Evaluasi Model.

Penelitian ini dilatarbelakangi oleh pentingnya informasi curah hujan dalam mendukung pengambilan keputusan di berbagai sektor, khususnya di bidang pertanian, infrastruktur, dan mitigasi bencana. Permasalahan utama yang diangkat adalah bagaimana menerapkan algoritma *Decision Tree C4.5* untuk mengklasifikasikan curah hujan berdasarkan data historis cuaca. Tujuan dari penelitian ini adalah untuk mengetahui hasil dan tingkat akurasi klasifikasi curah hujan menggunakan algoritma *C4.5*. Penelitian dilakukan dengan pendekatan kuantitatif, dimulai dari pengumpulan data cuaca harian Kabupaten Bogor dari BMKG, dilanjutkan dengan proses preprocessing data, pemilihan fitur, pelatihan model klasifikasi menggunakan algoritma *C4.5*, serta evaluasi performa model menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score. Pengujian dilakukan dalam lima skenario pembagian data latih dan data uji (90:10 hingga 50:50) untuk mengukur konsistensi kinerja model. Hasil penelitian menunjukkan bahwa skenario dengan rasio 90:10 memberikan akurasi tertinggi sebesar 77%. Fitur seperti suhu rata-rata, lama penyinaran matahari, dan kecepatan angin rata-rata merupakan variabel yang paling berpengaruh dalam klasifikasi. Berdasarkan hasil tersebut, algoritma *C4.5* dapat dijadikan solusi yang efektif dalam membangun sistem klasifikasi curah hujan berbasis data historis cuaca.

ABSTRACT

Khassogi, Abdurrahman Wahid. 2025. **Rainfall Classification Using the C4.5 Decision Tree Algorithm.** Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Advisor: (I) Dr.Zainal Abidin, M.Kom (II) Tri Mukti Lestari, M.Kom.

This research is motivated by the importance of rainfall information in supporting decision making in various sectors, especially in agriculture, infrastructure, and disaster mitigation. The main problem raised is how to apply the Decision Tree C4.5 algorithm to classify rainfall based on historical weather data. The purpose of this study is to determine the results and accuracy level of rainfall classification using the C4.5 algorithm. The research was conducted with a quantitative approach, starting from collecting daily weather data of Bogor Regency from BMKG, followed by data preprocessing, feature selection, classification model training using C4.5 algorithm, and model performance evaluation using evaluation metrics such as accuracy, precision, recall, and F1-score. Tests were conducted in five scenarios of training and test data distribution (90:10 to 50:50) to measure the consistency of model performance. The results showed that the scenario with a 90:10 ratio gave the highest accuracy of 77%. Features such as average temperature, sunshine duration, and average wind speed were the most influential variables in the classification. Based on these results, the C4.5 algorithm can be used as an effective solution in building a rainfall classification system based on historical weather data.

Keywords: Rainfall, Decision Tree, C4.5 Algorithm, Classification, Model Evaluation.

مستخلص البحث

كاصوجي، عبد الرحمن وحيد. 2025. **تصنيف هطول الأمطار باستخدام خوارزمية شجرة القرار C4.5**. أطروحة. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (١) الدكتور زين العابدين، الماجستير في علوم الحاسوب (٢) تري موكتي ليستاري، الماجستير في علوم الحاسوب.

الكلمات المفتاحية: هطول الأمطار، شجرة القرار، خوارزمية C4.5، التصنيف، تقييم النموذج.

تأتي هذه الدراسة في إطار أهمية المعلومات عن هطول الأمطار في دعم اتخاذ القرار في مختلف القطاعات، خاصة في مجالات الزراعة والبنية التحتية والتخفيف من الكوارث. المشكلة الرئيسية التي تطرحها هذه الدراسة هي كيفية تطبيق خوارزمية شجرة القرار C4.5 لتصنيف هطول الأمطار بناءً على بيانات الطقس التاريخية. تهدف هذه الدراسة إلى معرفة نتائج ومستوى دقة تصنيف هطول الأمطار باستخدام خوارزمية C4.5.أجريت الدراسة باستخدام المنهج الكمي، بدءًا من جمع بيانات الطقس اليومية لمقاطعة بوجور من الهيئة العامة للأرصاد الجوية، ثم عملية المعالجة المسبقة للبيانات، واختيار الميزات، وتدريب نموذج التصنيف باستخدام خوارزمية C4.5، وتقييم أداء النموذج باستخدام مقاييس التقييم مثل الدقة والضبط والاسترجاع و F1-score. تم تنفيذ عملية التدريب والاختبار في خمسة سيناريوهات مختلفة لتقسيم البيانات، وهي ٩٠٪ للتدريب و١٠٪ للاختبار، ٨٠٪ للتدريب و٢٠٪ للاختبار، ٧٠٪ للتدريب و٣٠٪ للاختبار، ٦٠٪ للتدريب و٤٠٪ للاختبار، و٥٠٪ للتدريب و٥٠٪ للاختبار.أظهرت النتائج أن السيناريو بنسبة ٩٠٪ للتدريب و١٠٪ للاختبار أعطى أعلى دقة بنسبة ٧٧٪. كانت الميزات مثل متوسط درجة الحرارة ومدة سطوع الشمس ومتوسط سرعة الرياح هي المتغيرات الأكثر تأثيرًا في التصنيف. بناءً على هذه النتائج، يمكن اعتبار خوارزمية C4.5 حلاً فعالاً في بناء نظام تصنيف

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah memberikan kontribusi besar dalam pengolahan dan analisis data di berbagai bidang, termasuk dalam bidang klimatologi dan lingkungan hidup. Salah satu fenomena penting yang menjadi perhatian banyak pihak adalah curah hujan. Informasi mengenai klasifikasi curah hujan sangat krusial untuk mendukung berbagai aktivitas, seperti pertanian, perencanaan pembangunan infrastruktur, hingga mitigasi bencana alam. Ketidakpastian kondisi cuaca yang semakin meningkat akibat perubahan iklim global menuntut adanya metode ilmiah yang dapat membantu memetakan dan mengklasifikasikan pola curah hujan secara akurat dan sistematis(Nasrullah et al., 2024).

Dalam menghadapi tantangan tersebut, pemanfaatan teknologi *data mining* menjadi solusi yang menjanjikan. Salah satu algoritma yang banyak digunakan dalam klasifikasi data adalah Algoritma *C4.5*. Algoritma ini memiliki keunggulan dalam membentuk model pohon keputusan yang mudah dipahami, mampu menangani data numerik maupun kategorikal, serta menghasilkan aturan klasifikasi yang transparan(Risnawati et al., 2023). Dengan menggunakan atribut cuaca seperti suhu, kelembaban, tekanan udara, dan kecepatan angin, algoritma *C4.5* dapat membentuk pola klasifikasi yang memudahkan interpretasi terhadap kategori curah hujan, baik rendah, sedang, maupun tinggi.

Penelitian sebelumnya menunjukkan bahwa algoritma *C4.5* mampu memberikan tingkat akurasi yang tinggi dalam klasifikasi data curah hujan. Dalam studi yang di lakukan oleh (Al Arif et al., 2022), algoritma *C4.5* menghasilkan akurasi sebesar 88,03%, lebih tinggi dibandingkan algoritma *Naïve Bayes* dan *K-Nearest Neighbors*. Hasil ini menguatkan potensi *C4.5* sebagai alat bantu yang efektif dalam pemrosesan data iklim. Demikian pula, penelitian yang dilakukan oleh (Siregar, N. A., 2024)menunjukkan keberhasilan penggunaan algoritma ini dalam mengklasifikasikan curah hujan di Kota Medan dengan hasil klasifikasi yang akurat dan bermanfaat untuk perencanaan mitigasi.

Pentingnya memahami dan mengantisipasi kondisi cuaca, termasuk curah hujan, juga sejalan dengan nilai-nilai dalam ajaran Islam yang menekankan pentingnya ilmu pengetahuan dan kesiapsiagaan dalam kehidupan. Hal ini tercermin dalam firman Allah SWT dalam Surah *An-Nur* ayat 43:

اَلَمْ تَرَ اَنَّ اللّٰهَ يُزْجِيْ سَحَابًا ثُمَّ يُؤَلِّفُ بَيْنَهٗ ثُمَّ يَجْعَلُهٗ رُكَامًا فَتَرَى الْوَدْقَ يَخْرُجُ مِنْ خِلٰلِهٖۚ وَيُنَزِّلُ مِنَ السَّمَاۤءِ مِنْ جِبَالٍ فِيْهَا مِنْۢ بَرَدٍ فَيُصِيْبُ بِهٖ مَنْ يَّشَاۤءُ وَيَصْرِفُهٗ عَنْ مَّنْ يَّشَاۤءُۗ يَكَادُ سَنَا بَرْقِهٖ يَذْهَبُ بِالْاَبْصَارِ ﴿٤٣﴾

"Tidakkah engkau melihat bahwa sesungguhnya Allah mengarahkan awan secara perlahan, kemudian mengumpulkannya, lalu menjadikannya bertumpuk-tumpuk. Maka, engkau melihat hujan keluar dari celah-celahnya. Dia (juga) menurunkan (butiran-butiran) es dari langit, (yaitu) dari (gumpalan-gumpalan awan seperti) gunung-gunung. Maka, Dia menimpakannya (butiran-butiran es itu) kepada siapa yang Dia kehendaki dan memalingkannya dari siapa yang Dia kehendaki. Kilauan kilatnya hampir-hampir menghilangkan penglihatan", *(Q.S. An-Nur: 43).*

Ayat ini menunjukkan bahwa hujan merupakan salah satu tanda kekuasaan Allah SWT yang terjadi melalui proses yang dapat diamati dan dipelajari. Dengan demikian, mengkaji curah hujan melalui pendekatan ilmiah tidak hanya merupakan

upaya akademis, tetapi juga bagian dari refleksi terhadap kebesaran ciptaan-Nya. Dalam konteks ini, upaya klasifikasi dan pemodelan curah hujan tidak hanya penting secara praktis, tetapi juga mencerminkan semangat Islam dalam menggabungkan ilmu dan iman untuk kemaslahatan umat.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi curah hujan dengan menggunakan algoritma *C4.5* berdasarkan data iklim Indonesia. Dengan memanfaatkan data historis dari berbagai variabel meteorologis, penelitian ini diharapkan dapat menghasilkan sistem klasifikasi yang akurat dan dapat digunakan sebagai acuan dalam pengambilan keputusan, khususnya dalam bidang pertanian, mitigasi bencana, dan pengelolaan sumber daya air. Pendekatan ini diharapkan tidak hanya memberikan kontribusi bagi dunia akademik, tetapi juga mendukung pembangunan berkelanjutan yang berbasis pada data.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, bagaimana tingkat akurasi algoritma *C4.5* dalam mengklasifikasikan curah hujan data historis di Indonesia?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan, di antaranya.

a. Data yang digunakan dalam penelitian ini merupakan data curah hujan harian yang diperoleh dari situs resmi BMKG.

b. Implementasi program menggunakan bahasa pemrograman *Python*.

1.4 Tujuan Penelitian

Untuk mengetahui hasil dan mendapatkan tingkat keakuratan dari algoritma *C4.5* dalam mengklasifikasikan data historis curah hujan di Indonesia.

1.5 Manfaat Penelitian

Dengan melakukan penelitian ini, diharapkan dapat memberikan manfaat dan keuntungan di kemudian hari. Beberapa contoh manfaat yang diharapkan termasuk sebagai berikut.

a. Menyediakan Solusi yang lebih akurat dan andal untuk memprediksi curah hujan, sehingga dapat mendukung perkiraan cuaca yang lebih tepat.
b. Penelitian ini akan memperkaya literatur ilmiah tentang penerapan algoritma *Decision Tree C4.5* dalam prediksi curah hujan berbasis data online, serta berkontribusi pada pengembangan metodologi baru dalam analisis cuaca.
c. Hasil dari penelitian ini, khususnya dalam klarifikasi curah hujan yang di kategorikan, dapat menjadi acuan atau model awal bagi penelitian selanjutnya yang berfokus pada pengembangan model prediksi dengan lebih banyak variabel atau penggunaan algoritma yang berbeda.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Bagian ini menguraikan sejumlah penelitian terdahulu yang memiliki relevansi dengan topik yang sedang diteliti. Beragam studi sebelumnya telah menerapkan algoritma *Decision Tree C4.5* dalam berbagai bidang dan konteks yang berbeda. Untuk memberikan gambaran yang lebih sistematis mengenai penelitian-penelitian tersebut Tabel 2.1.

Tabel 2.1 Penelitian Terkait

No	Judul	Metode	Variable/ Fitur	Hasil penelitian	Perbandingan penelitian
1	*Comparison of data mining methods for prediction of rainfall with C4.5, Naïve Bayes, and KNN Algorithm* (Al Arif et al., 2022)	*C4.5, Naïve Bayes, KNN*	Variabel target: status curah hujan (hujan atau tidak hujan). Variabel prediktor: *Tavg*(suhu rata-rata), *RH_avg*(kelembapan rata-rata), *sunshine duration* (lama penyinaran) dan *wind speed* (kecepatan angin).	Penelitian membandingkan akurasi tiga algoritma klasifikasi (*C4.5*, *Naïve Bayes, KNN*) menggunakan data BMKG. *C4.5* memberikan akurasi tertinggi sebesar 88,03%, menjadikannya metode terbaik untuk prediksi curah hujan.	Penelitian yang dilakukan (Al Arif et al., 2022) menggunakan variabel prediktor berupa suhu rata-rata (*Tavg*), kelembapan rata-rata (*RH_avg*), durasi penyinaran matahari (*sunshine duration*), dan kecepatan angin (*wind speed*). Sementara itu, penelitian ini menggunakan variabel *Tavg, RH_avg*, lama penyinaran matahari (*ss*), arah angin maksimum (*ddd_x*), dan kecepatan angin rata-rata (*ff_avg*). Perbedaan utama terletak pada penambahan atribut *ddd_x* dan *ff_avg* yang diharapkan

No	Judul	Metode	Variable/ Fitur	Hasil penelitian	Perbandingan penelitian
					mampu memperkaya informasi dalam proses klasifikasi.
2	Klasifikasi *data mining* untuk mengestimasi potensi curah hujan berdampak banjir daerah menggunakan *algoritma C4.5* (Risnawati et al., 2023)	*C4.5*	Variabel target: Potensi banjir berdasarkan curah hujan (dengan kategori hujan). Variabel prediktor: *Tavg*(suhu rata-rata), *RH_avg*(kelembapan rata-rata), *wind speed* (kecepatan angin) dan *Rainfall* (curah hujan kategori: rendah, sedang, tinggi dan ekstrem).	*C4.5* digunakan untuk klasifikasi curah hujan yang menyebabkan banjir di Jawa Tengah dan Timur. Dengan data dari BMKG dan NASA, model menghasilkan akurasi sebesar 83,33% dan membantu identifikasi daerah rawan banjir.	Penelitian yang dilakukan (Risnawati et al., 2023) memanfaatkan variabel *Tavg, RH_avg,* kecepatan angin (*wind speed*), dan curah hujan dalam bentuk kategori (rendah hingga ekstrem). Adapun penelitian ini menggunakan variabel *Tavg, RH_avg, ss, ddd_x,* dan *ff_avg* dengan target dua kelas, yakni hujan dan tidak hujan. Perbandingan menunjukkan bahwa penelitian ini menekankan pada atribut tambahan seperti arah angin maksimum dan kecepatan angin rata-rata, serta pendekatan klasifikasi *biner*.
3	Implementasi *Data Mining* pada *Dataset* Prakiraan Cuaca Menggunakan *Algoritma C4.5* (Fikry Ubaidillah et al., 2023)	*C4.5*	Variabel target: Prakiraan Cuaca (kategori cuaca). Variabel prediktor: Suhu, kelembapa, kecepatan angin dan arah angin.	Menggunakan 2.400 data prakiraan cuaca dari *accuweather*, model *C4.5* membentuk pohon keputusan dengan kelembapan sebagai atribut utama. Akurasi mencapai 81,94%, cukup baik	Penelitian yang dilakukan (Fikry Ubaidillah et al., 2023) mengimplementasikan algoritma *C4.5* dengan variabel suhu, kelembapan, kecepatan angin, dan arah angin sebagai atribut prediktor. Variabel-variabel tersebut memiliki kesamaan dengan yang digunakan dalam penelitian

No	Judul	Metode	Variable/ Fitur	Hasil penelitian	Perbandingan penelitian
				dalam memprediksi kondisi cuaca harian.	ini,namun penelitian ini menambahkan atribut lama penyinaran matahari(*ss*) serta memanfaatkan *ddd_x & ff_avg* secara eksplisit sebagai variabel yang relevan terhadap klasifikasi curah hujan.
4	Penerapan Algoritma *Decision Tree C4.5* untuk Prediksi Cuaca di Kota Semarang (Nafisatun Najah et al., 2025)	*C4.5*	Variabel target: Status curah hujan (hujan/tidak hujan). Variable prediktor: Temperatur rata-rata, kelembapan rata-rata, lama penyinaran dan kecepatan angin rata-rata	Dengan data cuaca dari BMKG (2016–2023), algoritma *C4.5* menghasilkan akurasi sangat tinggi yaitu 99,12%, membuktikan efektivitas metode ini dalam klasifikasi cuaca untuk mendukung perencanaan kebijakan daerah.	Penelitian yang dilakukan (Nafisatun Najah et al., 2025) menggunakan atribut temperatur, kelembapan, lama penyinaran, dan kecepatan angin rata-rata. Penelitian ini memiliki kesamaan dalam hal penggunaan *Tavg, RH_avg,* dan *ss*, namun melibatkan tambahan variabel arah angin saat kecepatan maksimum (*ddd_x*) untuk memperluas cakupan analisis dalam klasifikasi cuaca, khususnya curah hujan.
5	Perbandingan Akurasi *Algoritma C4.5* dan *K- Nearest Neighbors* untuk klasifikasi curah hujan berdasarkan iklim Indonesia (Nasrullah et al., 2024)	*C4.5, KNN*	Variabel target: Kategorik curah hujan. Variabel prediktor: *Tavg*(suhu rata-rata), *RH_avg*(kelembapan rata-rata), *RR*(curah hujan),	Studi membandingkan *C4.5* dan *KNN* dengan *tuning* parameter. Setelah optimasi, *KNN* mencapai akurasi 83,37% dan *C4.5* 82,56%, menunjukkan bahwa *tuning* meningkatkan	Penelitian yang dilakukan (Nasrullah et al., 2024) mengkaji variabel *Tavg, RH_avg, RR, sunshine duration,* dan arah angin sebagai atribut prediktor. Sementara itu, penelitian ini menghilangkan *RR* sebagai fitur karena telah didiskretisasi

No	Judul	Metode	Variable/ Fitur	Hasil penelitian	Perbandingan penelitian
			sunshine duration dan *wind direction*.	performa dan *KNN* sedikit lebih unggul.	menjadi target kelas, dan menggunakan kombinasi *Tavg, RH_avg, ss, ddd_x,* dan *ff_avg* sebagai variabel utama. Hal ini menunjukkan pendekatan variabel prediktor yang lebih disesuaikan dengan proses klasifikasi berbasis atribut meteorologis yang saling berkaitan.

2.2 *Machine Learning* (ML)

Machine Learning (ML) merupakan pendekatan komputasi yang memungkinkan sistem belajar dari data untuk mengenali pola dan melakukan prediksi tanpa perlu diprogram secara eksplisit (I.D. Id, 2021). Dalam bidang klimatologi, ML semakin banyak dimanfaatkan untuk menganalisis dan memprediksi fenomena cuaca, termasuk klasifikasi curah hujan.

Penerapan ML untuk klasifikasi curah hujan telah dibuktikan dalam berbagai penelitian. Seperti pada penelitian yang dilakukan Nafisatun Najah et al. (2025) menggunakan algoritma *Decision Tree C4.5* untuk mengklasifikasikan status curah hujan di Kota Semarang berdasarkan data cuaca dari BMKG tahun 2016–2023. Model dibangun menggunakan *RapidMiner* dan menghasilkan akurasi sangat tinggi, yaitu 99,12%, menunjukkan efektivitas ML dalam mengenali pola meteorologis yang kompleks. Penelitian lain oleh Fikry Ubaidillah et al. (2023) menggunakan *dataset* dari situs prakiraan cuaca untuk wilayah Pantura, dan juga menerapkan algoritma *C4.5*. Dengan atribut kelembapan, suhu, kecepatan dan arah

angin, model menghasilkan akurasi 81,94%, cukup baik untuk sistem prakiraan berbasis data.

Dari dua penelitian ini, dapat disimpulkan bahwa pendekatan ML sangat potensial dalam klasifikasi data cuaca. Hasil yang diperoleh tidak hanya berguna secara ilmiah, tetapi juga secara praktis, misalnya dalam pengambilan keputusan sektor pertanian, mitigasi bencana, dan perencanaan pembangunan.

2.3 Klasifikasi

Klasifikasi adalah metode dalam supervised learning yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang ditemukan dari data historis (Sarker, 2021). Dalam klasifikasi curah hujan, metode ini digunakan untuk menentukan apakah suatu kondisi cuaca akan menghasilkan hujan atau tidak, atau mengelompokkannya dalam intensitas tertentu.

Penelitian oleh Al Arif et al. (2022) membandingkan tiga algoritma klasifikasi (*C4.5, KNN,* dan *Naïve Bayes*) menggunakan data BMKG. Hasilnya, algoritma *C4.5* memberikan performa terbaik dengan akurasi 88,03%, lebih unggul dibanding dua metode lainnya. Hal ini menunjukkan bahwa klasifikasi menggunakan *Decision Tree* mampu mengelompokkan data curah hujan secara efektif dan logis.

Sementara itu, dalam studi yang dilakukan oleh Nasrullah et al. (2024), perbandingan dilakukan antara *C4.5* dan *KNN* pada data iklim Indonesia. Setelah dilakukan *tuning* parameter, *C4.5* berhasil mencapai akurasi 82,56%. Hasil ini memperkuat bahwa klasifikasi sangat bergantung pada pemilihan metode yang tepat serta preprocessing data yang optimal.

Berdasarkan hasil-hasil tersebut, dapat disimpulkan bahwa metode klasifikasi merupakan pendekatan yang efektif untuk menangani data cuaca yang bersifat numerik dan fluktuatif. Dengan dukungan data historis yang akurat, model klasifikasi dapat menjadi alat bantu prediksi yang andal dalam sistem prakiraan hujan.

2.4 Metode *Decision Tree*

Decision Tree atau pohon Keputusan adalah salah satu metode dalam *data mining* yang digunakan untuk klasifikasi dan prediksi. Metode ini menyajikan model dalam bentuk struktur pohon, dimana setiap simpul (*node*) mempresentasikan atribut, setiap cabang mewakili nilai dari atribut tersebut, dan setiap daun (*leaf node*) menunjukkan hasil klasifikasi untuk keputusan akhir.

Menurut Sartika et al. (2017) , dalam penelitiannya menjelaskan bahwa pohon keputusan bekerja dengan membagi data secara rekursif berdasarkan atribut yang dipilih hingga mencapai kondisi homogen atau tidak bisa dibagi lagi. Proses ini memudahkan dalam memahami dan menginterpretasi hasil prediksi karena bentuknya menyerupai alur keputusan logis dan sistematis. Komponen utama dalam *Decision Tree* terdiri dari.

a. *Root node*: titik awal yang mencerminkan keseluruhan data dan mewakili atribut dengan pengaruh paling besar terhadap klasifikasi.
b. *Internal node*: titik percabangan yang menunjukkan pengujian terhadap nilai atribut tertentu.
c. *Leaf node*: titik akhir yang merepresentasikan label kelas atau hasil akhir klasifikasi, berupa label kelas yang dituju.

d. *Branch*: jalur penghubung antara node berdasarkan hasil uji terhadap suatu atribut.

Setiap jalur dari akar hingga daun pada pohon dapat ditafsirkan sebagai satu aturan klasifikasi, sehingga seluruh pohon terdiri atas sekumpulan aturan yang dapat digunakan untuk mengklasifikasikan data baru.

Berdasarkan struktur dan cara kerjanya, *decision tree* tidak hanya memudahkan proses klasifikasi, tetapi juga menghasilkan model yang mudah diinterpretasikan karena menyerupai alur logika pengambilan keputusan. Pendekatan ini membentuk struktur hierarkis berupa pohon yang menyusun aturan-aturan klasifikasi dari data yang tersedia. Salah satu algoritma yang banyak digunakan untuk membangun model pohon keputusan adalah algoritma *C4.5*(Azwanti, 2018). Algoritma ini dikenal memiliki keunggulan dalam menangani atribut bertipe kontinu, menangani data dengan nilai yang hilang, serta mampu menghasilkan model yang lebih optimal melalui proses pemangkasan (*pruning*).

2.4.1 Algoritma *C4.5*

Dalam penelitian ini, algoritma *C4.5* digunakan sebagai salah satu metode klasifikasi untuk membangun model pohon keputusan yang mampu mengelompokkan data secara efektif. merupakan salah satu algoritma *decision tree* yang paling populer dalam bidang *data mining* dan *machine learning*. Algoritma ini dikembangkan oleh Ross Quinlan sebagai penyempurnaan dari algoritma *ID3* (*Iterative Dichotomiser 3*), dengan mengatasi beberapa keterbatasan *ID3*, seperti ketidakmampuan menangani atribut kontinu, nilai yang hilang (*missing values*),

serta kecenderungan untuk menghasilkan model yang terlalu kompleks (*overfitting*).

Keunggulan utama algoritma *C4.5* terletak pada fleksibilitas dan ketahanannya dalam berbagai kondisi data. Algoritma ini mampu bekerja dengan baik pada data yang memiliki atribut bertipe kontinu maupun diskrit, serta dapat menangani nilai yang hilang *(missing values)*. Selain itu, algoritma ini juga melakukan pemangkasan pohon (*pruning*) untuk menyederhanakan model dan meningkatkan generalisasi terhadap data baru(Lakshmi et al., 2016).

A. Proses kerja algoritma *C4.5*

Secara umum, langkah-langkah algoritma *C4.5* dalam membangun pohon keputusan adalah sebagai berikut.

a. Menghitung *entropy* dan *gain* ratio untuk setiap atribut yang tersedia.

b. Memilih atribut dengan *gain* ratio tertinggi sebagai *node* (simpul) keputusan.

c. Membagi *dataset* ke dalam subset berdasarkan nilai atribut yang dipilih.

d. Mengulangi proses secara rekursif untuk setiap subset hingga.

   a) Seluruh data dalam subset termasuk dalam kelas yang sama.

   b) Tidak ada atribut tersisa yang dapat digunakan untuk pemisahan.

B. *Entropy* dan *information gain*

Salah satu proses krusial dalam algoritma *C4.5* adalah pemilihan atribut akar, yaitu atribut terbaik yang akan ditempatkan sebagai simpul utama pada pohon keputusan. Pemilihan atribut ini didasarkan pada nilai *Information Gain*, yang

menunjukkan seberapa besar suatu atribut mampu mengurangi tingkat ketidakpastian dalam data. Sebelum menghitung *Information Gain*, terlebih dahulu dihitung nilai *Entropy* dari *dataset*, yang mengukur tingkat ketidakteraturan atau keragaman kelas.

a. *Entropy*

Entropy digunakan untuk mengukur keragaman data dalam satu set. Semakin homogen suatu data, maka nilai *entropy*-nya semakin kecil, dan sebaliknya.

$$\text{Entropy}(S) = -\sum_{i=1}^{n} pi \log_2 pi \qquad (2.1)$$

Keterangan:
S = satu set data (misalnya data cuaca).
N = jumlah kelas (misal hujan dan tidak hujan).
Pi = proporsi data dari kelas ke- i.
log_2 = logaritma basis 2.

Catatan: Jika semua data dalam satu kelas, maka *Entropy* = 0 (murni); jika data terbagi rata di semua kelas, maka *Entropy* maksimum (tidak murni).

b. *Information gain*

Setelah *entropy* awal dihitung, langkah selanjutnya adalah menghitung *Information Gain* untuk setiap atribut. *Information Gain* menunjukkan seberapa besar pengurangan ketidakpastian setelah data dibagi berdasarkan atribut tertentu.

$$Gain(S, A) = Entropy(S) - \sum_{v=1}^{v} \frac{|Sv|}{|S|} Entropy(Sv) \qquad (2.2)$$

Keterangan:
$Gain(S, A)$ = *information gain* dari atribut A.
$Entropy(S)$ = *entropy* dari set data awal.
Sv = subset dari S untuk nilai v.
$\frac{|Sv|}{|S|}$ = proporsi subset Sv terhadap S.

Semakin besar nilai *gain*, maka semakin baik atribut tersebut dalam memisahkan data.

2.5 Curah Hujan

Curah hujan adalah jumlah air hujan yang jatuh ke permukaan bumi selama periode waktu tertentu, biasanya diukur dalam satuan milimeter (*mm*) per satuan waktu (Ruswanti, 2020). Curah hujan memainkan peran penting dalam sistem iklim global dan sangat mempengaruhi ketersediaan air, pertanian, serta pola cuaca harian dan musiman. Variabilitas curah hujan di wilayah tropis seperti Indonesia sangat dipengaruhi oleh faktor-faktor atmosferik, seperti suhu udara, kelembapan, tekanan udara, serta fenomena iklim global seperti *El Niño* dan *La Niña*(Hidayat et al., 2014).

Selain faktor atmosfer, kondisi lokal seperti topografi dan penggunaan lahan juga berkontribusi terhadap distribusi curah hujan di suatu wilayah. Daerah pegunungan, misalnya, cenderung menerima curah hujan lebih tinggi akibat proses pengangkatan orografis. Dalam konteks operasional (Wandayantolis et al., 2023), Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) mengelompokkan intensitas curah hujan menjadi lima kategori, yaitu:

a. Tidak hujan (0 mm).
b. Hujan sangat ringan (1 – 5 mm).
c. Hujan ringan (5 – 20 mm).
d. Hujan sedang (20 – 50 mm).
e. Hujan lebat (50 – 100 mm).
f. Hujan sangat lebat (>100 mm).

Informasi mengenai klasifikasi intensitas curah hujan sangat penting untuk berbagai keperluan, seperti mitigasi bencana banjir, pengelolaan sumber daya air, dan perencanaan aktivitas pertanian. Dengan kemajuan teknologi, penggunaan metode berbasis *machine learning* untuk prediksi curah hujan semakin berkembang karena mampu mengolah data historis secara lebih efektif dan menghasilkan estimasi yang lebih akurat (Hidayat et al., 2014).

Mengingat pentingnya informasi tentang curah hujan dalam perencanaan dan mitigasi bencana, maka upaya prediksi curah hujan yang akurat menjadi sangat relevan. Oleh karena itu, pada bagian selanjutnya akan dibahas mengenai konsep dan teknik prediksi curah hujan.

2.5.1 Prediksi Curah Hujan

Prediksi adalah proses memperkirakan kejadian di masa depan dengan memanfaatkan data historis secara ilmiah, baik melalui pendekatan kuantitatif maupun kualitatif (Panggabean et al., 2020). Dalam bidang meteorologi, khususnya prediksi curah hujan, pendekatan kuantitatif semakin banyak digunakan karena tersedianya data historis yang melimpah dan kemajuan teknologi komputasi.

Salah satu pendekatan kuantitatif yang umum digunakan adalah metode klasifikasi, yang mengelompokkan data berdasarkan atribut cuaca untuk menentukan intensitas curah hujan. Algoritma *decision tree C4.5* sering diterapkan karena dapat menangani data kategorikal dan numerik, serta menghasilkan aturan klasifikasi yang mudah dipahami. Lou et al. (2021) dalam penelitiannya menunjukkan bahwa *C4.5* efektif dalam memprediksi variabilitas curah hujan jangka panjang melalui pemrosesan data meteorologis seperti suhu, kelembapan,

dan data historis curah hujan. Keunggulan algoritma ini terletak pada kemampuannya menangani data hilang, melakukan *pruning* untuk menghindari *overfitting*, serta menghasilkan model yang efisien dan interpretatif.

Oleh karena itu, penggunaan algoritma klasifikasi seperti *C4.5* menjadi solusi yang efektif dalam sistem prediksi curah hujan berbasis data. Kemampuannya dalam menangani data hilang, melakukan pruning untuk menghindari *overfitting*, serta menghasilkan model yang akurat dan mudah diinterpretasikan, menjadikannya alat yang andal untuk mendukung pengambilan keputusan, terutama dalam mitigasi risiko bencana dan pengelolaan sumber daya alam.

2.6 Evaluasi model dengan *Confusion Matrix*

Evaluasi model merupakan proses penting dalam *machine learning* untuk mengukur seberapa baik model dapat melakukan klasifikasi terhadap data yang belum pernah dilihat sebelumnya. Salah satu metode evaluasi yang umum digunakan adalah *confusion matrix*, yang memberikan representasi visual atas performa klasifikasi model dengan membandingkan nilai aktual dan nilai prediksi (Sathyanarayanan, 2024).

a. Struktur *Confusion Matrix*

Pada klasifikasi biner, struktur *confusion matrix* dijabarkan pada tabel 2.2.

Tabel 2.2 Struktur *Confusion Matrix* untuk Evaluasi Model

		Prediksi	
		Positif	*Negatif*
Aktual	*Positif*	*True Positive* (TP)	*False Negative* (FN)
	Negatif	*False Positive* (FP)	*True Negative* (TN)

Keterangan :
TP = Data aktual positif yang diprediksi positif.
FP = Data aktual negatif yang diprediksi positif (kesalahan tipe I).
FN = Data aktual positif yang diprediksi negatif (kesalahan tipe II).
TN = Data aktual negatif yang diprediksi negative.

Untuk klasifikasi multikelas, matriks diperluas menjadi bentuk persegi sesuai jumlah kelas(Sathyanarayanan, 2024).

b. Metrik evaluasi

Dari *confusion matrix*, berbagai metrik evaluasi dapat dihitung.

a) *Accuracy*

Mengukur seberapa besar proporsi klasifikasi yang benar terhadap data:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.3}$$

b) *Precision*

Mengukur ketepatan prediksi positif seluruh prediksi positif:

$$Precision = \frac{TP}{TP + FP} \tag{2.4}$$

c) *Recall/Sensitivty*

Mengukur seberapa baik model mendeteksi kasus positif:

$$Recall = \frac{TP}{TP + FN} \tag{2.5}$$

d) *F1-Score*

Merupakan rata-rata harmonis dari *precision* dan *recall*:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{2.6}$$

Metrik seperti *precision* dan *recall* sangat penting digunakan dalam konteks data yang tidak seimbang, karena akurasi saja bisa menyesatkan jika salah satu kelas mendominasi(Sathyanarayanan, 2024).

c. Peran evaluasi dalam pengembangan model

Evaluasi menggunakan *confusion matrix* memungkinkan analisis terhadap kesalahan klasifikasi yang dilakukan oleh model, serta membantu mendeteksi bias terhadap kelas tertentu. Selain itu, metrik evaluasi juga dapat digunakan untuk memperbaiki model, seperti melalui penyeimbangan data atau tuning parameter(Han et al., 2011).

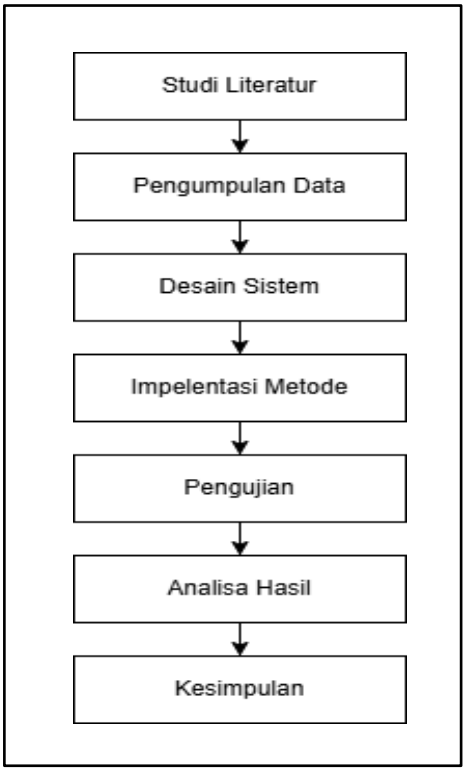
Dengan demikian, penggunaan *confusion matrix* dan metrik-metrik turunannya tidak hanya membantu mengukur akurasi model, tetapi juga memberikan wawasan yang lebih mendalam terhadap kinerja dan keandalan model klasifikasi yang digunakan(Sathyanarayanan, 2024)

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

Penelitian ini bertujuan untuk membangun sistem klasifikasi curah hujan berdasarkan data historis cuaca menggunakan algoritma *Decision Tree C4.5*. Sistem dikembangkan dengan bahasa pemrograman *Python* dan didukung oleh pustaka *machine learning* seperti *Scikit-Learn*. Pada subbab ini dijelaskan rancangan penelitian secara keseluruhan, yang mencakup alur proses mulai dari pengumpulan data, *preprocessing*, transformasi fitur, perancangan sistem klasifikasi, hingga tahap implementasi dan pengujian untuk mengevaluasi kinerja model yang dibangun.



Gambar 3.1 Desain Penelitian

Alur desain penelitian yang terdiri dari tujuh tahapan utama seperti pada gambar 3.1. Penelitian diawali dengan studi literatur untuk mengkaji teori-teori dan penelitian terdahulu yang relevan sebagai dasar pengembangan sistem. Selanjutnya dilakukan pengumpulan data cuaca harian dari BMKG yang mencakup parameter-parameter seperti suhu, kelembaban, durasi penyinaran matahari, angin, dan curah hujan. Setelah data diperoleh, tahap desain sistem dilakukan untuk merancang arsitektur klasifikasi menggunakan algoritma *Decision Tree C4.5*. Tahap berikutnya adalah implementasi metode, di mana algoritma diolah dan diimplementasikan menggunakan bahasa pemrograman *Python* dengan pustaka *Scikit-Learn*. Model yang telah dibangun kemudian diuji melalui skenario pembagian data latih dan data uji, dan hasilnya dianalisis menggunakan metrik evaluasi seperti akurasi, *precision, recall,* dan *F1-score*. Terakhir, penelitian ditutup dengan tahap kesimpulan yang merangkum hasil yang diperoleh serta memberikan evaluasi terhadap performa model dan efektivitas metode yang digunakan.

3.2 Pengumpulan Data

Proses pengumpulan data merupakan langkah awal dalam menjalankan penelitian ini. Data yang digunakan dalam penelitian ini berasal dari *dataset* resmi BMKG (Badan Meteorologi, Klimatologi, dan Geofisika Indonesia) yang diunduh langsung melalui platform Data *Online* BMKG https://dataonline.bmkg.go.id/. *Dataset* ini berisi catatan harian kondisi iklim di Kabupaten Bogor selama kurun waktu tahun 2019 hingga 2020, yang mencakup berbagai variabel meteorologis yang berhubungan dengan curah hujan. Penggunaan data resmi dari BMKG

memastikan akurasi dan validitas informasi klimatologis yang digunakan dalam penelitian ini. Berikut adalah data yang terdapat dalam *dataset* laporan curah hujan Kabupaten Bogor:

Tabel 3.1 *Dataset* Laporan Curah Hujan Kab.Bogor

Date	***Tn***	***Tx***	***Tavg***	***RH_avg***	***Ss***	***ff_x***	***ddd_x***	***ff_avg***	***ddd_car***	**Target**
01-01-2019	19	24,5	20,2	94	0,5	3	280	1	C	Hujan
02-01-2019	19,5	26,3	21,2	89	0,3	2	280	1	C	Hujan
03-01-2019	19,6	26	21,8	82	1,5	3	270	1	C	Hujan
04-01-2019	19,6	26,8	22,2	75	5,9	2	310	0	C	Tidak Hujan
05-01-2019	18,6	26,8	22,1	79	4,2	3	300	0	C	Hujan
06-01-2019	19,2	27,2	22,7	82	5,5	4	320	1	C	Hujan
07-01-2019	18,8	27,6	21,9	88	4,5	3	270	1	C	Tidak Hujan
08-01-2019	18,7	24,7	21,5	90	3,4	1	250	0	C	Hujan
09-01-2019	19	25	21,3	91	0	2	350	0	C	Hujan
10-01-2019	19,3	25,7	21,4	94	1,1	2	310	0	C	Tidak Hujan

Keterangan :

Tn	= Temperatur minimal	ff_x	= Kecepatan angin maksimum
Tx	= Temperatur maksimal	ddd_x	= Arah angin saat kecepatan maksimum
Tavg	= Temperatur rata-rata	ff_avg	= Kecepatan angin rata-rata
RH_avg	= Kelembapan rata-rata	ddd_car	= Arah angin dominan
ss	= Lama penyinaran matahari	Target	= Hujan dan Tidak Hujan

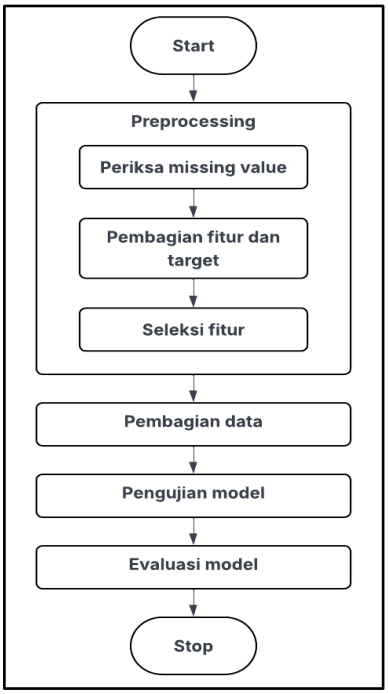
3.3 *Preprocessing* Data

Sebelum data digunakan dalam proses pelatihan menggunakan algoritma *Decision Tree C4.5*, diperlukan tahapan *preprocessing* data untuk memastikan bahwa data yang digunakan memiliki kualitas yang memadai. Tahap ini berperan penting dalam meningkatkan akurasi serta efisiensi proses klasifikasi yang akan dilakukan. *Preprocessing* data merupakan langkah fundamental dalam proses penemuan pengetahuan karena kualitas keputusan yang dihasilkan sangat bergantung pada kualitas data yang digunakan (Merdekawati & Ismail, 2019). Proses ini mencakup berbagai kegiatan seperti penanganan *missing values*, konversi tipe data jika diperlukan, serta seleksi atribut-atribut yang relevan untuk proses klasifikasi (Pebdika et al., 2023).

Salah satu permasalahan umum dalam preprocessing adalah keberadaan data yang tidak lengkap (*missing values*), yang dapat disebabkan oleh gangguan pencatatan, kerusakan peralatan pengamatan, atau inkonsistensi selama pengumpulan data, termasuk pada atribut-atribut penting seperti suhu, kelembapan, dan curah hujan. Oleh karena itu, tahap preprocessing tidak hanya bersifat teknis, tetapi juga strategis dalam memastikan integritas dan representasi data sebelum digunakan dalam proses pelatihan model klasifikasi.

3.4 Desain Sistem

Desain sistem ini menjelaskan tahapan klasifikasi curah hujan menggunakan algoritma *Decision Tree C4.5*, mulai dari input data, pra-pemrosesan, pelatihan, pengujian, hingga evaluasi akurasi. Setiap tahap disusun secara sistematis untuk memperoleh hasil prediksi yang optimal. Berikut penjelasan tiap tahap pada *flowchart*:



Gambar 3.2 *Flowchart* Desain Sistem

Flowchart Desain Sistem menunjukkan urutan tahapan yang dilakukan mulai dari awal hingga akhir proses seperti yang digambarkan pada gambar 3.2. Proses diawali dengan tahap *Start* yang menandakan dimulainya sistem. Selanjutnya masuk ke tahap *Preprocessing*, yaitu proses pra-pemrosesan data yang meliputi pemeriksaan nilai yang hilang (*missing value*), seleksi fitur yang relevan, serta pembagian data menjadi fitur dan target. Langkah ini bertujuan untuk memastikan bahwa data yang digunakan bersih dan siap untuk dianalisis. Setelah *preprocessing*, data dibagi menjadi data latih dan data uji pada tahap Pembagian Data, sesuai dengan rasio tertentu seperti 80:20 atau 70:30. Data latih digunakan pada tahap Pengujian Model untuk membangun model klasifikasi menggunakan algoritma *Decision Tree C4.5,* sedangkan data uji digunakan untuk mengevaluasi performa model yang telah dilatih. Evaluasi dilakukan pada tahap Evaluasi Model, yang mencakup pengukuran metrik seperti akurasi, *precision*, *recall*, dan *f1-score* untuk menilai seberapa baik model dalam melakukan klasifikasi. Setelah seluruh proses selesai, sistem ditutup dengan tahap *Stop* sebagai penanda bahwa alur klasifikasi curah hujan telah selesai dijalankan secara menyeluruh.

3.5 Perhitungan Manual Algoritma *C4.5*

Untuk memahami proses kerja algoritma *C4.5* secara menyeluruh, dilakukan perhitungan manual menggunakan sampel data berikut sebanyak 10 baris. *Dataset* ini terdiri dari enam atribut yaitu *Tavg, RH_avg, ss, ddd_x, ff_avg,* dan Target. Atribut Target dibagi menjadi dua kelas yakni Hujan dan Tidak Hujan.

Tabel 3.2 Contoh *Dataset*

No	*Tavg*	*RH_avg*	*ss*	*ddd_x*	*ff_avg*	Target
1	20,2	94	0,5	280	1	Hujan
2	21,2	89	0,3	280	1	Hujan
3	21,8	82	1,5	270	1	Hujan
4	22,2	75	5,9	310	0	Tidak Hujan
5	22,1	79	4,2	300	0	Hujan
6	22,7	82	5,5	320	1	Hujan
7	21,9	88	4,5	270	1	Tidak Hujan
8	21,5	90	3,4	250	0	Hujan
9	21,3	91	0	350	0	Hujan
10	21,4	94	1,1	310	0	Tidak Hujan

Distribusi kelas:

a. Hujan = 7 data

b. Tidak hujan = 3 data

Diketahui jumlah data adalah 10 data dengan kelas hujan sebanyak 7 data dan kelas tidak hujan sebanyak 3 data.

A. Menghitung Nilai *Entropy* Total

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

$$\begin{aligned}\text{Entropy}(7,3) &= -\left(\frac{7}{10}\right) log_2\left(\frac{7}{10}\right) - \left(\frac{3}{10}\right) log_2\left(\frac{3}{10}\right) \\ &= -0{,}7\ log_2(0{,}7) - 0{,}3\ log_2(0{,}3) \\ &= -0{,}7 \cdot (-0{,}515) - 0{,}3 \cdot (-1{,}737) \\ &= 0{,}361 + 0{,}521 \\ &= 0{,}882\end{aligned}$$

B. Menghitung Nilai *Entropy* dan *Gain* dari Setiap Atribut

a. Atribut *Tavg (threshold = 21,65)*

Tavg ≤ 21,65:

Kategori: hujan (4) dan tidak hujan (1)

$$\begin{aligned}\text{Entropy}(S1) &= -\left(\frac{4}{5}\right) log_2\left(\frac{4}{5}\right) - \left(\frac{1}{5}\right) log_2\left(\frac{1}{5}\right)\\ &= -0{,}8 \cdot (-0{,}322) - 0{,}2 \cdot (-2{,}322)\\ &= 0{,}258 + 0{,}464\\ &= 0{,}722\end{aligned}$$

Tavg > 21,65:

Kategori: hujan (3) dan tidak hujan (2)

$$\begin{aligned}\text{Entropy}(S2) &= -\left(\frac{3}{5}\right) log_2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) log_2\left(\frac{2}{5}\right)\\ &= -0{,}6 \cdot (-0{,}737) - 0{,}4 \cdot (-1{,}322)\\ &= 0{,}422 + 0{,}529\\ &= 0{,}971\\ \text{Gain}(Tavg) &= 0{,}882 - \left(\frac{5}{10} \cdot 0{,}722 + \frac{5}{10} \cdot 0{,}971\right)\\ &= 0{,}882 - \ (0{,}361 + 0{,}486)\\ &= 0{,}882 - \ 0{,}847\\ &= 0{,}035\end{aligned}$$

b. Atribut *RH_avg (threshold = 88,5)*

RH_avg ≤ 88,5:

Kategori: hujan (4) dan tidak hujan (2)

$$\begin{aligned}\text{Entropy}(S1) &= -\left(\frac{4}{6}\right) log_2\left(\frac{4}{6}\right) - \left(\frac{2}{6}\right) log_2\left(\frac{2}{6}\right)\\ &= -0{,}667 \cdot (-0{,}585) - 0{,}333 \cdot (-1{,}585)\\ &= 0{,}390 + 0{,}528\\ &= 0{,}918\end{aligned}$$

Rh_avg > 88,5:

Kategori: hujan (3) dan tidak hujan (1)

$$\begin{aligned}\text{Entropy}(S2) &= -\left(\frac{3}{4}\right) log_2\left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) log_2\left(\frac{1}{4}\right)\\ &= -0{,}75 \cdot (-0{,}415) - 0{,}25 \cdot (-2)\\ &= 0{,}311 + 0{,}500\end{aligned}$$

$$= 0{,}811$$

$$\text{Gain}(RH_avg) = 0{,}882 - \left(\frac{6}{10} \cdot 0{,}918 + \frac{6}{10} \cdot 0{,}811\right)$$
$$= 0{,}882 - \ (0{,}551 + 0{,}324)$$
$$= 0{,}882 - \ 0{,}875$$
$$= 0{,}007$$

c. Atribut *ss (threshold = 3,95)*

ss ≤ 3,95:

Kategori: hujan (5) dan tidak hujan (1)

$$\text{Entropy}(S1) = -\left(\frac{5}{6}\right) log_2\left(\frac{5}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right)$$
$$= -0{,}833 \cdot (-0{,}263) - 0{,}167 \cdot (-2{,}585)$$
$$= 0{,}219 + 0{,}431$$
$$= 0{,}650$$

ss > 3,95:

Kategori: hujan (2) dan tidak hujan (2)

$$\text{Entropy}(S2) = -\left(\frac{2}{4}\right) log_2\left(\frac{2}{4}\right) - \left(\frac{2}{4}\right) log_2\left(\frac{2}{4}\right)$$
$$= -0{,}5 \cdot (-1) - 0{,}5 \cdot (-1)$$
$$= 1.000$$

$$\text{Gain}(ss) = 0{,}882 - \left(\frac{6}{10} \cdot 0{,}650 + \frac{6}{10} \cdot 1{,}000\right)$$
$$= 0{,}882 - \ (0{,}390 + 0{,}400)$$
$$= 0{,}882 - \ 0{,}790$$
$$= 0{,}092$$

d. Atribut *ddd_x (threshold = 290)*

ddd_x ≤ 290:

Kategori: hujan (4) dan tidak hujan (1)

$$\text{Entropy}(S1) = -\left(\frac{4}{5}\right) log_2\left(\frac{4}{5}\right) - \left(\frac{1}{5}\right) log_2\left(\frac{1}{5}\right)$$
$$= -0{,}8 \cdot (-0{,}322) - 0{,}2 \cdot (-2{,}322)$$
$$= 0{,}258 + 0{,}464$$

$$= 0{,}722$$

ddd_x > 290:

Kategori: hujan (3) dan tidak hujan (2)

$$\text{Entropy}(S2) = -\left(\frac{3}{5}\right) log_2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) log_2\left(\frac{2}{5}\right)$$
$$= -0{,}6 \cdot (-0{,}737) - 0{,}4 \cdot (-1{,}322)$$
$$= 0{,}442 + 0{,}529$$
$$= 0{,}971$$

$$\text{Gain}(ddd_x) = 0{,}882 - \left(\frac{5}{10}.0{,}722 + \frac{5}{10}.0{,}971\right)$$
$$= 0{,}882 - (0{,}361 + 0{,}486)$$
$$= 0{,}882 - 0{,}847$$
$$= 0{,}035$$

e. Atribut *ff_avg*

ff_avg terdiri dari 2 nilai unik yakni 0 dan 1.

ff_avg = 0 :

Kategori: hujan (3) dan tidak hujan (2)

$$\text{Entropy}(S1) = -\left(\frac{3}{5}\right) log_2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) log_2\left(\frac{2}{5}\right)$$
$$= -0{,}6 \cdot (-0{,}737) - 0{,}4 \cdot (-1{,}322)$$
$$= 0{,}442 + 0{,}529$$
$$= 0{,}971$$

ff_avg = 1:

Kategori: hujan (4) dan tidak hujan (1)

$$\text{Entropy}(S2) = -\left(\frac{4}{5}\right) log_2\left(\frac{4}{5}\right) - \left(\frac{1}{5}\right) log_2\left(\frac{1}{5}\right)$$
$$= -0{,}8 \cdot (-0{,}322) - 0{,}2 \cdot (-2{,}322)$$
$$= 0{,}258 + 0{,}464$$

$$= 0{,}722$$

$$\text{Gain}(ff_avg) = 0{,}882 - \left(\frac{5}{10}.0{,}971 + \frac{5}{10}.0{,}722\right)$$
$$= 0{,}882 - (0{,}486 + 0{,}361)$$
$$= 0{,}882 - 0{,}847$$
$$= 0{,}035$$

Berdasarkan hasil perhitungan *information gain,* fitur lama penyinaran matahari *(ss)* memiliki nilai *Gain* tertinggi, diikuti oleh suhu rata-rata *(Tavg),* kecepatan angin rata-rata *(ff_avg),* dan arah angin maksimum *(ddd_x)* yang memiliki nilai *Gain* sama besar. Sementara itu, atribut kelembapan rata-rata *(RH_avg)* menunjukkan kontribusi yang paling rendah terhadap pemisahan kelas. Dengan demikian, kita dapat memprioritaskan fitur-fitur seperti *ss, Tavg,* dan *ff_avg* dalam membangun model prediksi curah hujan, karena memiliki pengaruh paling signifikan terhadap hasil klasifikasi.

3.6 Skenario Pengujian

Skenario uji pada penelitian ini bertujuan untuk mengevaluasi performa model klasifikasi curah hujan yang dibangun menggunakan algoritma *C4.5*. Pengujian dilakukan dengan menggunakan data uji (*testing* data) yang telah dipisahkan dari data latih (*training* data) pada tahap sebelumnya. Skenario pengujian ini dilakukan untuk mengukur seberapa baik model yang telah dibentuk mampu mengklasifikasikan intensitas curah hujan berdasarkan data historis. Evaluasi model menggunakan *confusion matrix*.

3.6.1 *Confusion matrix*

Confusion matrix digunakan sebagai alat bantu visualisasi hasil klasifikasi model. Matriks ini menunjukkan distribusi prediksi benar dan salah untuk setiap kategori hujan (tidak hujan dan hujan).

Tabel 3.3 *Confusion Matrix* Hasil Klasifikasi Menggunakan Algoritma *C4.5*

		Prediksi	
		Hujan	Tidak Hujan
Aktual	Hujan	*True Positive* (TP)	*False Negative* (FN)
	Tidak Hujan	*False Positive* (FP)	*True Negative* (TN)

Keterangan:
TP(*True Positive*) = Data hujan yang diprediksi sebagai hujan.
TN(*True Negative*) = Data tidak hujan yang diprediksi sebagai tidak hujan.
FP(*False Positive*) = Data tidak hujan yang diprediksi sebagai hujan.
FN(*False Negative*) = Data hujan yang diprediksi sebagai tidak hujan.

A. Pengujian Akurasi Model

Akurasi digunakan untuk mengetahui seberapa besar proporsi klasifikasi yang dilakukan dengan benar oleh model terhadap seluruh data uji. Semakin tinggi akurasi, semakin baik performa model dalam mengklasifikasikan data.

Akurasi dihitung menggunakan formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan:
TP = *True Positive.*
TN = *True Negative.*
FP = *False Positive.*
FN = *False Negative.*

B. Pengujian Presisi *(Precision)*

Presisi menunjukkan proporsi data yang diprediksi sebagai hujan dan benar-benar hujan. Nilai ini penting terutama dalam sistem yang mengutamakan minimisasi kesalahan positif palsu.

Presisi dihitung menggunakan formula:

$$Precision = \frac{TP}{TP + FP}$$

C. Pengujian *Recall*

Recall mengukur sebarapa baik model dalam menangkap semua data yang memang tergolong dalam kategori hujan.semakin besar nilai *recall*, semakin baik model dalam mengenali data yang benar-benar hujan, meskipun presisinya belum tentu tinggi. *Recall* dihitung menggunakan formula:

$$Recall = \frac{TP}{TP + FN}$$

D. Pengujian *F1-Score*

F1-Score adalah rata-rata harmonis dari precision dan recall. Digunakan sebagai ukuran gabungan untuk mengevaluasi keseimbangan antara presisi dan recall. *F1-Score* dihitung menggunakan formula:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan evaluasi sistem klasifikasi curah hujan menggunakan algoritma *Decision Tree C4.5*. Data yang digunakan berasal dari catatan iklim harian Kabupaten Bogor yang telah melalui tahapan *preprocessing*, seleksi fitur, pelatihan, dan pengujian model. Sistem dikembangkan menggunakan bahasa pemrograman *python* di *Google Colab* dengan pustaka *Scikit-learn* dan *Pandas*. Evaluasi dilakukan berdasarkan lima skenario pembagian data latih dan uji untuk mengukur konsistensi performa model dalam membedakan antara kondisi 'Hujan' dan 'Tidak Hujan'. Analisis dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* guna menilai efektivitas dan keseimbangan prediksi.

4.1 Implementasi Sistem

Implementasi sistem dilakukan melalui beberapa tahapan utama, yang secara sistematis mendukung proses pembangunan model klasifikasi curah hujan berbasis algoritma *Decision Tree C4.5*. Setiap tahapan dirancang untuk memastikan bahwa data yang digunakan dapat diolah secara efektif dan menghasilkan model yang akurat. Tahapan-tahapan implementasi meliputi:

A. *Import Library* dan *Load Dataset*

Langkah awal dalam pengembangan sistem klasifikasi curah hujan adalah melakukan inisialisasi lingkungan kerja dengan mengimpor berbagai pustaka

(*library*) yang dibutuhkan. Pustaka tersebut mencakup modul standar *python* serta pustaka pihak ketiga seperti *pandas*, *scikit-learn*, dan *matplotlib*, yang memiliki peran penting dalam proses analisis data, implementasi algoritma *machine learning*, serta penyajian visualisasi data. Pemanfaatan pustaka ini bertujuan untuk menunjang efisiensi dan akurasi dalam pengolahan data iklim. Adapun data yang digunakan berasal dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dengan rentang waktu tahun 2019 hingga 2020, yang terdiri atas 735 entri data yang merepresentasikan kondisi cuaca di wilayah Kabupaten Bogor.

Setelah tahap inisialisasi selesai dilakukan, proses selanjutnya difokuskan pada akses dan pembacaan *dataset* yang telah disimpan di *Google Drive*. Mengingat proyek ini dikembangkan menggunakan *platform Google Colab*, maka diperlukan proses autentikasi dan integrasi terhadap akun *Google Drive* pengguna agar file data dalam format *xlsx* dapat diakses secara langsung. Rangkaian proses tersebut meliputi.

a. Mengimpor pustaka yang diperlukan untuk menghubungkan *Google Colab* dengan *Google Drive*, seperti (*google.colab.drive*).
b. Membuat fungsi atau skrip untuk memastikan bahwa file *dataset* (xlsx) tersedia dan dapat diakses dari direktori *Google Drive* yang telah di-*mount*.
c. Membaca *file xlsx* menggunakan pustaka *pandas* dan memuatnya ke dalam struktur *DataFrame* untuk memudahkan proses eksplorasi, manipulasi, dan analisis data selanjutnya.

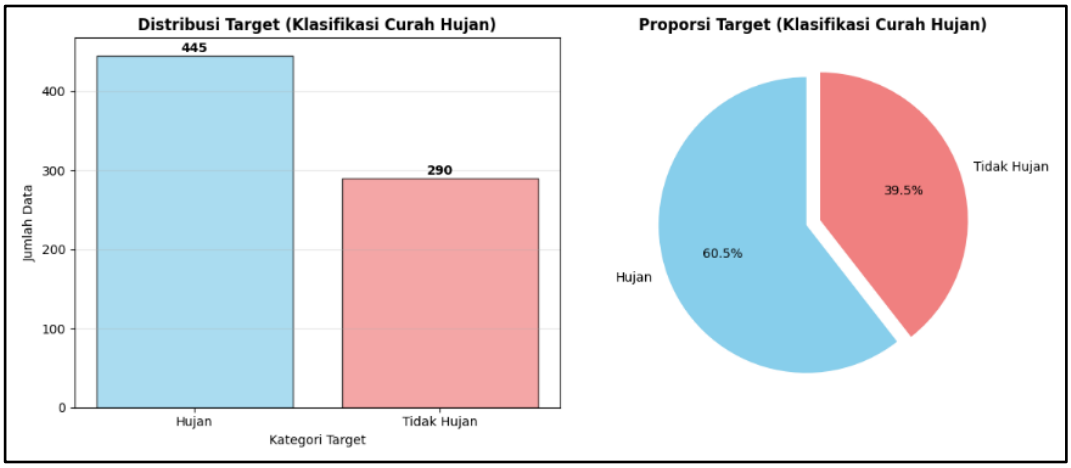
B. EDA (*Exploratory Data Analysis*)

Tahap *Exploratory Data Analysis* bertujuan untuk memahami struktur, karakteristik, dan distribusi data sebelum memasuki proses pemodelan. Berdasarkan informasi awal *dataset*, diketahui bahwa data terdiri dari 735 entri dengan 11 kolom yang mencakup data numerik dan kategorikal. Seluruh kolom tidak memiliki nilai yang hilang, sehingga data dinyatakan lengkap dan siap untuk dianalisis lebih lanjut. Kolom-kolom seperti *Tn, Tx, Tavg, RH_avg, ss, ddd_x,* dan *ff_avg* merupakan data numerik bertipe *float* atau *integer*, sementara kolom *date*, *ddd_car*, dan Target bertipe objek.

Analisis statistik deskriptif menunjukkan bahwa rata-rata temperatur (*Tavg*) berada di kisaran 21,6°C dengan variasi antara 19,3°C hingga 26,5°C. Kelembapan rata-rata (*RH_avg*) berkisar antara 81% hingga 99%, mencerminkan kondisi kelembapan tinggi yang umum terjadi di wilayah tropis. Sementara itu, durasi penyinaran matahari (*ss*) menunjukkan rentang dari 0 hingga 10,6 jam, yang menunjukkan variasi signifikan dalam intensitas penyinaran. Parameter angin seperti arah maksimum (*ddd_x*) memiliki distribusi dari 0° hingga 360°, sedangkan kecepatan rata-rata angin (*ff_avg*) relatif rendah, dengan rata-rata 0,9 m/s.

Selanjutnya, dilakukan analisis terhadap distribusi kelas target, yaitu klasifikasi kondisi cuaca menjadi ‘Hujan’ dan ‘Tidak Hujan’. Hasilnya menunjukkan bahwa dari 735 data, sebanyak 445 (60,5%) termasuk dalam kelas ‘Hujan’ dan 290 (39,5%) termasuk dalam kelas ‘Tidak Hujan’. Proporsi ini ditampilkan dalam bentuk diagram batang dan diagram lingkaran untuk memperjelas persebaran kelas target. Meskipun terdapat sedikit

ketidakseimbangan, distribusi tersebut masih tergolong cukup proporsional dan layak untuk digunakan dalam model klasifikasi tanpa memerlukan teknik penyeimbangan data tambahan.



Gambar 4.1 Visualisasi Proses Distribusi dan Proporsi Target

C. *Preprocessing* Data

Tahapan *preprocessing* data bertujuan untuk menyiapkan data mentah agar siap digunakan dalam proses klasifikasi. Langkah pertama dilakukan dengan memeriksa adanya nilai kosong (*missing values*) pada setiap kolom dalam *dataset*. Hasil pemeriksaan menunjukkan bahwa seluruh atribut tidak memiliki nilai yang hilang, sehingga tidak diperlukan penanganan lebih lanjut terhadap data yang tidak lengkap. Selanjutnya, dilakukan pemilihan fitur (*feature selection*) yang relevan untuk proses klasifikasi. Fitur yang dipilih meliputi *Tavg* (temperatur rata-rata dalam °C), *RH_avg* (kelembapan rata-rata dalam %), *ss* (lamanya penyinaran matahari dalam jam), *ddd_x* (arah angin saat kecepatan maksimum dalam derajat), dan *ff_avg* (kecepatan angin rata-rata dalam m/s). Pemilihan fitur ini didasarkan pada pengaruhnya terhadap kondisi cuaca, khususnya curah hujan.

Setelah fitur ditentukan, data difilter agar hanya mencakup kolom-kolom yang dibutuhkan, yaitu lima fitur terpilih dan kolom target yang berisi label klasifikasi, yakni ‘Hujan’ dan ‘Tidak Hujan’. *Dataset* kemudian dibersihkan kembali dengan menghapus seluruh baris yang mungkin mengandung nilai kosong (meskipun dalam kasus ini tidak ditemukan). Setelah proses pembersihan, data yang tersisa sebanyak 735 baris. Selanjutnya, dilakukan pemisahan antara fitur (variabel independen) dan target (variabel dependen). Variabel fitur berisi lima atribut cuaca yang digunakan sebagai input, sedangkan target adalah kondisi curah hujan. Pada tahap ini juga diketahui bahwa data klasifikasi terdiri dari dua kelas, yaitu ‘Hujan’ dan ‘Tidak Hujan’.

D. Pembagian Data

Secara umum, hasil evaluasi menunjukkan bahwa semakin besar proporsi data latih, semakin baik performa model. Hal ini konsisten dengan prinsip bahwa model pembelajaran mesin cenderung lebih akurat ketika dilatih dengan data yang lebih banyak. Oleh karena itu, pemilihan skenario dengan proporsi data latih yang lebih besar, seperti 90:10 atau 80:20, direkomendasikan untuk menghasilkan model klasifikasi yang optimal.

Untuk mengevaluasi performa model klasifikasi curah hujan secara lebih menyeluruh, dilakukan pengujian dengan lima skenario berbeda berdasarkan rasio pembagian data pelatihan dan data pengujian. Pendekatan ini bertujuan untuk mengamati kestabilan model dalam berbagai proporsi data serta mengetahui sejauh mana perubahan jumlah data pelatihan dan pengujian dapat memengaruhi hasil klasifikasi.

Setiap skenario memiliki rasio pembagian data yang berbeda, yaitu:

a. Skenario 1: 90% data digunakan untuk pelatihan dan 10% untuk pengujian (rasio 90:10)
b. Skenario 2: 80% pelatihan dan 20% pengujian (rasio 80:20)
c. Skenario 3: 70% pelatihan dan 30% pengujian (rasio 70:30)
d. Skenario 4: 60% pelatihan dan 40% pengujian (rasio 60:40)
e. Skenario 5: 50% pelatihan dan 50% pengujian (rasio 50:50).

Pada tahap ini, dilakukan pemisahan data berdasarkan rasio yang telah ditentukan untuk masing-masing skenario. Selanjutnya, proses pelatihan dan pengujian algoritma *Decision Tree C4.5* dilakukan pada setiap pembagian. Hasil dari masing-masing skenario disimpan dalam sebuah struktur data yang mencakup informasi rasio pembagian, ukuran data pelatihan dan pengujian, serta nilai-nilai evaluasi seperti akurasi, *precision*, *recall*, dan *f1-score*. Pendekatan ini memungkinkan peneliti untuk membandingkan performa model secara objektif dan memilih konfigurasi pembagian data yang paling optimal untuk kasus klasifikasi curah hujan.

E. Pelatihan Model

Setelah data dibagi ke dalam lima skenario pengujian, tahap selanjutnya adalah melakukan pelatihan model klasifikasi menggunakan algoritma *Decision Tree C4.*5. Pada implementasi ini, model dikembangkan dengan memanfaatkan pustaka *DecisionTreeClassifier* dari *scikit-learn*, di mana kriteria pemilihan split menggunakan *entropy* untuk mencerminkan metode *information gain* yang menjadi dasar dari algoritma *C4.5*.

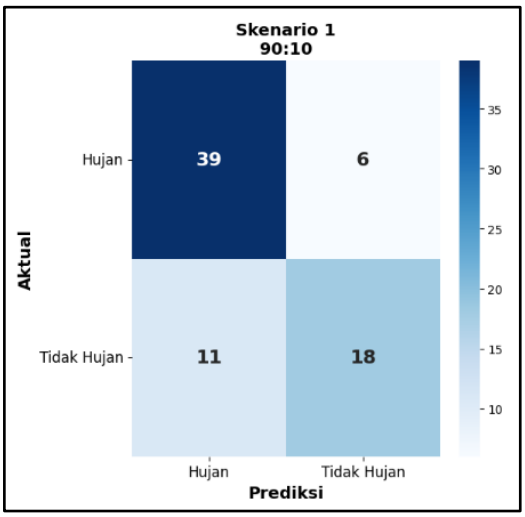
Setiap skenario pengujian diawali dengan proses pemisahan data pelatihan dan pengujian sesuai rasio masing-masing. Selanjutnya, model dilatih menggunakan data pelatihan yang telah ditentukan.

4.2 Hasil Uji Coba

Pengujian dilakukan sebanyak lima kali dengan cara mengubah proporsi pembagian data latih dan data uji. Tujuan dari pengujian ini adalah untuk mengetahui pengaruh rasio pembagian data terhadap akurasi model klasifikasi curah hujan menggunakan algoritma *Decision Tree C4.5*. Nilai *test_size* yang digunakan secara berturut-turut adalah 0.1, 0.2, 0.3, 0.4, dan 0.5.

A. Hasil Evaluasi 1 (Rasio 90:10)

Pengujian model pertama dilakukan dengan rasio pembagian data 90% untuk data latih sebesar 661 data dan 10% untuk data uji sebessar 74 data. Model menggunakan lima atribut input yaitu suhu rata-rata (*Tavg*), kelembapan rata-rata (*RH_avg*), durasi penyinaran (*ss*), arah angin saat kecepatan maksimum (*ddd_x*) dan kecepatan angin rata-rata (*ff_avg*). Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 77%. Detail hasil evaluasi model ditunjukkan pada Gambar 4.2.

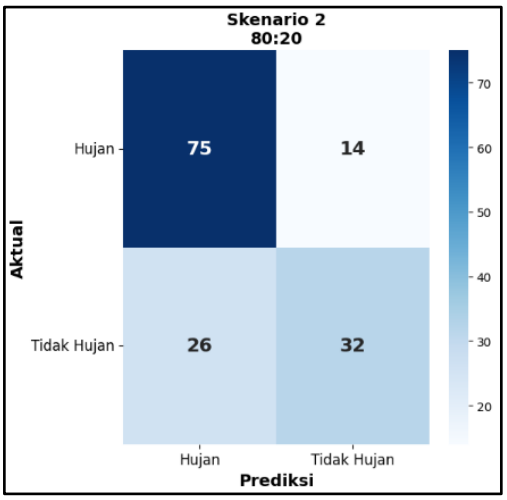


Gambar 4.2 Hasil Pengujian Model 1

Berdasarkan Gambar 4.2 yang menunjukkan hasil pengujian Model Pertama, terdapat 18 data kategori "Tidak Hujan" yang berhasil diprediksi dengan benar oleh sistem (*True Negative*), dan 39 data kategori "Hujan" yang juga berhasil diklasifikasikan dengan tepat (*True Positive*). Namun, terdapat 6 data aktual "Hujan" yang salah diklasifikasikan sebagai "Tidak Hujan" (*False Negative*), serta 11 data aktual "Tidak Hujan" yang salah diklasifikasikan sebagai "Hujan" (*False Positive*). Berdasarkan distribusi hasil klasifikasi tersebut, sistem kemudian dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Hasil pengujian menunjukkan bahwa model memperoleh nilai akurasi sebesar 77,03%, *precision* sebesar 78,00%, *recall* sebesar 86,67%, dan *f1-score* sebesar 82,08%. Nilai-nilai evaluasi ini mengindikasikan bahwa model memiliki performa yang baik dan seimbang dalam mengklasifikasikan data curah hujan, dengan tingkat kesalahan yang relatif rendah.

B. Hasil Evaluasi 2 (Rasio 80:20)

Pengujian model kedua dilakukan dengan rasio pembagian data 80% untuk data latih sebesar 588 data dan 20% untuk data uji sebesar 147 data. Model menggunakan lima atribut input yaitu suhu rata-rata (*Tavg*), kelembapan rata-rata (*RH_avg*), durasi penyinaran (*ss*), arah angin saat kecepatan maksimum (*ddd_x*) dan kecepatan angin rata-rata (*ff_avg*). Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 73%. Detail hasil evaluasi model ditunjukkan pada Gambar 4.3.



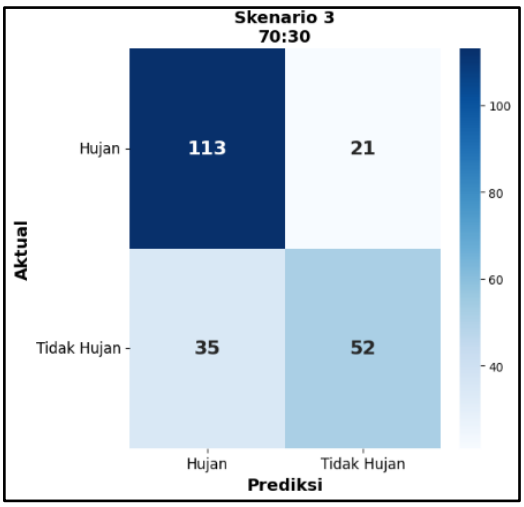
Gambar 4.3 Hasil Pengujian Model 2

Berdasarkan Gambar 4.3 yang menunjukkan hasil pengujian Model kedua, terdapat 32 data kategori "Tidak Hujan" yang berhasil diprediksi dengan benar oleh sistem (*True Negative*), dan 75 data kategori "Hujan" yang juga berhasil diklasifikasikan dengan tepat (*True Positive*). Namun, terdapat 14 data aktual "Hujan" yang salah diklasifikasikan sebagai "Tidak Hujan" (*False Negative*), serta 26 data aktual "Tidak Hujan" yang salah diklasifikasikan sebagai "Hujan" (*False Positive*). Dari hasil klasifikasi tersebut, dilakukan evaluasi performa model

menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Berdasarkan hasil pengujian tersebut, diperoleh nilai akurasi sebesar 72,79%, *precision* sebesar 74,26%, *recall* sebesar 84,27%, dan *f1-score* sebesar 78,95%. Nilai evaluasi ini menunjukkan bahwa model memiliki performa yang baik dan seimbang dalam mengklasifikasikan data curah hujan, meskipun terdapat sedikit penurunan akurasi dibandingkan skenario sebelumnya.

C. Hasil Evaluasi 3 (Rasio 70:30)

Pengujian model ketiga dilakukan dengan rasio pembagian data 70% untuk data latih sebesar 514 data dan 30% untuk data uji sebesar 221. Model menggunakan lima atribut input yaitu suhu rata-rata (*Tavg*), kelembapan rata-rata (*RH_avg*), durasi penyinaran (*ss*), arah angin saat kecepatan maksimum (*ddd_x*) dan kecepatan angin rata-rata (*ff_avg*). Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 75%. Detail hasil evaluasi model ditunjukkan pada Gambar 4.4.

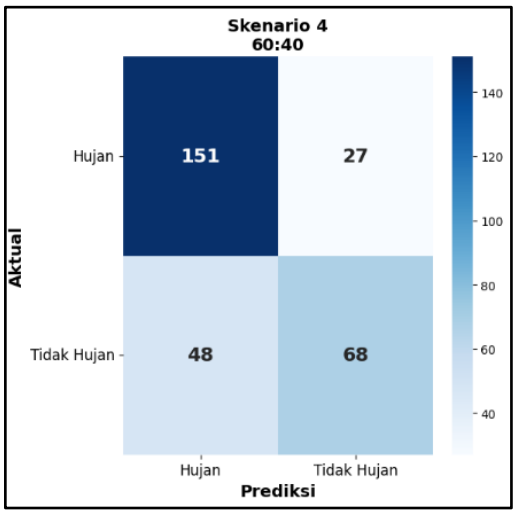


Gambar 4.4 Hasil Pengujian Model 3

Berdasarkan Gambar 4.4 yang menunjukkan hasil pengujian Model ketiga, terdapat 52 data kategori "Tidak Hujan" yang berhasil diprediksi dengan benar oleh sistem (*True Negative*), dan 113 data kategori "Hujan" yang juga berhasil diklasifikasikan dengan tepat (*True Positive*). Namun, terdapat 21 data aktual "Hujan" yang salah diklasifikasikan sebagai "Tidak Hujan" (*False Negative*), serta 35 data aktual "Tidak Hujan" yang salah diklasifikasikan sebagai "Hujan" (*False Positive*). Berdasarkan hasil klasifikasi tersebut, dilakukan evaluasi performa model menggunakan beberapa metrik penilaian. Dari evaluasi tersebut, diperoleh nilai akurasi sebesar 74,77%, *precision* sebesar 76,35%, *recall* sebesar 84,33%, dan *f1-score* sebesar 80,13%. Nilai evaluasi ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang cukup baik dan konsisten, dengan keseimbangan yang cukup stabil antara ketepatan prediksi dan kemampuan mengenali data hujan.

D. Hasil Evaluasi 4 (Rasio 60:40)

Pengujian model keempat dilakukan dengan rasio pembagian data 60% untuk data latih sebesar 441 data dan 40% untuk data uji sebesar 294 data. Model menggunakan lima atribut input yaitu suhu rata-rata (*Tavg*), kelembapan rata-rata (*RH_avg*), durasi penyinaran (*ss*), arah angin saat kecepatan maksimum (*ddd_x*) dan kecepatan angin rata-rata (*ff_avg*). Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 74%. Detail hasil evaluasi model ditunjukkan pada Gambar 4.5.

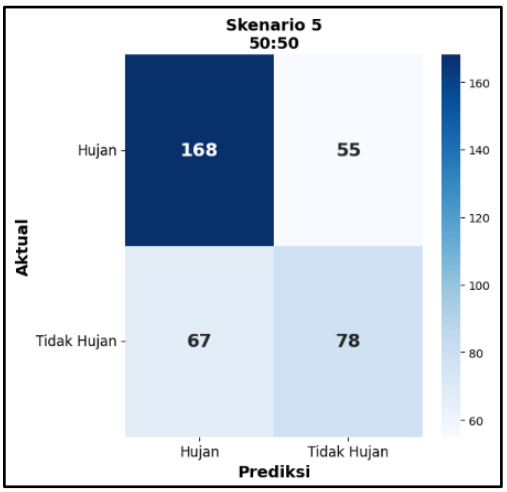


Gambar 4.5 Hasil Pengujian Model 4

Berdasarkan Gambar 4.5 yang menunjukkan hasil pengujian Model keempat, terdapat 68 data kategori "Tidak Hujan" yang berhasil diprediksi dengan benar oleh sistem (*True Negative*), dan 151 data kategori "Hujan" yang juga berhasil diklasifikasikan dengan tepat (*True Positive*). Namun, terdapat 27 data aktual "Hujan" yang salah diklasifikasikan sebagai "Tidak Hujan" (*False Negative*), serta 48 data aktual "Tidak Hujan" yang salah diklasifikasikan sebagai "Hujan" (*False Positive*). Berdasarkan hasil klasifikasi tersebut, dilakukan evaluasi terhadap performa model menggunakan sejumlah metrik pengukuran. Dari pengujian tersebut, diperoleh nilai akurasi sebesar 74,49%, *precision* sebesar 75,88%, *recall* sebesar 84,83%, dan *f1-score* sebesar 80,07%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan data curah hujan secara efektif, dengan performa yang relatif stabil meskipun akurasi sedikit menurun.

E. Hasil Evaluasi 5 (Rasio 50:50)

Pengujian model kelima dilakukan dengan rasio pembagian data 50% untuk data latih sebesar 367 data dan 50% untuk data uji sebesar 368 data. Model menggunakan lima atribut input yaitu suhu rata-rata (*Tavg*), kelembapan rata-rata (*RH_avg*), durasi penyinaran (*ss*), arah angin saat kecepatan maksimum (*ddd_x*) dan kecepatan angin rata-rata (*ff_avg*). Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 67%. Detail hasil evaluasi model ditunjukkan pada Gambar 4.6.



Gambar 4.6 Hasil Pengujian Model 5

Berdasarkan Gambar 4.6 yang menunjukkan hasil pengujian Model kelima, terdapat 78 data kategori "Tidak Hujan" yang berhasil diprediksi dengan benar oleh sistem (*True Negative*), dan 168 data kategori "Hujan" yang juga berhasil diklasifikasikan dengan tepat (*True Positive*). Namun, terdapat 55 data aktual "Hujan" yang salah diklasifikasikan sebagai "Tidak Hujan" (*False Negative*), serta 67 data aktual "Tidak Hujan" yang salah diklasifikasikan sebagai "Hujan" (*False Positive*). Berdasarkan distribusi klasifikasi tersebut, dilakukan evaluasi terhadap

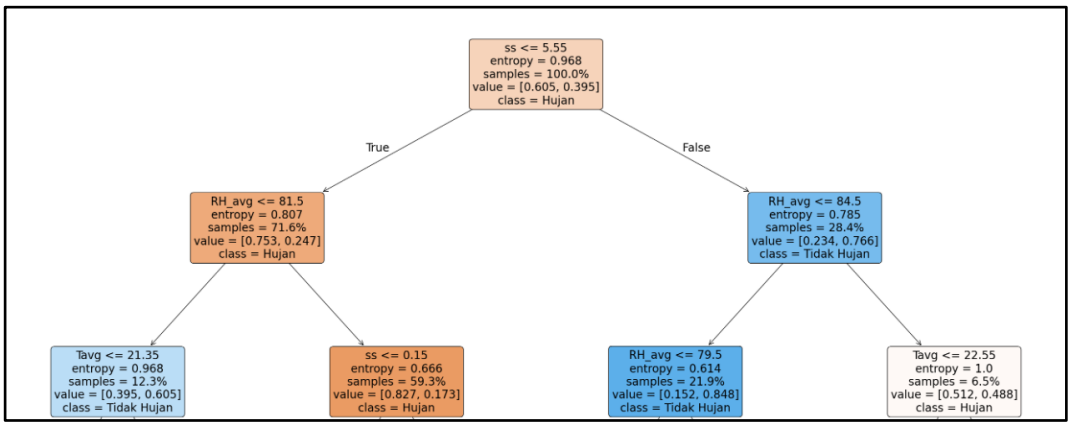
performa model menggunakan metrik pengujian. Dari hasil pengujian tersebut, diperoleh nilai akurasi sebesar 66,85%, *precision* sebesar 71,49%, *recall* sebesar 75,34%, dan *f1-score* sebesar 73,31%. Nilai evaluasi ini menunjukkan bahwa performa model cenderung menurun pada proporsi data pelatihan dan pengujian yang seimbang, namun model masih dapat mempertahankan keseimbangan yang cukup baik antara presisi dan sensitivitas terhadap data hujan.

4.2.1 Visualisasi *Decision Tree*

Visualisasi pohon keputusan bertujuan untuk memperlihatkan bagaimana model klasifikasi berbasis algoritma *Decision Tree C4.5* memetakan data masukan ke dalam kelas target, yaitu “Hujan” dan “Tidak Hujan”. Dalam penelitian ini, visualisasi pohon keputusan dilakukan sebanyak lima kali dengan variasi rasio pembagian data latih dan data uji, yaitu: 90:10, 80:20, 70:30, 60:40, dan 50:50.

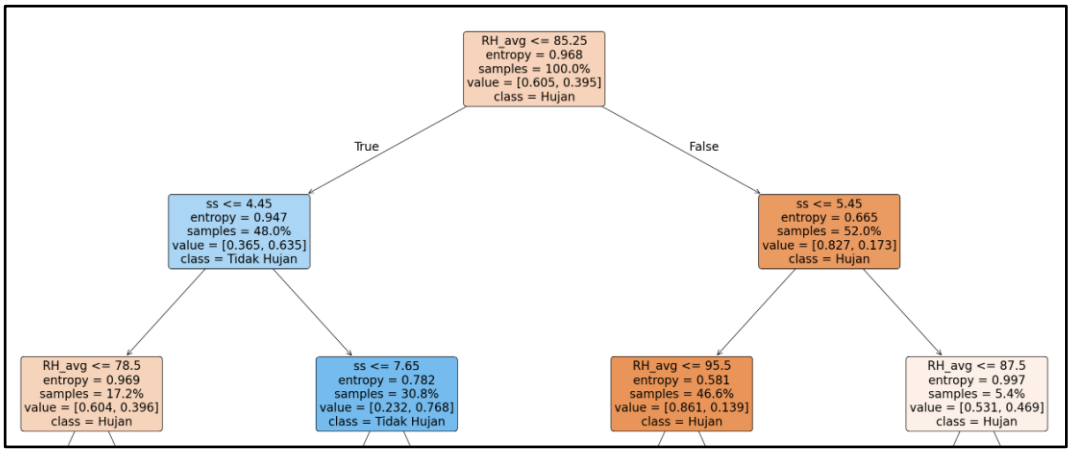
Setiap visualisasi menunjukkan struktur pohon keputusan hasil pelatihan model, dengan node yang mewakili atribut dan *threshold* pemisahan. Setiap cabang terbentuk berdasarkan nilai *information gain* tertinggi, dan proses ini berlanjut hingga pohon mencapai kondisi maksimum sesuai kedalaman *(max_depth = 10)* yang ditentukan dalam pelatihan.

Berikut adalah hasil visualisasi *decision tree* untuk masing-masing rasio data.



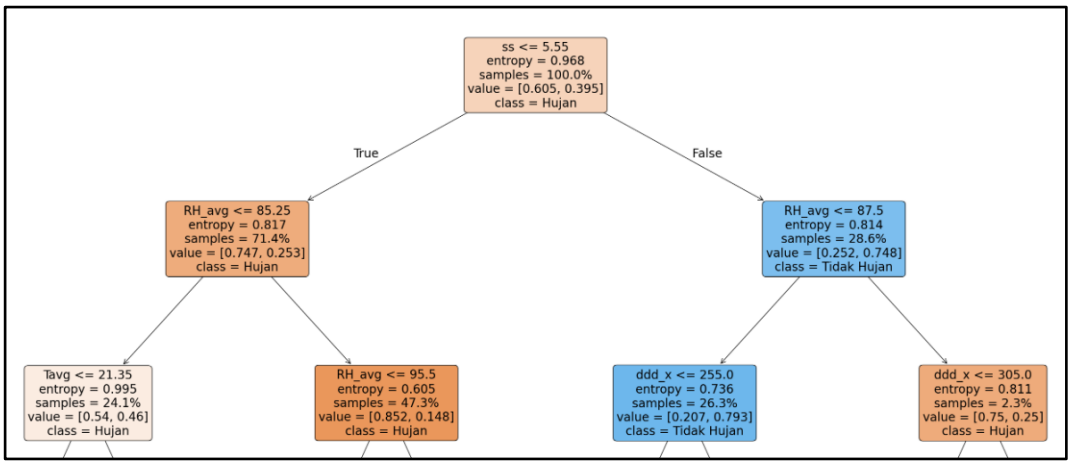
Gambar 4.7 Visualisasi Pohon Keputusan (Rasio 90:10)

Visualisasi pohon keputusan pada Gambar 4.7 dengan rasio pembagian data 90:10 menunjukkan struktur pohon yang stabil dan cukup mendalam, dengan atribut *ss* (lama penyinaran matahari) sebagai akar pohon. Visualisasi lengkap pohon keputusan dapat dilihat pada Lampiran 1. Setelah akar pohon, fitur *RH_avg*, *Tavg*, dan *ddd_x* digunakan dalam pemisahan lanjutan. Proporsi data latih yang besar (90%) memungkinkan model mempelajari pola secara lebih menyeluruh, sehingga menghasilkan pohon yang kompleks namun tetap akurat. Proses pemisahan kelas tampak tajam dan efisien, mencerminkan performa pembelajaran yang baik.



Gambar 4.8 Visualisasi Pohon Keputusan (Rasio 80:20)

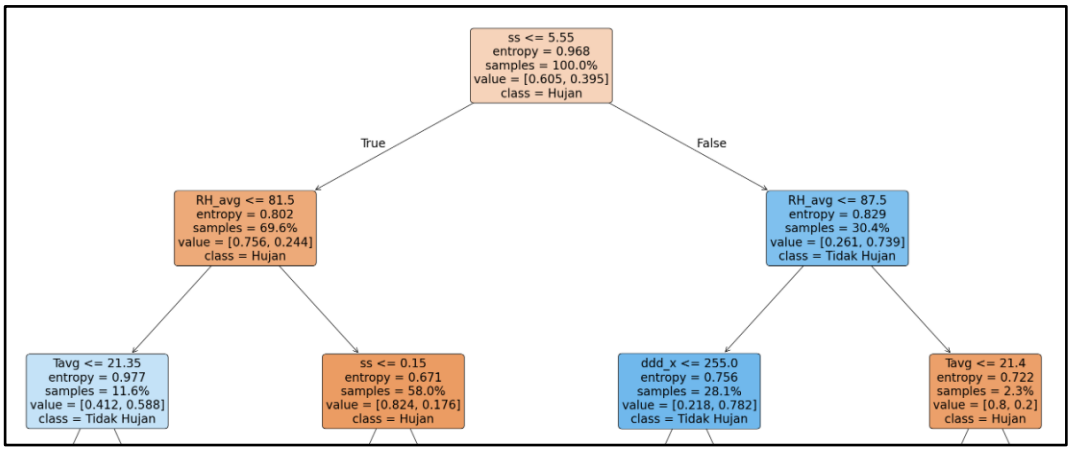
Visualisasi pohon keputusan pada Gambar 4.8 dengan rasio pembagian data 80:20 memperlihatkan struktur pohon yang masih cukup stabil, dengan atribut *RH_avg* (kelembapan rata-rata) sebagai akar pohon. Visualisasi lengkap dari struktur pohon keputusan ini ditampilkan pada Lampiran 2. Selanjutnya, atribut *ss*, *Tavg*, dan *ddd_x* digunakan sebagai pemisah lanjutan dalam proses klasifikasi. Meskipun proporsi data latih sedikit berkurang dibandingkan skenario sebelumnya, model tetap mampu mengenali pola dengan baik dan menghasilkan struktur pohon yang konsisten serta efisien dalam memisahkan kelas hujan dan tidak hujan.



Gambar 4.9 Visualisasi Pohon Keputusan (Rasio 70:30)

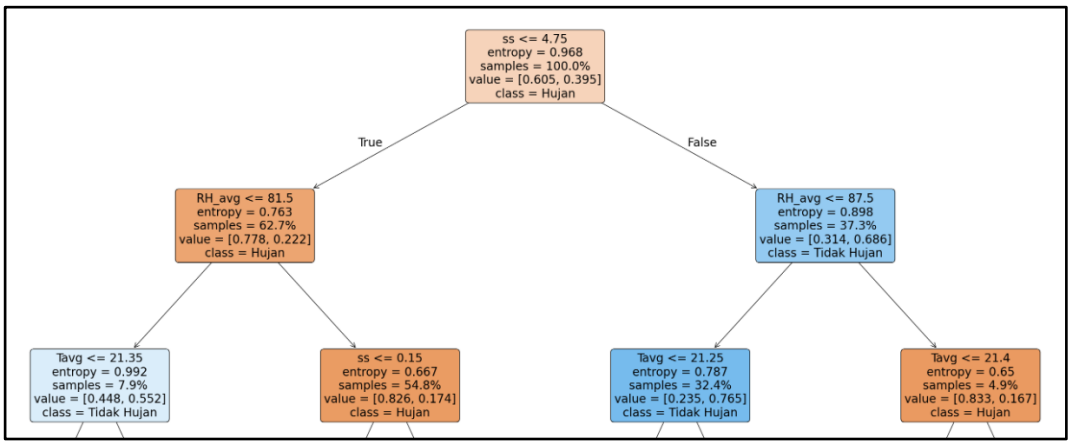
Visualisasi pohon keputusan pada Gambar 4.9 dengan rasio pembagian data 70:30 menunjukkan bahwa atribut *ss* (lama penyinaran matahari) menjadi akar pohon, menggantikan *RH_avg* yang muncul sebagai akar pada skenario sebelumnya. Visualisasi lengkap pohon keputusan ini dapat dilihat pada Lampiran 3. Perubahan ini mengindikasikan kemungkinan adanya fluktuasi dalam pemilihan fitur utama ketika proporsi data latih berkurang. Meskipun demikian, model tetap menunjukkan performa yang baik. Atribut seperti *ddd_x* dan *Tavg* muncul pada

cabang bawah pohon, menandakan bahwa model masih mampu mengenali pola relevan untuk klasifikasi curah hujan secara efektif.



Gambar 4.10 Visualisasi Pohon Keputusan (Rasio 60:40)

Visualisasi pohon keputusan pada Gambar 4.10 dengan rasio pembagian data 60:40 menunjukkan struktur yang relatif mirip dengan skenario sebelumnya, di mana atribut *ss* (lama penyinaran matahari) kembali muncul sebagai akar pohon. Visualisasi lengkap pohon keputusan ditampilkan pada Lampiran 4. Atribut *RH_avg*, *Tavg*, dan *ddd_x* digunakan sebagai pemisah lanjutan. Menariknya, meskipun proporsi data latih semakin kecil, model masih mampu menghasilkan klasifikasi yang cukup akurat. Kompleksitas struktur pohon pada skenario ini sedikit meningkat dibandingkan sebelumnya, mencerminkan upaya model dalam mengkompensasi keterbatasan informasi melalui pemisahan yang lebih rinci.



Gambar 4.11 Visualisasi Pohon Keputusan (Rasio 50:50)

Visualisasi pohon keputusan pada Gambar 4.11 dengan rasio pembagian data 50:50 menunjukkan bahwa struktur pohon tetap stabil meskipun jumlah data latih dan data uji seimbang. Atribut *ss* (lama penyinaran matahari) kembali menjadi akar pohon, dengan *RH_avg*, *Tavg*, dan *ddd_x* sebagai fitur pemisah lanjutan. Visualisasi lengkap pohon keputusan ditampilkan pada Lampiran 5. Model masih mampu mempelajari keterkaitan antar fitur meskipun data latih terbatas. Namun, penurunan akurasi yang cukup signifikan pada skenario ini mengindikasikan bahwa informasi yang tersedia dalam data latih belum cukup kuat untuk menghasilkan generalisasi yang optimal. Struktur pohon yang relatif dalam juga menunjukkan bahwa model memerlukan lebih banyak pemisahan untuk mencapai hasil klasifikasi yang baik.

Dari kelima visualisasi pohon keputusan pada skenario dengan rasio data latih dan uji yang berbeda, dapat disimpulkan bahwa atribut *RH_avg* (kelembapan rata-rata), *ss* (lama penyinaran matahari), dan *Tavg* (suhu rata-rata) secara konsisten muncul sebagai fitur pemisah utama, menunjukkan bahwa ketiganya memiliki peran penting dalam proses klasifikasi curah hujan. Konsistensi ini sesuai dengan

hasil perhitungan *information gain* yang menempatkan atribut-atribut tersebut sebagai yang paling informatif. Selain itu, perubahan struktur dan kedalaman pohon pada tiap skenario mencerminkan kemampuan adaptif algoritma *C4.5* dalam menyesuaikan kompleksitas model terhadap jumlah data latih yang tersedia. Pada skenario dengan data latih lebih banyak (seperti 90:10), pohon lebih stabil dan akurat, sedangkan pada rasio seimbang (50:50), model menghasilkan struktur pohon yang lebih dalam sebagai upaya untuk mempertahankan performa klasifikasi. Hal ini menunjukkan pentingnya proporsi data latih yang cukup untuk menghasilkan model yang optimal.

4.3 Pembahasan

Hasil pengujian model klasifikasi curah hujan menggunakan algoritma *Decision Tree C4.5* dilakukan melalui lima skenario pembagian data pelatihan dan pengujian dengan rasio sebagai berikut: 90:10, 80:20, 70:30, 60:40, dan 50:50. Setiap skenario menunjukkan performa model yang relatif stabil, meskipun terdapat variasi nilai akurasi dan metrik evaluasi lainnya.

Model mencapai akurasi tertinggi sebesar 77,03% pada Skenario 1 (90% data latih dan 10% data uji), sedangkan akurasi terendah tercatat sebesar 66,85% pada Skenario 5 (50% data latih dan 50% data uji). Hasil evaluasi lengkap dari setiap skenario dapat dilihat pada Tabel 4.1.

Tabel 4.1 Hasil Evaluasi Model Klasifiaksi *C4.5*

Skenario	**Akurasi**	***Precision***	***Recall***	***F1-score***
90:10	77.03%	78.00%	86.67%	82.11%
80:20	72.79%	74.26%	84.27%	78.95%
70:30	74.66%	76.35%	84.33%	80.14%
60:40	74.49%	75.88%	84.83%	80.11%
50:50	66.85%	71.49%	75.34%	73.36%

Dari hasil evaluasi tersebut, terlihat bahwa nilai *precision*, *recall*, dan *f1-score* pada skenario terbaik (90:10) menunjukkan kinerja yang seimbang. Hal ini menunjukkan bahwa model tidak hanya akurat, tetapi juga mampu mengenali kedua kelas target, yaitu Hujan dan Tidak Hujan, secara proporsional.

Selain itu, hasil visualisasi pohon keputusan menunjukkan bahwa atribut kelembapan rata-rata (*RH_avg*), lamanya penyinaran matahari (*ss*), dan temperatur rata-rata (*Tavg*) secara konsisten muncul sebagai node utama pada pohon keputusan di semua skenario. Hal ini sesuai dengan hasil perhitungan *information gain,* di mana ketiga fitur tersebut memberikan kontribusi paling besar dalam proses klasifikasi.

Dari sisi struktur, pohon keputusan yang dihasilkan pada skenario dengan data latih lebih besar (90:10 dan 80:20) cenderung lebih sederhana dan efisien. Sebaliknya, pada skenario dengan data latih yang lebih kecil (60:40 dan 50:50), struktur pohon menjadi lebih dalam dan kompleks. Hal ini mengindikasikan bahwa semakin sedikit data latih yang digunakan, semakin besar kemungkinan model mengalami *overfitting*, yang dapat menurunkan kemampuan generalisasi terhadap data baru.

Dengan demikian, dapat disimpulkan bahwa jumlah data latih yang memadai sangat berperan penting dalam membangun model klasifikasi yang andal. Skenario dengan rasio 90:10 terbukti memberikan performa terbaik dalam penelitian ini, baik dari segi akurasi maupun keseimbangan antar metrik evaluasi.

4.4 Integrasi Islam

Islam menempatkan ilmu pengetahuan sebagai elemen yang menyatu dengan keimanan. Hal ini ditegaskan melalui wahyu pertama yang diturunkan dalam Al-Qur'an, yaitu perintah *Iqra'* (bacalah), yang menjadi pondasi awal bagi lahirnya tradisi keilmuan dalam Islam. Umat Islam didorong untuk mempelajari dan merenungi alam semesta sebagai bentuk ibadah intelektual dan pengakuan atas kebesaran Allah SWT. Dalam kaitannya dengan penelitian ini, integrasi antara sains dan nilai-nilai Islam menjadi pijakan fundamental dalam pengembangan sistem klasifikasi curah hujan, sebagai bagian dari ikhtiar manusia dalam memahami serta merawat ciptaan-Nya.

Sebagai bentuk penerapan nilai-nilai tersebut dalam penelitian ini, pendekatan integrasi Islam dilakukan melalui tiga sudut pandang utama: (1) aspek teologis yang menggarisbawahi bahwa seluruh fenomena alam berada dalam kekuasaan Allah SWT, (2) aspek antropologis yang menekankan pentingnya teknologi sebagai sarana untuk memberikan manfaat sosial bagi kehidupan manusia, dan (3) aspek ekologis yang menekankan peran manusia dalam menjaga keseimbangan dan kelestarian lingkungan sebagai bentuk amanah dari Sang Pencipta.

4.4.1 Mu'amalah Ma'a Allah

Muamalah dengan Allah mencerminkan hubungan spiritual antara manusia dan Sang Pencipta. Dalam konteks penelitian ini, proses pengkajian fenomena alam seperti curah hujan merupakan bentuk penghambaan dan perenungan terhadap kekuasaan Allah SWT. Memahami dan mengklasifikasikan pola hujan melalui

pendekatan ilmiah merupakan manifestasi dari perintah untuk menggunakan akal dalam memahami tanda-tanda kebesaran Allah.

أَلَمْ تَرَ أَنَّ ٱللَّهَ يُزْجِى سَحَابًا ثُمَّ يُؤَلِّفُ بَيْنَهُۥ ثُمَّ يَجْعَلُهُۥ رُكَامًا فَتَرَى ٱلْوَدْقَ يَخْرُجُ مِنْ خِلَٰلِهِۦ وَيُنَزِّلُ مِنَ ٱلسَّمَآءِ مِن جِبَالٍ فِيهَا مِنۢ بَرَدٍ فَيُصِيبُ بِهِۦ مَن يَشَآءُ وَيَصْرِفُهُۥ عَن مَّن يَشَآءُ ۖ يَكَادُ سَنَا بَرْقِهِۦ يَذْهَبُ بِٱلْأَبْصَٰرِ

"Tidaklah kamu melihat bahwa Allah mengarak awan, kemudian mengumpulkan antara (bagian-bagian)nya, kemudian menjadikannya bertindih-tindih, maka kelihatanlah olehmu hujan keluar dari celah-celahnya dan Allah (juga) menurunkan (butiran-butiran) es dari langit, (yaitu) dari (gumpalan-gumpalan awan seperti) gunung-gunung, maka ditimpakan-Nya (butiran-butiran) es itu kepada siapa yang dikehendaki-Nya dan dipalingkan-Nya dari siapa yang dikehendaki-Nya. Kilauan kilat awan itu hampir-hampir menghilangkan penglihatan", (QS. An-Nur: 43).

Ayat ini menjelaskan bagaimana Allah SWT secara bertahap mengatur proses terbentuknya hujan, dimulai dari menggerakkan awan, mengumpulkannya, hingga menumpuk dan menurunkannya ke bumi (Tafsir Al-Muyassar, 2025). Penafsiran ini memberikan makna bahwa turunnya hujan merupakan hasil dari sistem ciptaan Allah yang sangat teratur dan dapat diamati secara ilmiah. Dalam konteks penelitian ini, penerapan algoritma untuk mengklasifikasikan curah hujan merupakan bagian dari ikhtiar manusia dalam memahami sunnatullah. Hal ini bukan dimaksudkan untuk menandingi kekuasaan-Nya, melainkan sebagai bentuk penghambaan serta pemanfaatan ilmu yang telah dianugerahkan oleh Allah SWT.

4.4.2 Mu'amalah Ma'a An-Nas

Pengetahuan yang diperoleh dari penelitian ini tidak hanya bermanfaat bagi diri sendiri, melainkan juga bagi masyarakat luas. Dengan memanfaatkan teknologi klasifikasi cuaca berbasis data, penelitian ini berpotensi besar membantu berbagai

sektor, seperti pertanian, perencanaan kota, mitigasi bencana, dan manajemen sumber daya air.

ذَٰلِكَ ٱلْكِتَٰبُ لَا رَيْبَ ۛ فِيهِ ۛ هُدًى لِّلْمُتَّقِينَ

"Kitab (Al Quran) ini tidak ada keraguan padanya; petunjuk bagi mereka yang bertakwa,", (QS. Al-Baqarah: 2).

Ayat ini menegaskan bahwa Al-Qur'an merupakan petunjuk yang menyeluruh bagi kehidupan manusia, mencakup urusan dunia dan akhirat (Tafsir Ibnu Katsir, 2025). Penafsiran ini memberikan makna bahwa di dalam Al-Qur'an terdapat dorongan untuk menuntut ilmu, mengembangkan potensi, serta memberikan manfaat bagi sesama. Dalam konteks penelitian ini, penerapan klasifikasi curah hujan mencerminkan upaya manusia dalam memanfaatkan ilmu pengetahuan untuk mendukung sektor pertanian, mitigasi bencana, dan perencanaan pembangunan. Semua ini bertujuan untuk mewujudkan kesejahteraan umat manusia, yang sejalan dengan prinsip Islam bahwa ilmu yang paling utama adalah ilmu yang memberikan manfaat bagi orang lain.

4.4.3 Mu'amalah Ma'a Alam

Islam menegaskan bahwa manusia adalah khalifah di muka bumi yang bertanggung jawab atas kelestarian lingkungan. Perubahan iklim global, yang memengaruhi pola curah hujan, sebagian besar disebabkan oleh aktivitas manusia. Penelitian ini merupakan bagian dari upaya untuk memahami dampak tersebut dan mencari solusi ilmiah dalam pengelolaan lingkungan.

وَلَا تُفْسِدُوا۟ فِى ٱلْأَرْضِ بَعْدَ إِصْلَٰحِهَا وَٱدْعُوهُ خَوْفًا وَطَمَعًا ۚ إِنَّ رَحْمَتَ ٱللَّهِ قَرِيبٌ مِّنَ ٱلْمُحْسِنِينَ

"Dan janganlah kamu membuat kerusakan di muka bumi, sesudah (Allah) memperbaikinya dan berdoalah kepada-Nya dengan rasa takut (tidak akan diterima) dan harapan (akan dikabulkan). Sesungguhnya rahmat Allah amat dekat kepada orang-orang yang berbuat baik." (QS. Al-A'raf: 56).

Ayat ini mengandung larangan bagi manusia untuk tidak membuat kerusakan di bumi setelah Allah menciptakannya dalam keadaan seimbang dan baik (Tafsir As-Sa'di, 2025). Penafsiran ini memberikan makna bahwa larangan tersebut tidak hanya mencakup kerusakan fisik, tetapi juga eksploitasi alam yang berlebihan serta penyalahgunaan ilmu pengetahuan. Oleh karena itu, penggunaan data iklim untuk klasifikasi curah hujan tidak hanya merupakan langkah ilmiah, tetapi juga bentuk tanggung jawab moral dan spiritual. Upaya ini mencerminkan kepedulian terhadap kelestarian lingkungan hidup, agar bumi tetap menjadi tempat yang layak bagi seluruh makhluk ciptaan-Nya.

Integrasi Islam dalam penelitian ini menjadi landasan filosofis bahwa ilmu pengetahuan bukanlah entitas yang berdiri sendiri, tetapi terikat erat dengan nilai-nilai spiritual dan etika. Penelitian tentang klasifikasi curah hujan ini merupakan bentuk pengamalan tauhid dalam bidang sains, di mana teknologi digunakan untuk memahami fenomena alam, memberi manfaat kepada manusia, dan menjaga keberlangsungan ciptaan Allah SWT.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan membangun model klasifikasi curah hujan berdasarkan data historis cuaca menggunakan algoritma *Decision Tree C4.5*. Berdasarkan implementasi dan pengujian terhadap lima skenario pembagian data latih dan data uji, model menunjukkan performa terbaik pada skenario 90:10 dengan akurasi 77,03%, *precision* 78,00%, *recall* 86,67%, dan *f1-score* 82,08%. Evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score* menunjukkan performa yang cukup stabil, terutama saat proporsi data latih lebih besar. Atribut kelembapan rata-rata (*RH_avg*), lama penyinaran matahari (*ss*), dan suhu rata-rata (*Tavg*) konsisten menjadi node pemisah utama, menandakan peran pentingnya dalam proses klasifikasi.

Visualisasi pohon keputusan memberikan gambaran yang jelas mengenai logika klasifikasi yang dilakukan model, sehingga memudahkan interpretasi hasil secara transparan. Secara keseluruhan, algoritma *Decision Tree C4.5* terbukti cukup efektif untuk klasifikasi curah hujan dan memiliki potensi untuk dikembangkan lebih lanjut dalam sistem prediksi berbasis data iklim di masa depan.

5.2 Saran

Penelitian ini masih memiliki ruang pengembangan yang cukup luas. Beberapa saran yang dapat diberikan untuk penelitian selanjutnya antara lain:

a. Menerapkan pendekatan *ensemble* seperti *Random Forest* atau *XGBoost* guna meningkatkan akurasi dan memperkuat ketahanan model terhadap *overfitting*.

b. Menambahkan atribut cuaca tambahan seperti tekanan udara, curah hujan hari sebelumnya, atau indeks fenomena iklim global seperti *El Niño–La Niña* untuk memperkaya fitur yang digunakan dalam klasifikasi.

c. Mengembangkan model klasifikasi multikelas,misalnya mengklasifikasikan curah hujan menjadi hujan ringan, sedang, lebat, hingga ekstrem, agar hasil klasifikasi lebih informatif dan bermanfaat dalam mitigasi bencana dan perencanaan pertanian.

d. Mengintegrasikan model ke dalam aplikasi berbasis web atau mobile, sehingga dapat digunakan secara langsung oleh masyarakat, lembaga pemerintah, atau pihak lain yang berkepentingan dalam pemantauan dan prediksi cuaca berbasis data.

Dengan pengembangan lebih lanjut, model ini berpotensi memberikan kontribusi nyata dalam mendukung sistem informasi cuaca dan pengambilan keputusan berbasis data di berbagai sektor.

DAFTAR PUSTAKA

Al Arif, A., Firdaus, M., Maruhawa, Y., AMIK Riau, S., & Purwodadi Panam, J. (2022). Comparison of Data Mining Methods for Prediction of Rainfall with C4.5, Naïve Bayes, and KNN Algorithm. https://journal.irpi.or.id/index.php/sentimas

Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning. In Jurnal Ilmiah Ilmu Komputer (Vol. 13, Issue 1).

Fikry Ubaidillah, Dedi Cahyadi, & Mochammad Rifai. (2023). Implementasi Data Mining pada Dataset Prakiraan Cuaca Menggunakan Algoritma C4.5. Jurnal Teknologi Dan Sistem Informasi, 4(2), 123–131.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems).

Hidayat, R. , & A. K. (2014). Variabilitas Curah Hujan Indonesia dan Hubungannya dengan ENSO/IOD: Estimasi Menggunakan Data JRA-25/JCDAS. Jurnal Agromet Indonesia, 28, 1–8. https://doi.org/http://journal.ipb.ac.id/index.php/agromet/article/view/9972

Ibnu Daqiqil Id. (2021). Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python.

Lakshmi, B. N., Indumathi, T. S., & Ravi, N. (2016). A Study on C.5 Decision Tree Classification Algorithm for Risk Predictions During Pregnancy. Procedia Technology, 24, 1542–1549. https://doi.org/10.1016/j.protcy.2016.05.128

Lou, D., Yang, M., Shi, D., Wang, G., Ullah, W., Chai, Y., & Chen, Y. (2021). K-means and c4.5 decision tree based prediction of long-term precipitation variability in the poyang lake basin, china. Atmosphere, 12(7). https://doi.org/10.3390/atmos12070834

Merdekawati, G. I., & Ismail. (2019). Prediksi Curah Hujan Di Jakarta Berbasis Algoritma Levenberg Marquardt. Jurnal Ilmiah Informatika Komputer, 24(2), 116–128. https://doi.org/10.35760/ik.2019.v24i2.2366

Nabila Amelia Siregar. (2024). Implementasi Algoritma Decision Tree C4.5 Dalam Mengklasifikasi Data Curah Hujan UPT. BMKG Kota Medan.

Nafisatun Najah, A., Nur Amalina, A., & Hidayat, R. (2025). Penerapan Algoritma Decision Tree C4.5 untuk Prediksi Cuaca di Kota Semarang. INDEXIA: Informatic and Computational Intelligent Journal, 07(01), 45–52.

Nasrullah, M. F., Saedudin, Rd. R., & Hamami, F. (2024a). Perbandingan Akurasi Algoritma C4.5 dan K-Nearest Neighbors untuk Klasifikasi Curah Hujan Berdasarkan Iklim Indonesia. JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika), 9(2), 628–638. https://doi.org/10.29100/jipi.v9i2.4655

Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. JURIKOM (Jurnal Riset Komputer), 7(1), 56. https://doi.org/10.30865/jurikom.v7i1.1947

Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip. In Jurnal Mahasiswa Teknik Informatika (Vol. 7, Issue 1). https://doi.org/10.36040/jati.v7i1.6303

Risnawati, I., Rahma, S., Kusuma, F., Salsabila4, N., Nourmansyah5, A., Farhan6, A., & Irfiani, E. (2023a). Klasifikasi Data Mining Untuk Mengestimasi Potensi Curah Hujan Berdampak Banjir Daerah Menggunakan Algoritma C4.5. Jurnal INSAN (Journal of Information Systems Management Innovation, 3(2). http://jurnal.bsi.ac.id/index.php/jinsan

Ruswanti, D. (2020). Pengukuran Performa Support Vector Machine Dan Neural Network Dalam Meramalkan Tingkat Curah Hujan. In Gaung Informatika (Vol. 13, Issue 1).

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. In SN Computer Science (Vol. 2, Issue 3). Springer. https://doi.org/10.1007/s42979-021-00592-x

Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian (Vol. 1, Issue 2).

Sathyanarayanan, S. (2024). Confusion Matrix-Based Performance Evaluation Metrics. African Journal of Biomedical Research, 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345

Tafsir Al-Muyassar. (2025, June 29). Tafsir QS. An-Nur Ayat 43. https://tafsirweb.com/5642-surat-an-nur-ayat-43.html

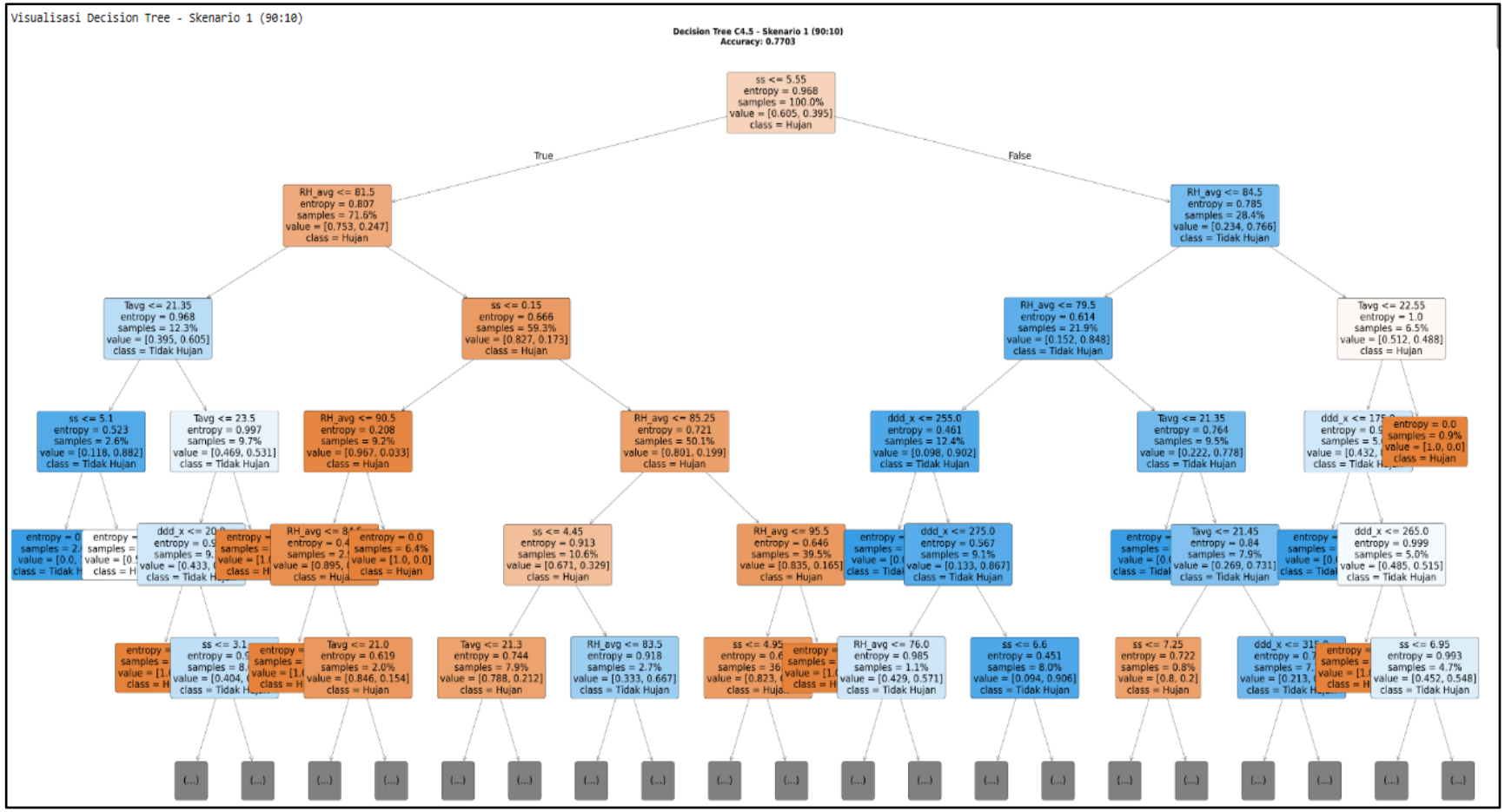
Tafsir As-Sa'di. (2025, June 29). Tafsir QS. Al-A'raf Ayat 56. https://tafsirweb.com/2526-surat-al-araf-ayat-56.html

Tafsir Ibnu Katsir. (2025, June 29). Tafsir QS. Al-Baqarah Ayat 2. https://tafsirweb.com/116-surat-al-baqarah-ayat-2.html

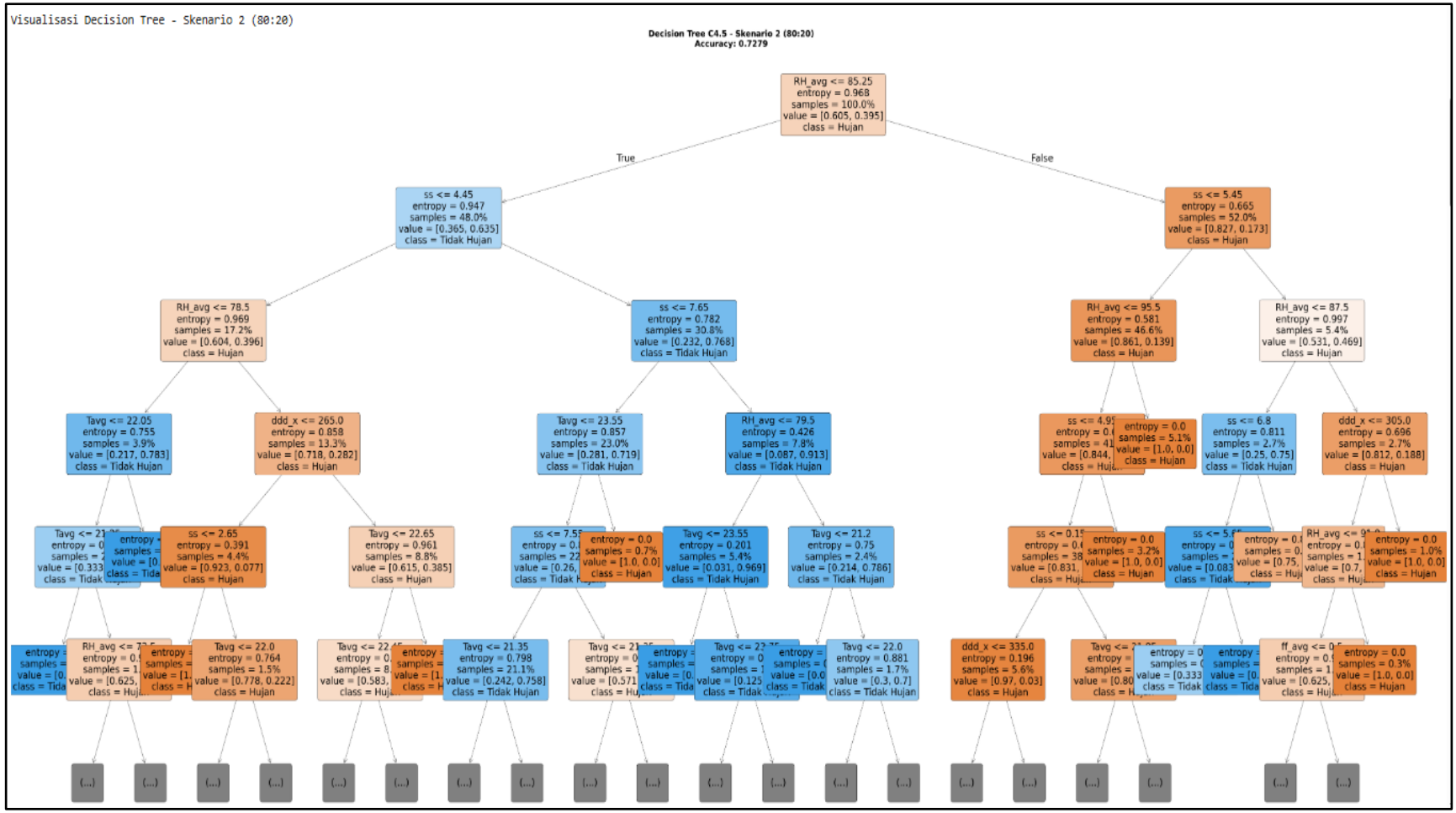
Wandayantolis, Nandang Pangaribowo, Sirajul Munir, Masagus Ismail Zulfiandy, Tenike Nanza Apria, Winesty Dewi Nurputri, Raga Ramanda Syailendra, Dwi Ratnawati, Shinta Mediany, Widyasari, Nikitasha Gema Yunanda, & Siska Masrury. (2023). Buletin Analisis Hujan November 2023 dan Prakiraan Hujan Januari, Februari, dan Maret 2024 di Sumatera Selatan. http://iklim.sumsel.bmkg.go.id

LAMPIRAN

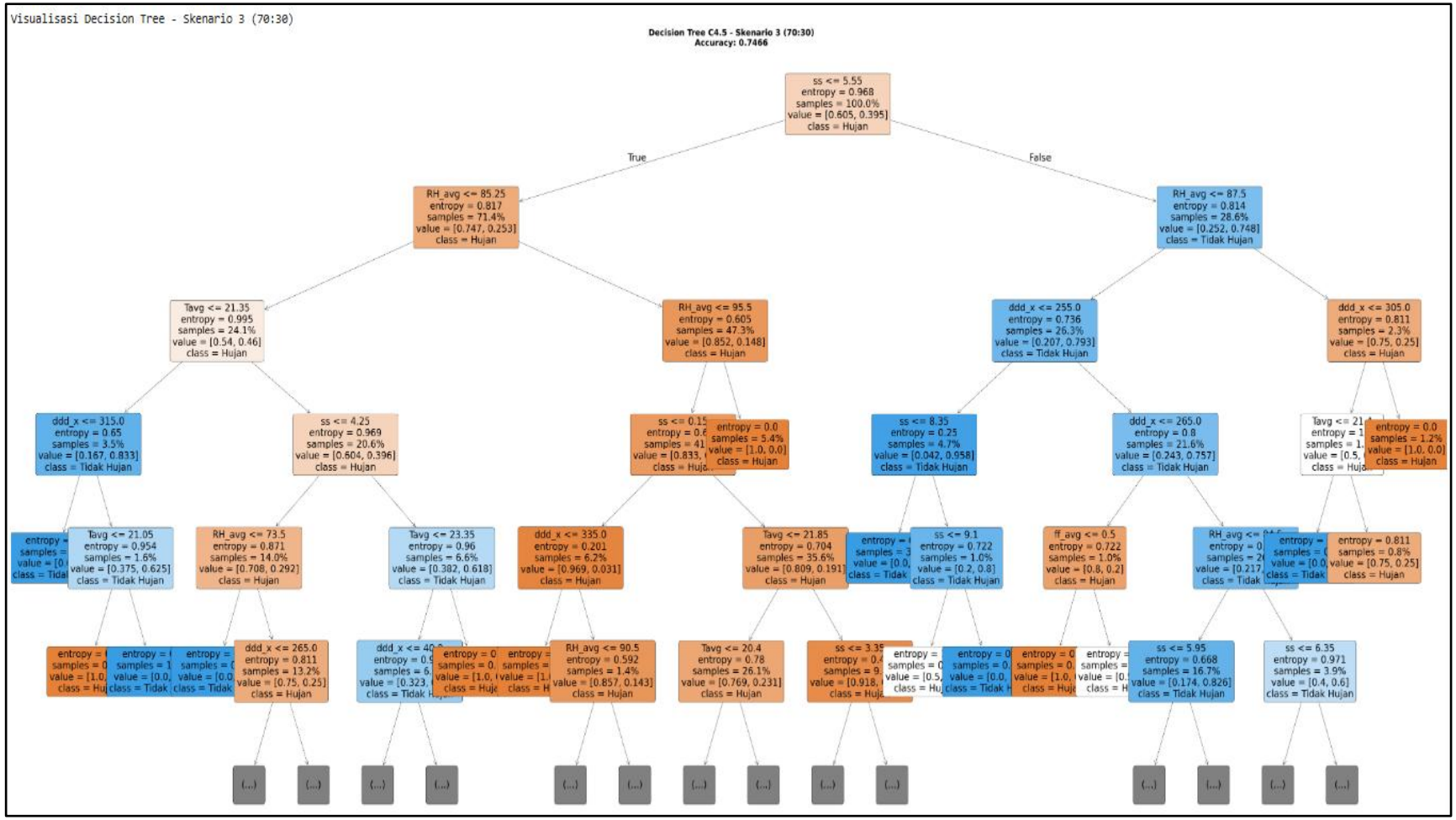
Lampiran 1. Visualisasi *Decision Tree C4.5* Skenario 1 Rasio (90:10)



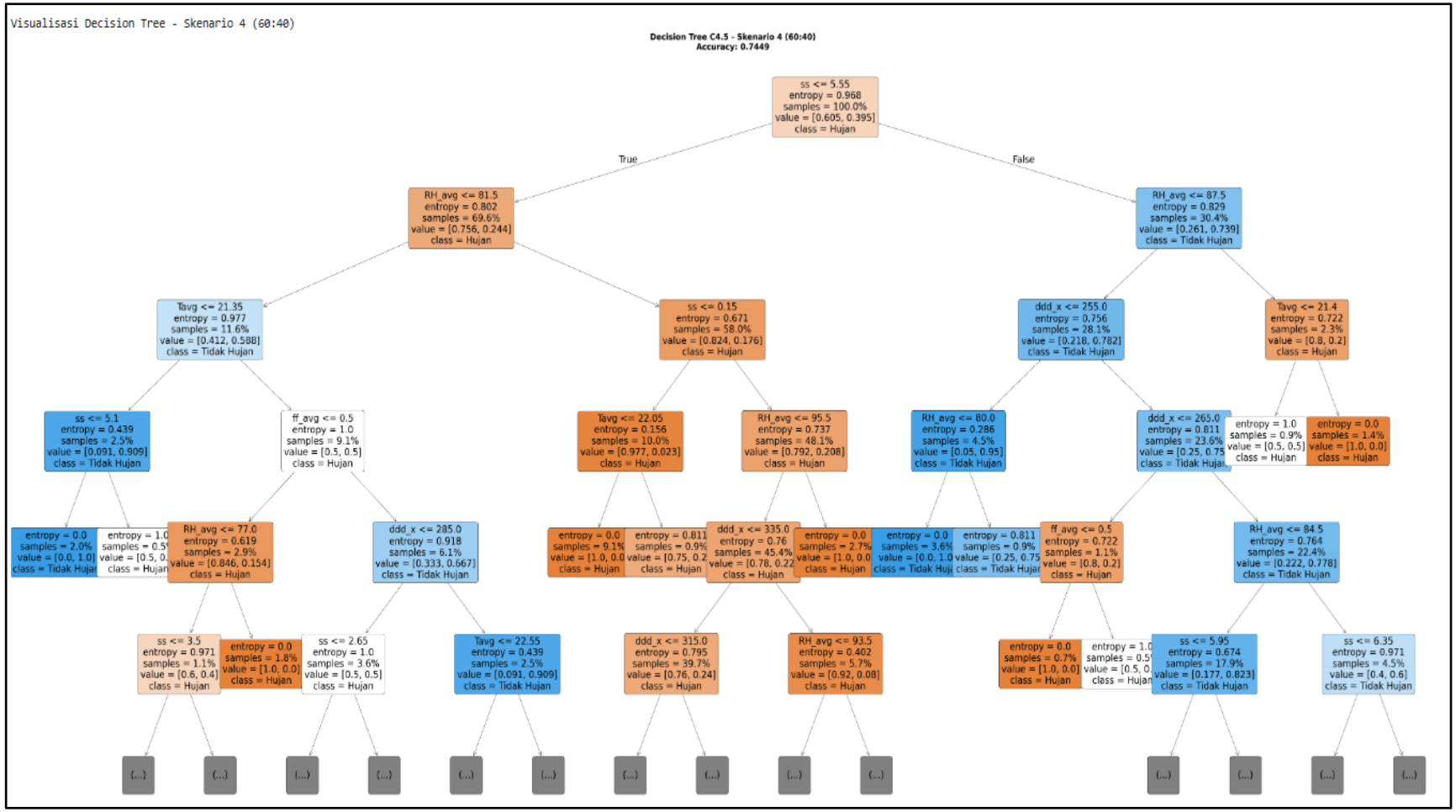
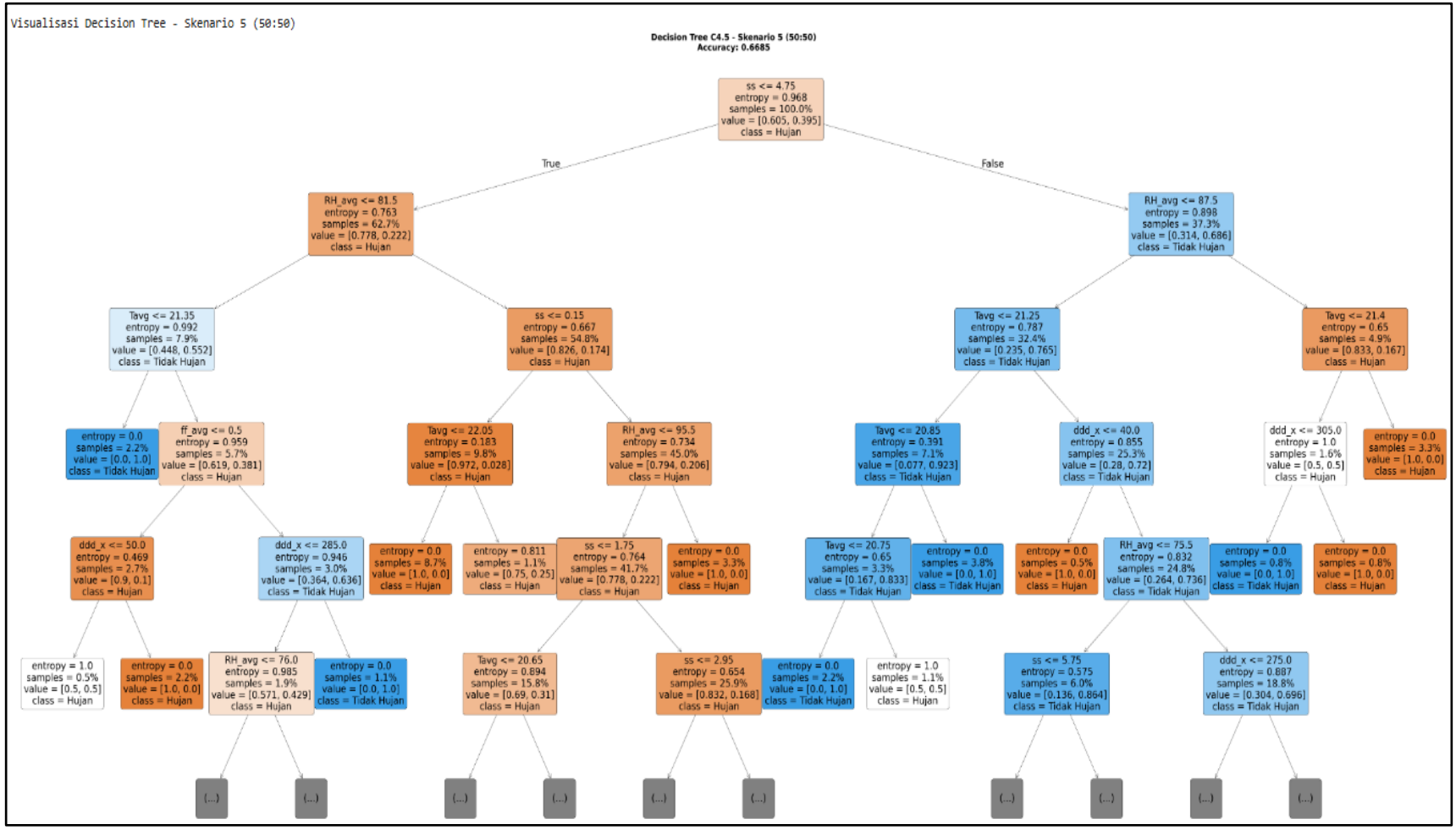
Lampiran 2. Visualisasi *Decision Tree C4.5* Skenario 2 Rasio (80:20)



Lampiran 3. Visualisasi *Decision Tree C4.5* Skenario 3 Rasio (70:30)



Lampiran 4. Visualisasi *Decision Tree C4.5* Skenario 4 Rasio (60:40)



Lampiran 5. Visualisasi *Decision Tree C4.5* Skenario 5 Rasio (50:50)