

Perbandingan Model Decision Tree dan Random Forest untuk Penentuan Kesesuaian Lahan Budidaya Cabai dan Terong

Astiah Amir¹, Fachruddin Fachruddin^{1*}, Fadli Idris¹, Meylis Safriani¹, Reskiana Saefuddin², Indera Sakti Nasution³, Sanusi⁴, Mawaddah Putri Arisma Siregar⁵

¹ Fakultas Teknik, Program Studi Teknik Sipil, Universitas Teuku Umar, Meulaboh, Indonesia

² Badan Riset dan Inovasi Nasional, Jl. Raya Jakarta-Bogor km 46, Bogor, Indonesia

³ Fakultas Pertanian, Program Teknik Pertanian, Universitas Syiah Kuala, Banda Aceh, Indonesia

⁴ Fakultas Teknik, Program Teknologi Informatika, Universitas Teuku Umar, Aceh Barat, Indonesia

⁵ Fakultas Pertanian, Program Studi Agroteknologi, Universitas Teuku Umar, Aceh Barat, Indonesia

Email: ¹astiahmir@utu.ac.id, ²fachruddin@utu.ac.id, ³fadliidris@utu.ac.id, ⁴meylissafriani@utu.ac.id, ⁵resk001@brin.go.id,

⁶i.nasution@usk.ac.id, ⁷sanusi@utu.ac.id, ⁸mawaddahputriarisma@utu.ac.id

Email Penulis Korespondensi: fachruddin@utu.ac.id

Submitted 28-05-2025; Accepted 23-07-2025; Published 31-10-2025

Abstrak

Kabupaten Aceh Barat memiliki potensi besar dalam budidaya tanaman hortikultura seperti cabai dan terong, meskipun karakteristik tanah gambut dengan tingkat keasaman tinggi dan variabilitas lingkungan menjadi tantangan utama. Penentuan kesesuaian lahan yang akurat memerlukan analisis berbagai variabel seperti pH tanah, kelembaban tanah dan udara, curah hujan, serta tekstur tanah. Penelitian ini bertujuan mengembangkan model klasifikasi kesesuaian lahan menggunakan algoritma Decision Tree dan Random Forest untuk tanaman cabai dan terong di wilayah tersebut. Data lingkungan dan karakteristik tanah dianalisis menggunakan kedua metode tersebut untuk mengevaluasi performa klasifikasi. Hasil penelitian menunjukkan bahwa algoritma Random Forest unggul dengan akurasi mencapai 99% pada klasifikasi lahan cabai, serta nilai precision dan recall yang lebih tinggi dibandingkan Decision Tree. Untuk klasifikasi lahan terong, kedua algoritma menunjukkan performa sempurna dengan akurasi dan metrik evaluasi mencapai 1.00 tanpa kesalahan klasifikasi. Keunggulan Random Forest terletak pada kemampuannya menangani variabel input yang kompleks dan mengurangi risiko overfitting melalui ensemble pohon keputusan, sehingga menghasilkan prediksi yang lebih stabil dan andal. Dengan demikian, Random Forest sangat cocok digunakan dalam sistem klasifikasi kesesuaian lahan berbasis data lingkungan di Aceh Barat, mendukung pengambilan keputusan budidaya yang lebih optimal dan berkelanjutan. Penelitian ini memberikan kontribusi penting dalam penerapan teknologi machine learning untuk meningkatkan efisiensi dan hasil produksi pertanian di wilayah dengan karakteristik tanah gambut yang menantang.

Kata Kunci: Kesesuaian Lahan; Random Forest; Decision Tree; Budidaya Cabai dan Terong

Abstract

Kabupaten Aceh Barat has great potential for cultivating horticultural crops such as chili peppers and eggplants, despite the challenges posed by peat soil characteristics with high acidity levels and environmental variability. Accurate land suitability determination requires analysis of various variables such as soil pH, soil and air moisture, rainfall, and soil texture. This study aims to develop land suitability classification models using Decision Tree and Random Forest algorithms for chili and eggplant crops in the region. Environmental data and soil characteristics were analyzed using both methods to evaluate classification performance. The results show that the Random Forest algorithm outperforms with an accuracy of up to 99% in chili land classification, as well as higher precision and recall values compared to Decision Tree. For eggplant land classification, both algorithms demonstrated perfect performance with accuracy and evaluation metrics reaching 1.00 without any misclassification. The advantage of Random Forest lies in its ability to handle complex input variables and reduce the risk of overfitting through ensemble decision trees, resulting in more stable and reliable predictions. Therefore, Random Forest is highly suitable for use in land suitability classification systems based on environmental data in West Aceh, supporting more optimal and sustainable cultivation decision-making. This study makes an important contribution to the application of machine learning technology to improve agricultural efficiency and production outcomes in regions with challenging peat soil characteristics.

Keywords: Land Suitability; Random Forest; Decision Tree; Chili and Eggplant Cultivation

1. PENDAHULUAN

Kabupaten Aceh Barat merupakan salah satu wilayah di Provinsi Aceh yang memiliki potensi besar dalam sektor pertanian, terutama dalam budidaya tanaman hortikultura seperti cabai dan terong. Sebagian besar wilayahnya memiliki karakteristik tanah gambut yang unik, yaitu tingkat keasaman tinggi, kandungan bahan organik berlebih [1]. Kondisi ini seringkali menjadi tantangan utama dalam menentukan kesesuaian lahan untuk budidaya berbagai jenis tanaman secara optimal. Penentuan kesesuaian lahan seharusnya mempertimbangkan berbagai variabel lingkungan seperti pH tanah, kelembaban tanah dan udara, curah hujan, serta tekstur tanah [2], [3]. Variabel-variabel ini sangat berpengaruh terhadap keberhasilan budidaya, terutama di wilayah dengan kondisi edafik ekstrem seperti Aceh Barat. pH tanah yang terlalu asam dapat menghambat penyerapan unsur hara oleh tanaman, sementara kelembaban dan curah hujan yang tidak sesuai dapat memicu serangan penyakit atau gagal panen [4], [5]. Selain itu, tekstur tanah menentukan kapasitas retensi air dan aerasi, yang sangat krusial bagi pertumbuhan akar tanaman.

Dalam praktiknya, petani di Aceh Barat masih banyak mengandalkan pengalaman tradisional dan observasi manual dalam menentukan lokasi budidaya, tanpa dukungan analisis berbasis data yang sistematis. Hal ini berpotensi

menyebabkan kesalahan dalam pemilihan lahan, yang berdampak pada penurunan hasil panen, inefisiensi penggunaan lahan, serta pemborosan sumber daya [6]. Kabupaten Aceh Barat memiliki potensi lahan yang sangat layak untuk dikembangkan, khususnya di sektor pertanian dan perkebunan. Kondisi geografis wilayah yang cenderung datar dengan sedikit area berbukit menjadikannya prospektif sebagai lahan budidaya [7]. Oleh karena itu, penerapan teknik machine learning dalam menganalisis kesesuaian lahan menjadi solusi yang menjanjikan untuk mendukung pengambilan keputusan berbasis data, sehingga dapat meningkatkan efisiensi dan hasil produksi pertanian secara menyeluruh. Pendekatan ini memungkinkan identifikasi pola-pola tersembunyi dalam data lingkungan dan agronomi, yang seringkali luput dari pengamatan manual, serta dapat digunakan untuk membangun model prediksi yang adaptif terhadap perubahan kondisi lahan maupun iklim. Dengan demikian, teknologi ini diharapkan dapat menjadi fondasi dalam pengembangan sistem pertanian cerdas yang berkelanjutan di wilayah Aceh Barat.

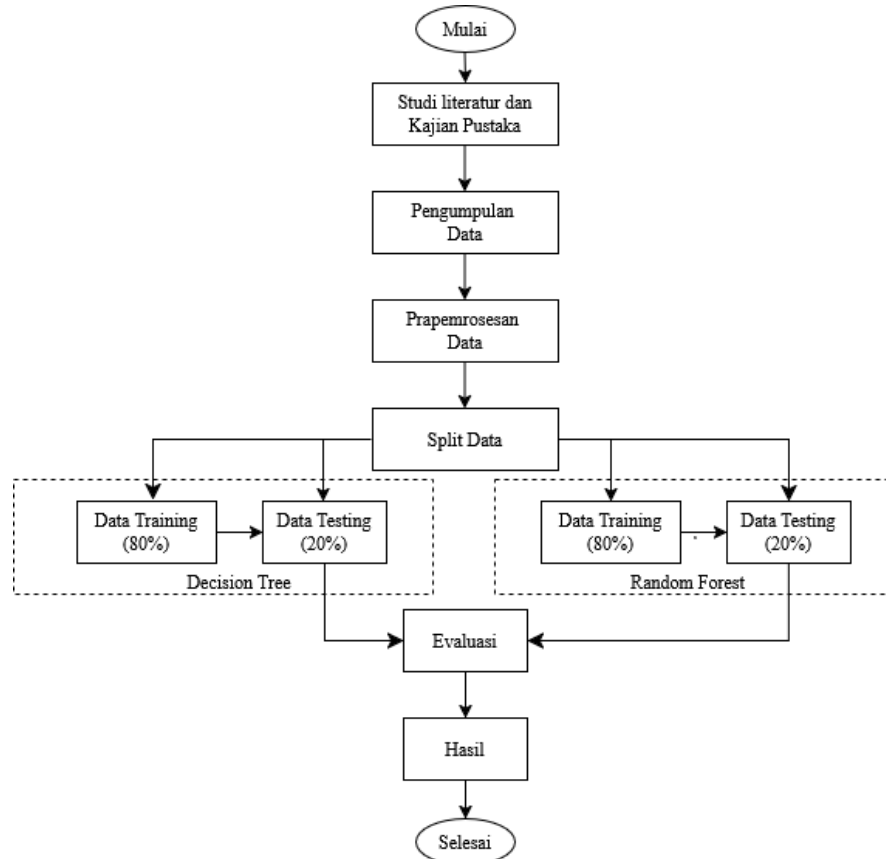
Seiring dengan perkembangan teknologi, pendekatan berbasis machine learning telah banyak diterapkan dalam bidang pertanian untuk membantu proses pengambilan keputusan secara lebih objektif dan akurat. Dua algoritma yang cukup populer dalam klasifikasi lahan pertanian adalah Decision Tree dan Random Forest. Penelitian sebelumnya yang dilakukan oleh Faiz membandingkan akurasi algoritma CART dan Random Forest dalam klasifikasi penggunaan lahan menggunakan citra Landsat 8 dan Sentinel-2A di platform *Google Earth Engine*. Hasilnya menunjukkan bahwa Random Forest memiliki akurasi lebih tinggi dibandingkan CART [8]. Selanjutnya penelitian Rodriguez-Galiano. Hasil penelitiannya menunjukkan bahwa algoritma RF menghasilkan klasifikasi tutupan lahan yang akurat, dengan akurasi keseluruhan 92% dan indeks Kappa sebesar 0,92 [9]. Selanjutnya penelitian Berhane, dkk. Hasil temuannya algoritma Random Forest (RF) menunjukkan kinerja yang lebih unggul dibandingkan Decision Tree (DT) dalam klasifikasi lahan basah, dengan akurasi keseluruhan (OA) lebih dari 81%. Meskipun kedua algoritma tergolong nonparametrik dan sama-sama layak digunakan, RF memberikan hasil klasifikasi yang lebih akurat dan stabil [10]. Selanjutnya penelitian Purwanto, dkk. Algoritma Random Forest (RF) secara umum menunjukkan kinerja yang lebih unggul dibandingkan Decision Tree (DT) dalam klasifikasi hutan mangrove. RF menghasilkan akurasi keseluruhan tertinggi sebesar 99,12% pada citra Landsat-7 ETM+, dibandingkan DT yang mencapai 92,82%. Namun, pada citra Landsat-8 OLI, DT sedikit lebih unggul dengan akurasi 98,34% dibandingkan RF sebesar 97,79% [11]. Penelitian lain juga mengatakan bahwa Random Forest efektif dalam meningkatkan akurasi prediksi kesesuaian lahan pertanian, karena kemampuannya dalam menangani data non-linear dan mengurangi risiko overfitting melalui ensemble dari pohon keputusan [12]. Selanjutnya Imam Khosravi. Penelitiannya membandingkan algoritma Decision Tree (DT), Random Forest (RF), Rotation Forest (RoF), Support Vector Machine (SVM), dan Dynamic Time Warping (DTW). Hasil menunjukkan bahwa algoritma DTW dan RF memberikan performa terbaik dengan akurasi sekitar 96%, serta mampu meningkatkan akurasi keseluruhan sebesar 8% dibandingkan metode lainnya [13].

Berdasarkan berbagai temuan penelitian sebelumnya, algoritma Random Forest (RF) dan Decision Tree menunjukkan performa yang sangat baik dalam klasifikasi lahan, sehingga sangat potensial untuk diterapkan dalam menentukan kesesuaian lahan untuk tanaman cabai dan terong di wilayah Aceh Barat. Wilayah ini memiliki karakteristik tanah gambut dengan keasaman tinggi, kelembaban yang bervariasi, serta curah hujan yang cukup tinggi, sehingga membutuhkan pendekatan klasifikasi yang mampu menangani data kompleks dan non-linear. Oleh karena itu, penelitian ini memiliki urgensi yang tinggi mengingat kebutuhan untuk memperoleh metode klasifikasi lahan yang akurat dan andal dalam menghadapi kompleksitas karakteristik tanah dan iklim di wilayah Aceh Barat. Implementasi algoritma Random Forest dan Decision tree berpotensi memberikan kontribusi signifikan dalam peningkatan presisi penentuan kesesuaian lahan untuk tanaman cabai dan terong, sehingga mendukung pengelolaan sumber daya lahan secara berkelanjutan dan optimal. Oleh karena itu, penelitian ini penting untuk dilakukan guna mengembangkan solusi berbasis data yang efektif dalam mendukung sektor pertanian di daerah tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

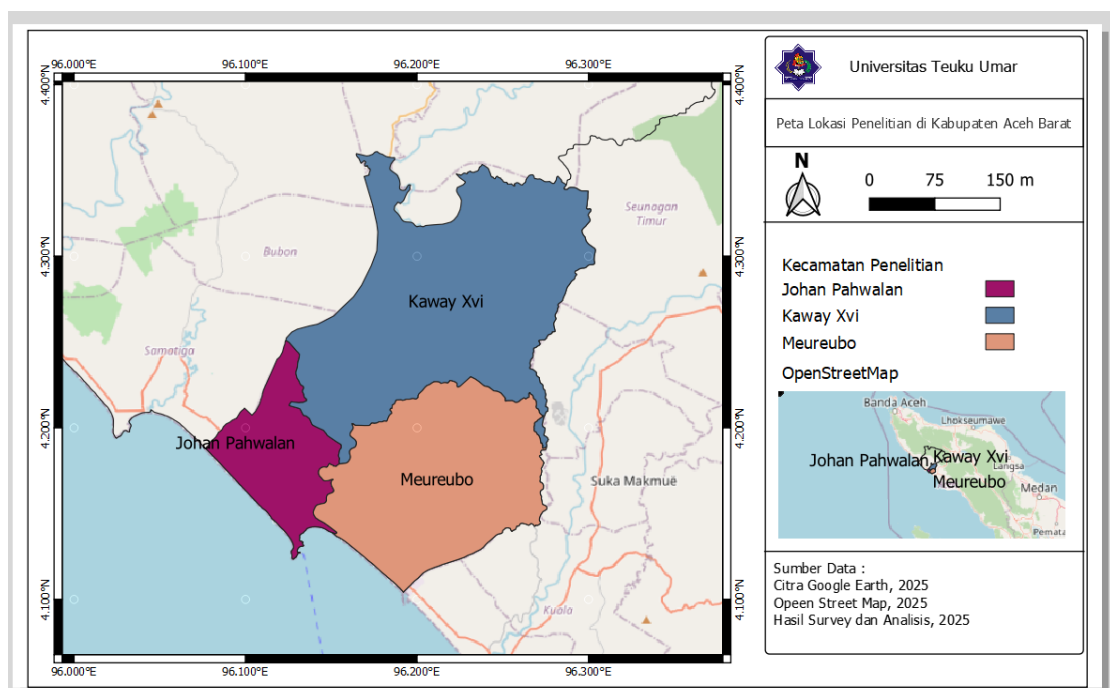
Tahapan penelitian merupakan serangkaian langkah sistematis yang dirancang untuk menggambarkan alur pelaksanaan penelitian dari tahap perencanaan awal hingga penyusunan kesimpulan akhir, guna memastikan keteraturan dan ketepatan dalam pencapaian tujuan penelitian. Adapun tahapan-tahapan yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

Berdasarkan tahapan penelitian pada gambar 1, dapat dijelaskan masing-masing tahapan sebagai berikut:

- Studi literatur dan kajian pustaka: bagian ini mengidentifikasi permasalahan terkait keseusian lahan untuk tanaman cabai dan terong. Langkah ini dilakukan melalui survei lokasi, wawancara petani pada tiga kecamatan di Aceh Barat yaitu Johan Pahlawan, Kaway XVI, dan Meureubo (Gambar 2). Lokasi tersebut dipilih karena merupakan wilayah potensial pertanian di Aceh Barat dengan ketersediaan lahan yang masih luas namun belum dimanfaatkan secara optimal. Kemudian kajian pustaka dengan cara pengumpulan referensi pendukung dari jurnal dan buku ilmiah supaya hasil penelitian ini mempunyai dasar yang kuat dan dapat dipertanggungjawabkan.



Gambar 2. Lokasi Penelitian Kecamatan Johan Pahlawan, Kaway XVI, dan Meureubo

- b. Pengumpulan Data: pengumpulan data dalam penelitian ini dari berbagai sumber, antara lain data tanah dari ISRIC-World Soil Information, data tutupan lahan dari WWF/USGS, data elevasi dari SRTM, serta data kategori tanaman cabai dan terong yang diperoleh dari Repositori Kementerian Pertanian.
- c. Prapemrosesan Data: data dalam penelitian ada yang kategorik dan ada yang numerik, maka perlu dilakukan prapemrosesan data sebelum dilakukan untuk penerapan model klasifikasi. Praproses dilakukan dengan mengubah data kategorik menjadi bentuk numerik melalui proses One Hot Encoding. One hot encoding merupakan metode untuk mengonversi data kategorikal menjadi representasi numerik yang dapat diproses oleh algoritma pembelajaran mesin [14], [15]. One hot encoding berguna untuk menyederhanakan pengolahan dan analisis data oleh algoritma machine learning dalam penelitian ini untuk kasus klasifikasi. Teknik ini juga membantu meningkatkan kualitas data serta memastikan kompatibilitas data bersih, relevan, dan sesuai format.
- d. Split Data: Data dibagi menjadi dua bagian yaitu data latih sebesar 80% dan data uji sebesar 20% untuk masing-masing target klasifikasi, yaitu cabai dan terong.
- e. Penerapan Model Klasifikasi: proses klasifikasi kesesuaian lahan cabai dan terong menggunakan dua algoritma, yaitu Decision Tree C4.5 dan Random Forest.

1. Decision Tree C4.5

Algoritma Decision Tree C4.5 penyempurnaan dari algoritma ID3 dan juga dikembangkan oleh J. R. Quinlan [16]. Algoritma C4.5 menggunakan rasio gain informasi sebagai metrik untuk membagi pohon keputusan. Algoritma ini secara rekursif mengklasifikasikan data hingga data tersebut dapat dikategorikan seakurat mungkin [17], [18], [19]. Algoritma C4.5 secara rekursif menggunakan konsep entropi untuk memilih atribut terbaik hingga tidak ada lagi cabang yang dapat dibentuk [20]. Algoritma C4.5 membangun pohon keputusan dengan memilih atribut sebagai akar, membuat cabang untuk tiap nilai, lalu membagi kasus ke cabang tersebut, dan mengulangi proses hingga semua kasus memiliki kelas yang sama [21], [22]. Menghitung gain dengan menggunakan persamaan berikut:

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

Keterangan:

S: Himpunan kasus

A: Atribut

n: jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke-i

$|S|$: jumlah kasus dalam S

Sementara itu perhitungan nilai entropi dapat dilihat pada persamaan berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (2)$$

Keterangan:

S: himpunan kasus

n: jumlah partisi S

p_i : proporsi dari S_i terhadap S

Kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma C4.5 adalah gain ratio, yang diformulasikan oleh persamaan berikut:

$$Gain\ Ratio(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)} \quad (3)$$

Keterangan:

S: dataset

j: atribut (fitur)

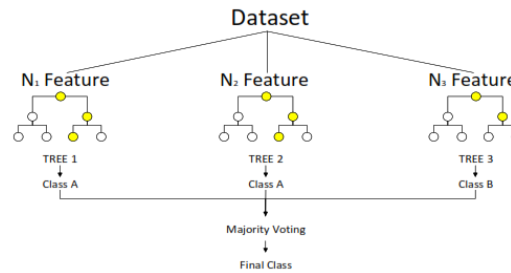
2. Random Forest

Random Forest merupakan algoritma pembelajaran ensemble yang menggabungkan sejumlah pohon keputusan (decision tree) yang bekerja secara bersama-sama untuk meningkatkan akurasi dalam tugas klasifikasi maupun regresi [23], [24]. Selain itu, random forest juga termasuk kedalam model pembelajaran supervised learning yang secara luas diterapkan dalam berbagai disiplin ilmu untuk keperluan klasifikasi [25], [26], [27]. Menghitung entropi dan gini dengan menggunakan rumus berikut:

$$Entropy = \sum_i^c p_i * \log_2 p_i \quad (4)$$

$$Gini = 1 - \sum_i^c p_i^2 \quad (5)$$

Deteksi dilakukan secara bertahap melalui struktur pohon keputusan yang bercabang ke bawah, seperti ditampilkan pada Gambar 3.



Gambar 3. Split Random Forest

- f. Evaluasi: evaluasi dilakukan dengan menggunakan classification report dan confusion matrix untuk menilai performa masing-masing algoritma. Akhirnya, dilakukan analisis terhadap hasil evaluasi guna mengetahui efektivitas model dalam mengklasifikasikan data tanaman cabai dan terong.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset dalam penelitian ini berjumlah 360 data, berupa data historis bulanan yang tercatat dari tahun 1995 hingga 2024. Data yang digunakan sebagai input ini terdiri dari curah hujan, suhu, penyinaran matahari, kelembapan udara, kelembapan tanah, diperoleh dari BMKG Cut Nyak Dhien, Kabupaten Nagan Raya. Selanjutnya Ph tanah, tekstur tanah, elevasi, drainase diperoleh dari website ISRIC-World Soil Information, WWF/USGS, dan SRTM. Sedangkan data yang dijadikan sebagai target yaitu kategori kesesuaian lahan untuk cabai dan terong diperoleh dari Repositori Kementerian Pertanian. Sampel data yang digunakan pada penelitian ini, dapat dilihat pada tabel 1 berikut.

Tabel 1. Dataset Penelitian

No	Thn	Blm	pHT	KT	KU	PM	TT	CH	S	E	D	C	T
1	1995	Jan	4.8	41.8	82	5	Liat	306	27.7	24	Baik	S	TS
2	1995	Feb	5.6	29.9	91	5.5	Pasir	367	24.9	31	Sedang	TS	S
3	1995	Mar	5.1	26.3	75.3	6	Pasir	330	30	38	Sedang	TS	S
4	1995	Apr	7.5	38	85.7	6.5	Lempung	497	27.1	45	Sedang	S	TS
5	1995	Mei	4.8	16.6	91.3	6	Pasir	223	24.3	11	Sedang	TS	TS
...	...	...	...	...	...	...	...	...	...	...	...	...	...
356	2024	Agu	5.4	42.7	71.5	5	Liat	127	30.6	49	Baik	TS	S
357	2024	Sep	4.6	41.5	81.9	5.5	Liat	421	28.1	15	Sedang	S	TS
358	2024	Okt	6	50.6	75.2	6	Liat	475	30.5	22	Buruk	TS	TS
359	2024	Nov	6	26.1	73.7	5.5	Pasir	201	30.1	29	Baik	TS	S
360	2024	Des	5.4	34.3	77.7	5	Lempung	357	29.3	36	Baik	TS	S

Keterangan:

Thn: Tahun, pHT: pH Tanah, KT: Kelembapan Tanah, KU: Kelembapan Udara, PM: Penyinaran Matahari, TT: Tekstur Tanah, CH: Curah Hujan, S: Suhu, E: Elevasi, D: Drainase, C: Cabai, T: Terong, S: Sesuai, TS: Tidak Sesuai.

Tabel 1. menyajikan sampel data iklim dan karakteristik tanah yang digunakan dalam penelitian untuk menentukan kesesuaian lahan terhadap komoditas cabai dan terong. Dataset terdiri atas data historis bulanan dari tahun 1995 hingga 2024, dengan total sebanyak 360 data. Sebelum digunakan dalam proses pemodelan klasifikasi, dilakukan tahapan praproses data, salah satunya dengan mengeliminasi kolom No, Thn, dan Blm karena tidak memberikan kontribusi langsung terhadap prediksi model. Selanjutnya, dilakukan proses encoding untuk mengubah data kategorikal (teks) menjadi format numerik agar dapat diterima oleh algoritma pembelajaran mesin.

3.2 Prapemrosesan Data

Sebelum membangun model klasifikasi, dilakukan prapemrosesan data guna memastikan bahwa data yang digunakan bersih, relevan, dan sesuai format. Data dalam penelitian ini berupa numerik dan kategorikal maka perlu dilakukan Encoding pada data kateirik khususnya pada data tekstur tanah, drainase, cabai, dan terong. Encoding adalah proses mengubah data kategorikal menjadi format numerik agar dapat diproses oleh komputer dalam analisis data dan machine learning [28]. Hasil dari encoding tersebut disajikan dalam Tabel 2 berikut.

Tabel 2. Hasil Encoding Data

Data ke-	pHT	KT	KU	PM	TT	CH	S	E	D	C	T
1	4.8	41.8	82	5	1	306	27.7	24	0	1	0
2	5.6	29.9	91	5.5	2	367	24.9	31	2	0	1
3	5.1	26.3	75.3	6	2	330	30	38	2	0	1

4	7.5	38	85.7	6.5	0	497	27.1	45	2	1	0
5	4.8	16.6	91.3	6	2	223	24.3	11	2	0	0
...	...	...	...	...	...	...	...	...	...	...	...
356	5.4	42.7	71.5	5	1	127	30.6	49	0	0	1
357	4.6	41.5	81.9	5.5	1	421	28.1	15	2	1	0
358	6	50.6	75.2	6	1	475	30.5	22	1	0	0
359	6	26.1	73.7	5.5	2	201	30.1	29	0	0	1
360	5.4	34.3	77.7	5	0	357	29.3	36	0	0	1

Tabel 2. Menjelaskan kolom yang tidak relevan untuk pemodelan klasifikasi seperti No, Thn, dan Bln telah dihilangkan sehingga hanya menyisakan atribut-atribut numerik dan target klasifikasi. Seluruh variabel kategorikal seperti tekstur tanah (TT), drainase (D), serta variabel target yaitu kesesuaian lahan untuk cabai (C) dan terong (T) telah dikonversi ke dalam format numerik menggunakan teknik label encoding. Sebagai contoh, nilai tekstur tanah seperti Liat, Pasir, dan Lempung masing-masing dikodekan sebagai 0, 1, dan 2, sementara kategori drainase (Baik, Sedang, Buruk) dan kelas kesesuaian lahan (Sesuai = 1, Tidak Sesuai = 0) juga diubah ke dalam representasi numerik. Hasil encoding ini menghasilkan dataset yang sepenuhnya numerik dan siap digunakan sebagai input dalam algoritma klasifikasi berbasis machine learning. Proses ini penting untuk memastikan bahwa model dapat menginterpretasikan seluruh atribut secara matematis dalam proses pelatihan dan prediksi.

3.3 Model Klasifikasi

Algoritma yang digunakan untuk klasifikasi kesesuaian lahan cabai dan terong yaitu Decision Tree dan Random Forest, karena kapabilitasnya dalam menangani data berdimensi tinggi dan kompleks. Decision Tree membangun satu struktur pohon keputusan yang membagi data berdasarkan fitur-fitur paling informatif. Pengaturan hyperparameter seperti kedalaman maksimum dan jumlah minimum sampel pada pemisahan digunakan untuk mengontrol kompleksitas model dan mencegah overfitting. Sementara Random Forest mengadopsi pendekatan ensemble dengan menggabungkan sejumlah pohon keputusan guna meningkatkan akurasi klasifikasi serta mengurangi potensi overfitting.

3.1.1 Decision Tree

Klasifikasi dengan model Decision Tree mula-mula dataset hasil encoding pada tabel 2 dibagi menjadi data latih dan data uji menggunakan fungsi `train_test_split` dari pustaka Scikit-Learn. Ilustrasi kode Python. Proses pembagian data latih dan data uji ditampilkan pada Gambar 4 berikut.

```
# === 4. Split Data ===
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 4. Kode Split Data

Kode pada Gambar 4. di atas berfungsi untuk membagi dataset menjadi dua bagian, yaitu data latih (training set) dan data uji (testing set), dengan menggunakan fungsi `train_test_split` dari pustaka Scikit-Learn. Variabel `X_train` (pHT, KT, KU, PM, TT, CH, S, E, D) dan `y_train` (C, T). Sedangkan `X_test`, dan `y_test` masing-masing merepresentasikan fitur dan label dari dataset itu sendiri, hanya saja membedakan parameter `test_size=0.2` menunjukkan bahwa 20% dari data akan digunakan sebagai data uji, sedangkan 80% sisanya digunakan untuk pelatihan model. Penggunaan `random_state=42` bertujuan untuk memastikan bahwa pembagian data dilakukan secara konsisten setiap kali program dijalankan, sehingga hasil eksperimen dapat direproduksi. Selanjutnya tahap membuat model decision tree, evaluasi dan pohon keputusan dapat diperlihatkan coding python pada Gambar 5 berikut.

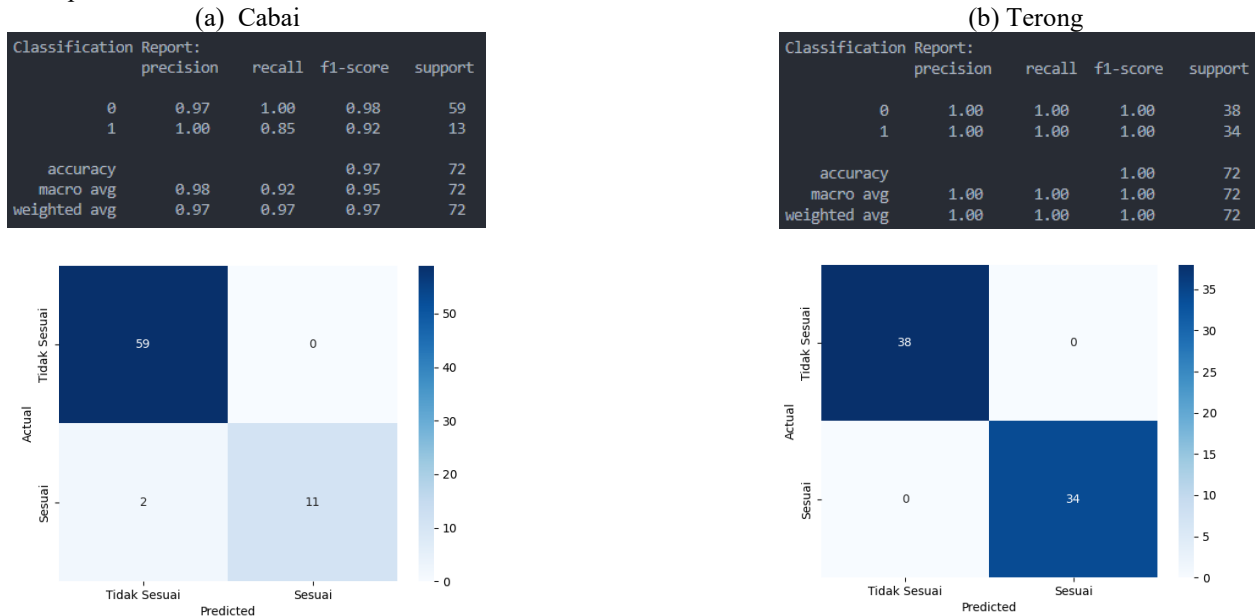
```
# === 5. Buat Model Decision Tree ===
model = DecisionTreeClassifier(criterion='entropy', max_depth=5, random_state=42)
model.fit(X_train, y_train)

# === 6. Evaluasi ===
y_pred = model.predict(X_test)
print("Akurasi:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))

# === 7. Confusion Matrix ===
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=['Tidak Sesuai', 'Sesuai'],
            yticklabels=['Tidak Sesuai', 'Sesuai'])
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```

Gambar 5. Koding Model Decision Tree

Kode pada Gambar 5. model dikembangkan dengan kriteria pemisahan entropy dan kedalaman maksimum 5. Evaluasi model dilakukan menggunakan metrik akurasi dan classification report yang mencakup presisi, recall, dan F1-score. Hasil klasifikasi divisualisasikan dalam bentuk confusion matrix guna melihat performa prediksi terhadap kelas Sesuai (1) dan Tidak Sesuai (0) untuk lahan cabai dan terong. Hasil klasifikasi kesesuaian lahan cabai dan terong dapat dilihat pada Gambar 6 berikut.



Gambar 6. Hasil Klasifikasi Model Decision Tree

Gambar 6. menjelaskan bahwa model klasifikasi lahan untuk cabai (a) menunjukkan akurasi tinggi (97%) dengan presisi kelas Sesuai mencapai 1.00, namun recall-nya masih 0.85, menandakan beberapa lahan Sesuai salah diklasifikasikan. F1-score sebesar 0.92 tetap mencerminkan kinerja baik, meski recall dapat ditingkatkan. Sementara itu, model untuk terong (b) menunjukkan performa sempurna dengan akurasi, presisi, recall, dan F1-score semuanya 1.00. Kedua model dinilai andal, sehingga lahan yang diprediksi Sesuai direkomendasikan untuk budidaya cabai dan terong.

3.1.1 Random Forest

Random Forest merupakan algoritma ensemble learning yang menggabungkan sejumlah pohon keputusan (decision trees) untuk bekerja secara bersama-sama guna meningkatkan akurasi dalam tugas klasifikasi maupun regresi [23]. Pengembangan model Random Forest dalam hal pembagian data latih dan data uji sama dengan decision tree yang dimana data dibagi 80% untuk pelatihan dan 20% digunakan untuk data uji. Secara lengkap coding untuk pengembangan model Random Forest dapat dilihat pada Gambar 7 berikut.

```
# === 4. Split data ===
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

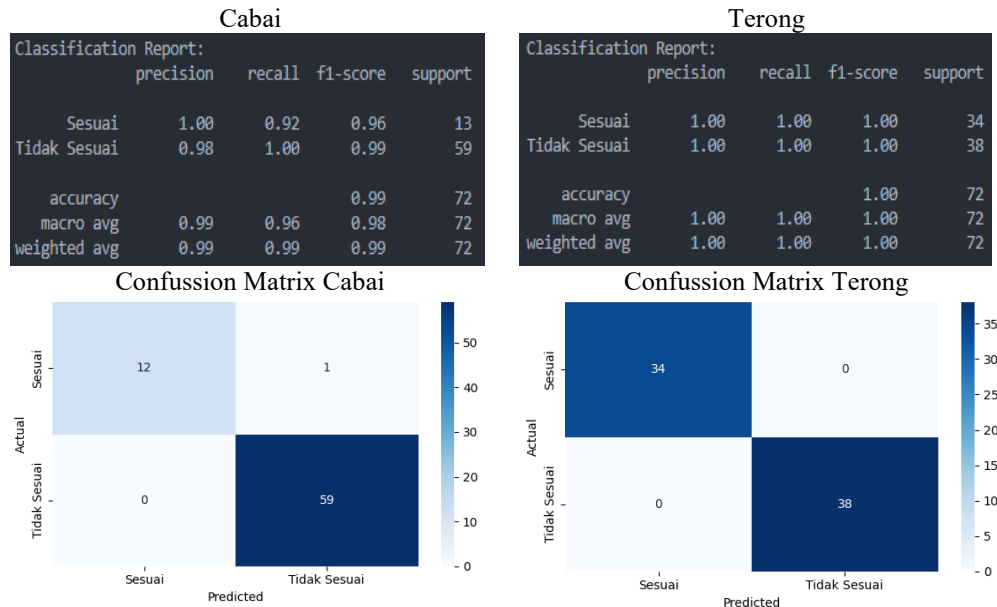
# === 5. Train Random Forest ===
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# === 6. Prediksi & Evaluasi ===
y_pred = model.predict(X_test)
print("Akurasi:", accuracy_score(y_test, y_pred))
print("\nClassification Report:")
print(classification_report(y_test, y_pred, target_names=label_encoders['Cabai'].classes_))

# === 7. Confusion Matrix ===
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=label_encoders['Cabai'].classes_,
            yticklabels=label_encoders['Cabai'].classes_)
plt.title("Confusion Matrix")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.tight_layout()
plt.show()
```

Gambar 7. Koding Model Random Forest

Kode pada Gambar 7 di atas merupakan model klasifikasi menggunakan algoritma Random Forest. Data terlebih dahulu dibagi menjadi data latih dan data uji dengan perbandingan 80:20 menggunakan fungsi `train_test_split`. Model Random Forest kemudian dilatih dengan data latih dan digunakan untuk memprediksi data uji. Hasil prediksi dievaluasi menggunakan metrik akurasi dan laporan klasifikasi yang menampilkan nilai `precision`, `recall`, dan `f1-score` untuk cabai dan terong. Selanjutnya, performa model divisualisasikan melalui confusion matrix menggunakan heatmap dari library Seaborn, dengan label kelas diperoleh dari objek `label_encoders` untuk interpretasi yang lebih mudah. Hasil model Random Forest dapat dilihat pada Gambar 8 berikut.



Gambar 8. Hasil Klasifikasi Model Random Forest

Pada Gambar 8 klasifikasi data cabai, model klasifikasi cabai menunjukkan akurasi tinggi (99%) dengan presisi sempurna untuk kelas Sesuai (1.00), namun `recall`-nya masih 0.92. Terdapat sedikit kesalahan dalam mendeteksi lahan Sesuai. Sebaliknya, model untuk terong menunjukkan performa sempurna (semua metrik 1.00) tanpa kesalahan klasifikasi. Model terong siap diterapkan, sementara model cabai hanya memerlukan sedikit perbaikan untuk meningkatkan `recall`. Keduanya sangat potensial untuk digunakan dalam sistem klasifikasi lahan.

3.4 Pembahasan

Berdasarkan hasil evaluasi performa model klasifikasi kesesuaian lahan untuk budidaya cabai dan terong, algoritma Random Forest terbukti sebagai pilihan yang paling tepat dan unggul dibandingkan Decision Tree. Pada klasifikasi lahan cabai, Random Forest mampu mencapai akurasi sebesar 99%, lebih tinggi dibandingkan Decision Tree yang memperoleh 97%. Selain itu, Random Forest menunjukkan nilai `precision` sebesar 1.00 dan `recall` sebesar 0.92 untuk kelas Sesuai, sedangkan Decision Tree hanya mencapai `recall` 0.85, yang menunjukkan bahwa Random Forest lebih andal dalam mengenali lahan yang benar-benar sesuai untuk budidaya cabai. Nilai F1-score yang lebih tinggi pada Random Forest (0.96 dibandingkan 0.92 pada Decision Tree) mengindikasikan keseimbangan yang lebih baik antara presisi dan kemampuan model dalam menangkap seluruh data positif. Jumlah kesalahan klasifikasi juga lebih sedikit pada Random Forest, yaitu hanya satu data Sesuai yang salah diklasifikasikan sebagai Tidak Sesuai, dibandingkan dua kesalahan pada Decision Tree. Untuk klasifikasi lahan terong, kedua algoritma menunjukkan performa sempurna, dengan akurasi dan seluruh metrik evaluasi mencapai nilai maksimal (1.00), serta tanpa kesalahan klasifikasi berdasarkan confusion matrix. Meskipun demikian, Random Forest tetap lebih direkomendasikan karena merupakan algoritma ensemble yang menggabungkan banyak pohon keputusan, sehingga lebih tahan terhadap overfitting, lebih stabil ketika digunakan pada data baru, serta memiliki kemampuan generalisasi yang lebih baik dibandingkan dengan Decision Tree tunggal. Keunggulan performa Random Forest dalam mengklasifikasikan kesesuaian lahan ini sangat erat kaitannya dengan kemampuannya dalam menangani variabel input yang kompleks, seperti pH tanah, kelembaban tanah, kelembaban udara, penyinaran matahari, tekstur tanah, curah hujan, suhu, elevasi, dan drainase, yang semuanya merupakan faktor penting dalam menentukan kecocokan lahan untuk budidaya tanaman. Random Forest mampu mengenali pola dan interaksi antar variabel tersebut secara lebih efektif melalui proses pembelajaran dari banyak pohon keputusan, sehingga hasil prediksi yang dihasilkan menjadi lebih akurat dan dapat diandalkan. Oleh karena itu, secara keseluruhan, Random Forest dinilai sangat cocok dan sangat baik untuk digunakan dalam sistem klasifikasi kesesuaian lahan berbasis data lingkungan, baik untuk tanaman cabai maupun terong, serta layak dijadikan dasar dalam pengambilan keputusan budidaya di lapangan.

4. KESIMPULAN

Berdasarkan hasil penelitian dan evaluasi yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest terbukti lebih unggul dibandingkan Decision Tree dalam klasifikasi kesesuaian lahan untuk budidaya cabai dan terong. Pada lahan cabai, Random Forest mencapai akurasi 99%, mengungguli Decision Tree yang hanya mencapai akurasi 97% untuk kelas Sesuai. Meskipun pada klasifikasi lahan terong kedua algoritma menunjukkan performa sempurna dengan akurasi dan seluruh metrik evaluasi bernilai 1.00, Random Forest tetap direkomendasikan karena sebagai algoritma ensemble, ia lebih stabil dan tahan terhadap overfitting. Keunggulan ini berkaitan dengan kemampuannya dalam mengenali pola dan interaksi kompleks antar variabel lingkungan seperti pH tanah, kelembaban tanah dan udara, penyinaran, tekstur tanah, curah hujan, suhu, elevasi, dan drainase. Dengan memanfaatkan banyak pohon keputusan, Random Forest dapat menangkap hubungan non-linear antar variabel secara lebih akurat dan andal dibandingkan model pohon tunggal, sehingga sangat layak digunakan sebagai dasar sistem pendukung keputusan berbasis data lingkungan untuk meningkatkan efektivitas dan keberhasilan budidaya tanaman cabai dan terong.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

- a. Lembaga Pengelola Dana Pendidikan (LPDP) atas dukungan pendanaan melalui Program Pendanaan Riset Pembangunan Berkelanjutan (PRPB) sehingga riset Stemini ini dapat dilaksanakan.
- b. Universitas Teuku Umar, khususnya kepada Rektor, Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM), Program Studi Teknik Sipil dan Pertanian Universitas Teuku Umar (UTU), serta UTU Farm atas dukungan fasilitas yang telah diberikan.
- c. Laboratorium Tanah Universitas Syiah Kuala atas kontribusi teknis yang telah diberikan selama pelaksanaan riset.
- d. Kelompok Tani Sama Rasa, Desa Mesjid Tuha, Kecamatan Meureubo, Kabupaten Aceh Barat, atas kerja sama dan dukungan yang diberikan selama pelaksanaan riset ini.

REFERENCES

- [1] S. Jailani, R. Ratnawaty, N. Nasruddin, F. Faisal, and I. Ismadi, "Respon Tanaman Terung (*Solanum melongena* L.) Pada Berbagai Media Tanaman dan Dosis Pupuk NPK," *J. Agrum*, vol. 16, no. 2, p. 151, 2019, doi: 10.29103/agrium.v16i2.5867.
- [2] R. Wirosoedarmo *et al.*, "101338-None-E82914Wirosoedarmo, R., Sutanahaji, A. T., Kurniati, E., Gdwd, V., Ehuksd, D., Noldpwrjrl, G., Dk, Z., Kdvlo, G. D. Q., Odsdjqdq, S., & Vlidw, W. (2011). 101338-None-E82914Ba. 31(1).Ba," vol. 31, no. 1, 2011.
- [3] P. D. Wijaya, "Evaluasi Kesesuaian Lahan untuk Pengembangan Budidaya Udang Windu (*Penaeus Monodon*) Sistem Silvofishery di Kecamatan Seruway," vol. 13, no. 2, pp. 111–119, 2024.
- [4] F. Fatkurahman, M. A. Azis, Y. Rahim, R. Rahman, and S. Apriliani, "Evaluasi Kesesuaian Lahan untuk Tanaman Jagung (*Zea mays* L.) di Kecamatan Taluditi, Kabupaten Pohuwato," *J. Lahan Pertan. Trop.*, vol. 2, no. 1, pp. 72–80, 2023, doi: 10.56722/jlpt.v2i1.20601.
- [5] U. Hasanah, Khusrizal, Muliana, H. Akbar, and Yusra, "Determinasi Kesesuaian Lahan Tanaman Padi Sawah Irigasi Pendahuluan Metode Penelitian," *J. Ilm. Mhs. Agroekoteknologi*, vol. 1, no. 4, pp. 81–86, 2022.
- [6] S. Nurwahdani, R. S. Handayani, L. Nazirah, and A. Inayatillah, "Perbandingan pertumbuhan dan hasil berbagai varietas bawang merah (," vol. 11, pp. 68–75, 2024.
- [7] A. Amir *et al.*, "Mapping Agricultural Land Use in Aceh Barat By Using Gis," *Communnity Dev. J.*, vol. 5, no. 2, pp. 3302–3307, 2024.
- [8] F. Al Farikhi and R. W. D. Pramono, "Perbandingan algoritma classification and regression tree (cart) dan random forest (rf) untuk klasifikasi penggunaan lahan pada google earth engine," *J. Spat. Wahana Komun. dan Inf. Geogr.*, vol. 23, no. 2, pp. 170–179, 2023, doi: 10.21009/spatial.232.09.
- [9] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, no. 1, pp. 93–104, Jan. 2012, doi: 10.1016/J.ISPRSJP.2011.11.002.
- [10] T. M. Berhane *et al.*, "Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory," *Remote Sens.*, vol. 10, no. 4, 2018, doi: 10.3390/rs10040580.
- [11] A. D. Purwanto, K. Wikantika, A. Deliar, and S. Darmawan, "Decision Tree and Random Forest Classification Algorithms for Mangrove Forest Mapping in Sembilang National Park, Indonesia," *Remote Sens.*, vol. 15, no. 1, 2023, doi: 10.3390/rs15010016.
- [12] B. A. Ganati, "Predicting Land Suitability for Wheat and Barley Crops using Machine Learning Techniques," pp. 1–19, 2025.
- [13] I. Khosravi, "Towards sustainable agriculture in Iran using a machine learning-driven crop mapping framework," *Eur. J. Remote Sens.*, vol. 58, no. 1, Dec. 2025, doi: 10.1080/22797254.2025.2490787.
- [14] A. F. Azmi, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix," *Komputa J. Ilm. Komput. dan Inform.*, vol. 13, no. 1, pp. 111–119, 2024, [Online]. Available: <https://ojs.unikom.ac.id/index.php/komputa/article/view/12639>
- [15] H. L. Smith, P. J. Biggs, N. P. French, A. N. H. Smith, and J. C. Marshall, "Lost in the Forest: Encoding categorical variables and the absent levels problem," *Data Min. Knowl. Discov.*, vol. 38, no. 4, pp. 1889–1908, 2024, doi: 10.1007/s10618-024-01019-w.
- [16] J. Quinlan, "C4.5 Programs for Machine Learning," Morgan Kaufmann Publishers, United States. Accessed: May 24, 2025. [Online]. Available: <https://www.scirp.org/reference/referencespapers?referenceid=2554378>
- [17] S. Bishnoi and B. K. Hooda, "Decision Tree Algorithms and their Applicability in Agriculture for Classification," *J. Exp. Agric.*

- Int.*, vol. 44, no. 7, pp. 20–27, 2022, doi: 10.9734/jeai/2022/v44i730833.
- [18] F. Fadhila and P. S. Hasugian, “Application of C4.5 Algorithm to Prediction Sales at PT. Sumber Sayur Segar,” *J. Intell. Decis. Support Syst.*, vol. 5, no. 1, pp. 10–19, 2022, doi: 10.35335/idss.v5i1.45.
 - [19] D. Istiawan and L. Khikmah, “Implementation of C4.5 Algorithm for Critical Land Prediction in Agricultural Cultivation Areas in Pemali Jratun Watershed,” *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 2, p. 67, 2019, doi: 10.24014/ijaidm.v2i2.7569.
 - [20] I. Rahmayuni, “Perbandingan Performansi Algoritma C4.5 dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang,” *Teknoif*, vol. 2, no. 1, pp. 40–46, 2014.
 - [21] Kusriani and Luthfi E.T, *Algoritma Data Mining*. Yogyakarta: CV Andi Offset. Yogyakarta: CV Andi Offset, 2009.
 - [22] E. Prasetyo, *Data Mining Mengolah data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: ANDI. 2014.
 - [23] J. W. Yuda, H. Audytra, and N. Mahmudah, “Sistem Klasifikasi Tingkat Kerusakan Kunci Motor Menggunakan Random Forest dengan Hyperparameter Tuning,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 12, no. 2, pp. 84–94, 2025, doi: 10.30865/jurikom.v12i2.8517.
 - [24] N. G. Ramadhan, F. D. Adhinata, A. Jala, T. Segara, and D. Putra, “Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, pp. 251–256, 2022, doi: 10.30865/jurikom.v9i2.3979.
 - [25] N. G. Ramadhan and A. Romadhony, “Preprocessing Handling to Enhance Detection of Type 2 Diabetes Mellitus based on Random Forest,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, pp. 223–228, 2021.
 - [26] S. Pareeth, P. Karimi, M. Shafiei, and C. De Fraiture, “Mapping Agricultural Landuse Patterns from Time Series of Landsat 8 Using Random Forest Based Hierarchical Approach,” 2019, doi: 10.3390/rs11050601.
 - [27] A. Pang, M. W. L. Chang, and Y. Chen, “Evaluation of Random Forests (RF) for Regional and Local-Scale Wheat Yield Prediction in Southeast Australia,” pp. 1–19, 2022.
 - [28] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y. I. Cho, “Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms,” *Mathematics*, vol. 12, no. 16, 2024, doi: 10.3390/math12162553.