

Implementasi dan Evaluasi Algoritma Logika Fuzzy dan Random Forest untuk Sistem Rekomendasi Tanaman Berdasarkan Dataset Crop Recommendation



Disusun oleh:

SAMUEL YOHANES MICHAEL K.	103012300247
FADHILA M. WIRYASUKMA	103012300303

Program Studi S1 Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2025

BAB I

PENDAHULUAN

I.1. Latar Belakang

Pertanian memegang peranan strategis dalam mendukung ketahanan pangan dan perekonomian nasional. Keberhasilan budidaya tanaman sangat bergantung pada kesesuaian kondisi agroklimat, seperti kandungan nutrisi tanah (Nitrogen, Phosphorus, Kalium), suhu udara, kelembapan, pH tanah, dan curah hujan. Variasi faktor-faktor tersebut antar lokasi dan musim menyebabkan pemilihan jenis tanaman yang optimal menjadi tantangan tersendiri bagi petani. Kesalahan dalam memilih komoditas yang sesuai dapat menyebabkan penurunan hasil panen, kerugian ekonomi, serta ketidakpastian pasokan pangan nasional.

Dalam beberapa dekade terakhir, konsep precision agriculture (pertanian presisi) telah berkembang sebagai pendekatan modern yang memanfaatkan teknologi dan data untuk pengambilan keputusan agronomis. Penelitian-penelitian sebelumnya, seperti Crop Recommendation System for Precision Agriculture yang dilakukan oleh IEEE, menunjukkan bahwa pendekatan ini memungkinkan pemilihan tanaman yang sesuai berdasarkan data lingkungan seperti jenis tanah, kadar nutrisi, dan curah hujan melalui penerapan metode data mining seperti Random Tree, CHAID, K-Nearest Neighbor, dan Naive Bayes yang digabungkan melalui teknik ensemble majority voting. Sistem tersebut terbukti meningkatkan akurasi dalam rekomendasi tanaman sehingga membantu petani mengelola lahan dengan lebih efektif.

Selain itu, penelitian Crop Recommendation System oleh Ruchirawya T.H. menegaskan bahwa sistem rekomendasi tanaman dapat dibangun menggunakan integrasi perangkat keras dan kecerdasan buatan. Sistem ini dirancang dengan mengumpulkan data lingkungan menggunakan sensor dan mikrokontroler seperti Arduino, kemudian diproses menggunakan algoritma pembelajaran mesin seperti Naïve Bayes, Support Vector Machine (SVM), K-Means Clustering, dan Natural Language Processing (NLP) untuk analisis sentimen. Penelitian tersebut menekankan pentingnya deteksi kondisi lingkungan secara real-time guna merekomendasikan jenis tanaman yang paling sesuai untuk suatu lahan dengan tingkat akurasi dan efisiensi yang tinggi. Hal ini sangat relevan terutama dalam konteks keterbatasan lahan domestik dan ketidakpastian faktor iklim mikro yang berubah-ubah.

Kemajuan teknologi seperti sensor suhu dan kelembapan, kamera, drone, dan perangkat IoT lainnya telah merevolusi praktik pertanian modern dengan memungkinkan pemantauan kondisi tanaman dan lingkungan secara berkelanjutan. Data yang dihasilkan kemudian diolah menggunakan algoritma dan teknik statistik untuk mendukung pengambilan keputusan yang lebih tepat oleh petani, sehingga meningkatkan hasil panen melalui pemilihan tanaman yang lebih akurat berdasarkan kondisi spesifik lahan.

Namun, tantangan utama dalam pengolahan data agroklimat adalah sifatnya yang tidak pasti dan mengandung ambiguitas. Misalnya, nilai pH tanah yang berada di ambang batas antara asam dan netral, atau kelembapan yang dapat dikategorikan sebagai lembap maupun basah. Pendekatan logika konvensional yang bersifat biner (ya/tidak) sering kali tidak mampu menangani kondisi tersebut secara efektif.

Sebagai solusi, logika fuzzy hadir dengan kemampuannya mengakomodasi nilai-nilai “abu-abu” melalui pemodelan derajat kebenaran pada rentang $[0,1]$. Pendekatan utama dalam logika fuzzy—metode Mamdani—menawarkan mekanisme inferensi yang berbeda, di mana metode Mamdani menghasilkan output dalam bentuk himpunan fuzzy yang mudah diinterpretasikan, sedangkan metode Sugeno memberikan output berupa fungsi matematis atau nilai konstan yang lebih efisien secara komputasi.

Berdasarkan latar belakang tersebut, penelitian ini telah membangun Sistem Rekomendasi Tanaman berbasis logika fuzzy Mamdani dan Random Forest menggunakan *crop recommendation dataset* yang mencakup variabel N, P, K, suhu, kelembapan, pH, dan curah hujan. Implementasi dan evaluasi kedua metode tersebut telah dilakukan untuk menilai efektivitasnya dalam memberikan rekomendasi jenis tanaman yang paling sesuai. Hasil penelitian menunjukkan bahwa sistem berbasis logika fuzzy dapat memberikan rekomendasi yang lebih fleksibel dan akurat dibandingkan dengan metode konvensional, serta memberikan kontribusi penting terhadap pengembangan pertanian presisi yang adaptif terhadap perubahan lingkungan.

I.2. Rumusan Masalah

Berdasarkan persyaratan tugas besar dan latar belakang di atas, maka permasalahan penelitian ini dirumuskan sebagai berikut:

1. Bagaimana melakukan prapemrosesan data untuk memastikan kualitas *crop recommendation dataset*, meliputi pemilihan fitur, penanganan nilai hilang (missing value), dan normalisasi skala?
2. Bagaimana merancang dan mengimplementasikan sistem rekomendasi tanaman berbasis logika fuzzy dengan metode Mamdani yang memanfaatkan variabel agroklimat (N, P, K, suhu, kelembapan, pH, curah hujan)?
3. Bagaimana menetapkan fungsi keanggotaan (membership function) dan aturan linguistik (fuzzy rules) yang tepat untuk setiap variabel input dalam metode Mamdani?
4. Bagaimana mengevaluasi kinerja sistem fuzzy Mamdani dan Random Forest pada dataset studi kasus menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score?

5. Metode manakah—Fuzzy atau Random Forest—yang lebih efektif dan efisien dalam memberikan rekomendasi jenis tanaman optimal berdasarkan hasil evaluasi?

I.3. Tujuan

Berdasarkan rumusan masalah dan cakupan tugas besar ini, penelitian ini bertujuan untuk:

1. Melakukan prapemrosesan data pada *crop recommendation dataset*, meliputi seleksi fitur yang relevan, penanganan nilai hilang (missing value), dan normalisasi skala sehingga data siap digunakan dalam sistem fuzzy.
2. Merancang dan mengimplementasikan sistem rekomendasi tanaman berbasis logika fuzzy menggunakan metode Mamdani, dengan memanfaatkan variabel agroklimat (N, P, K, suhu, kelembapan, pH, curah hujan) sebagai masukan.
3. Menetapkan fungsi keanggotaan (membership function) dan menyusun aturan linguistik (fuzzy rules) yang sesuai untuk setiap variabel input dalam kedua metode fuzzy (Mamdani).
4. Mengevaluasi performa sistem fuzzy Mamdani dan Random Forest pada dataset studi kasus menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.
5. Menganalisis dan membandingkan efektivitas algoritma Fuzzy dan Random Forest dalam merekomendasikan jenis tanaman optimal berdasarkan hasil evaluasi, guna menentukan pendekatan yang paling efisien dan akurat.

BAB II

METODE

II.1. Paparan, statistik, dan sumber dari dataset yang digunakan

Penelitian ini menggunakan *crop recommendation dataset* yang diperoleh dari Kaggle (<https://www.kaggle.com/datasets/madhuraatmarambhagat/crop-recommendation-dataset>). Dataset tersebut terdiri atas tujuh variabel input, yaitu:

- **Nitrogen (N)**, kandungan Nitrogen dalam tanah (mg/kg)
- **Phosphorus (P)**, kandungan Phosphorus dalam tanah (mg/kg)
- **Potassium (K)**, kandungan Kalium dalam tanah (mg/kg)
- **Temperature**, suhu rata-rata udara (°C)
- **Humidity**, kelembaban relatif udara (%)
- **pH**, nilai keasaman tanah
- **Rainfall**, curah hujan (mm)

Dataset yang digunakan dalam penelitian ini terdiri dari 2.200 data dengan 22 label tanaman yang berbeda sebagai variabel output. Variabel output tersebut merepresentasikan jenis tanaman yang paling sesuai dengan kondisi agroklimat, seperti kandungan unsur hara (N, P, K), suhu, kelembapan, pH tanah, dan curah hujan. Contoh label tanaman yang terdapat dalam dataset ini antara lain padi (rice), jagung (maize), kapas (cotton), dan lain sebagainya. Namun, dalam penelitian ini, label target yang difokuskan adalah jagung (maize), kopi (coffee), dan kelapa (coconut), yang dipilih berdasarkan relevansi dan ketersediaan data.

Feature	Mean	Median	Std Dev	Min	Max	Skewness	Kurtosis
N	50.5518	37.0000	36.9173	0.0000	140.0000	0.5094	-1.0586
P	53.3627	51.0000	32.9859	5.0000	145.0000	1.0101	0.8556

K	48.1491	32.0000	50.6479	5.0000	205.0000	2.3735	4.4365
temperature	25.6162	25.5987	5.0637	8.8257	43.6755	0.1848	1.2270
humidity	71.4818	80.4731	22.2638	14.2580	99.9819	-1.0910	0.2987
ph	6.4695	6.4250	0.7739	3.5048	9.9351	0.2837	1.6491
rainfall	103.4637	94.8676	54.9584	20.2113	298.5601	0.9651	0.6030

Tabel diatas menyajikan hasil **statistik deskriptif** dari masing-masing fitur yang digunakan dalam sistem rekomendasi tanaman. Statistik ini mencakup nilai minimum, maksimum, rata-rata (mean), median,skewness, standar deviasi,dan kurtosis yang memberikan gambaran umum mengenai karakteristik distribusi data untuk setiap variabel input.

II.2. Pra-Pemrosesan Data

II.2.A. Mengatasi Missing Value

Kode berikut digunakan untuk memeriksa apakah terdapat nilai yang hilang (*missing value*) pada dataset:

```
print(df.isnull().sum())
```

🔥 Cek Missing Values:

```
N          0
P          0
K          0
temperature 0
humidity     0
ph           0
rainfall     0
label        0
```

Gambar diatas menunjukkan bahwa seluruh kolom dalam dataset—yaitu N, P, K, temperature, humidity, pH, rainfall, dan label—tidak mengandung nilai yang hilang. Hal ini ditunjukkan oleh nilai 0 pada setiap kolom, yang berarti data pada masing-masing atribut lengkap dan siap untuk digunakan dalam proses analisis lebih lanjut.

Kode berikut difokuskan pada kolom-kolom numerik untuk menangani data yang hilang (missing value) dengan cara mengisinya menggunakan nilai rata-rata (mean) dari masing-masing kolom.

```
numerik_cols = ['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall']
df[numerik_cols] = df[numerik_cols].fillna(df[numerik_cols].mean())
```

✓ Missing value diisi (mean substitution untuk numerik):

N	0
P	0
K	0
temperature	0
humidity	0
ph	0
rainfall	0
label	0

Gambar diatas menunjukkan bahwa seluruh kolom dalam dataset—meliputi N, P, K, temperature, humidity, pH, rainfall, dan label—tidak lagi mengandung nilai yang hilang setelah dilakukan proses imputasi menggunakan nilai rata-rata (mean) pada setiap kolom numerik.

II.2.B. Deteksi & Drop Outlier (IQR Method)

Untuk mendeteksi keberadaan outlier pada dataset, digunakan metode Interquartile ($1.5 \times IQR$), di mana Q1 adalah kuartil pertama dan Q3 adalah kuartil ketiga dari distribusi data. Proses ini dilakukan dengan iterasi terhadap setiap kolom numerik (N, P, K, temperature, humidity, ph, dan rainfall), lalu menghitung jumlah data yang memenuhi kriteria sebagai outlier.

```
for col in numerik_cols:
    Q1 = df[col].quantile(0.25)
```

```

Q3 = df[col].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

outlier_mask = (df[col] < lower_bound) | (df[col] > upper_bound)
banyak_outlier = outlier_mask.sum()
print(f" • {col}: {banyak_outlier} outlier ditemukan")

df = df[~outlier_mask]

df = df.reset_index(drop=True)

```

Berdasarkan hasil deteksi, ditemukan outlier sebanyak 138 data pada fitur P (phosphorus), 62 data pada fitur K (kalium), 58 data pada fitur temperature, 58 data pada fitur ph, dan 38 data pada fitur rainfall. Sementara itu, tidak ditemukan outlier pada fitur N dan humidity. Seluruh data yang teridentifikasi sebagai outlier dihapus dari dataset untuk menjaga kualitas dan integritas data yang akan digunakan pada tahap pemodelan.

Deteksi Outlier (IQR) pada fitur numerik:

- *N: 0 outlier ditemukan*
- *P: 138 outlier ditemukan*
- *K: 62 outlier ditemukan*
- *temperature: 58 outlier ditemukan*
- *humidity: 0 outlier ditemukan*
- *ph: 58 outlier ditemukan*
- *rainfall: 38 outlier ditemukan*

Data setelah drop outlier, shape sekarang: (1846, 8)

Setelah proses pembersihan data dari outlier, jumlah data mengalami pengurangan dari semula 2.200 menjadi 1.846 baris data, dengan tetap mempertahankan delapan atribut (tujuh fitur numerik dan satu label). Proses ini diharapkan dapat meningkatkan performa model dengan mengurangi potensi noise yang ditimbulkan oleh data ekstrem.

II.2.C. Mengatasi *skewness* pada fitur Rainfall

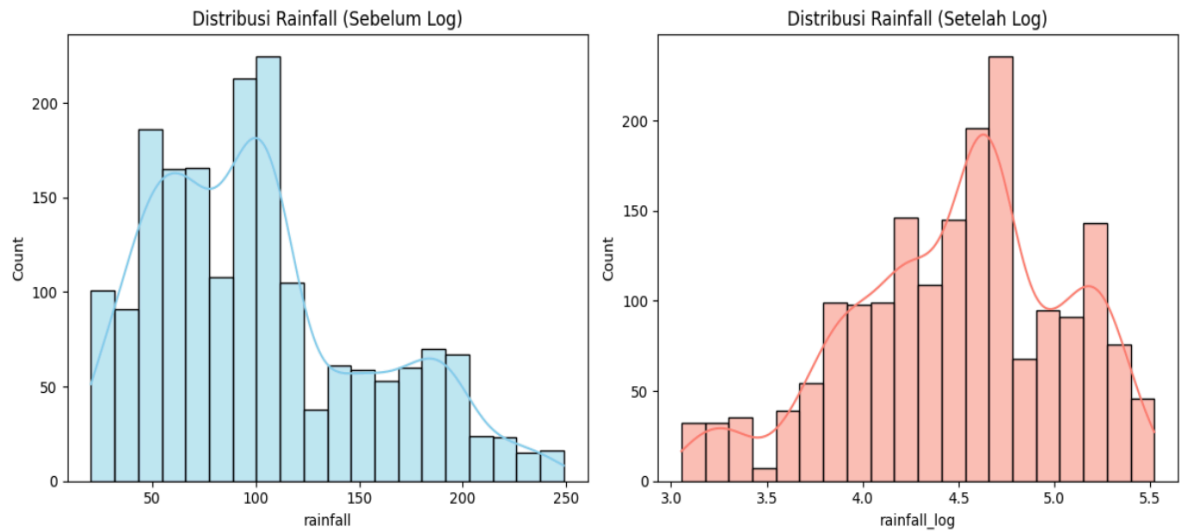
Untuk meningkatkan kualitas input pada sistem berbasis logika fuzzy, khususnya dalam tahap fuzzifikasi, dilakukan transformasi logaritmik terhadap fitur `rainfall`. Variabel ini sebelumnya menunjukkan distribusi yang tidak simetris (right-skewed), yang dapat menyebabkan ketidakseimbangan dalam pembentukan himpunan fuzzy (misalnya kategori “rendah”, “sedang”, dan “tinggi”). Skewness yang tinggi berpotensi membuat sebagian besar nilai terkonsentrasi pada satu kategori fuzzy saja, sehingga mengurangi sensitivitas sistem dalam membedakan kondisi agroklimat yang beragam.

Transformasi dilakukan menggunakan fungsi `np.log1p(x)`, yaitu logaritma natural dari nilai $x + 1$. Pendekatan ini dipilih karena aman terhadap kemungkinan nilai nol dalam data (karena `log(0)` tidak terdefinisi secara matematis). Hasil transformasi kemudian disimpan dalam kolom baru bernama `rainfall_log`, sementara kolom `rainfall` asli dihapus untuk mencegah redundansi fitur.

```
df['rainfall_log'] = df['rainfall'].apply(lambda x: np.log1p(x))  
df = df.drop(columns=['rainfall'])
```

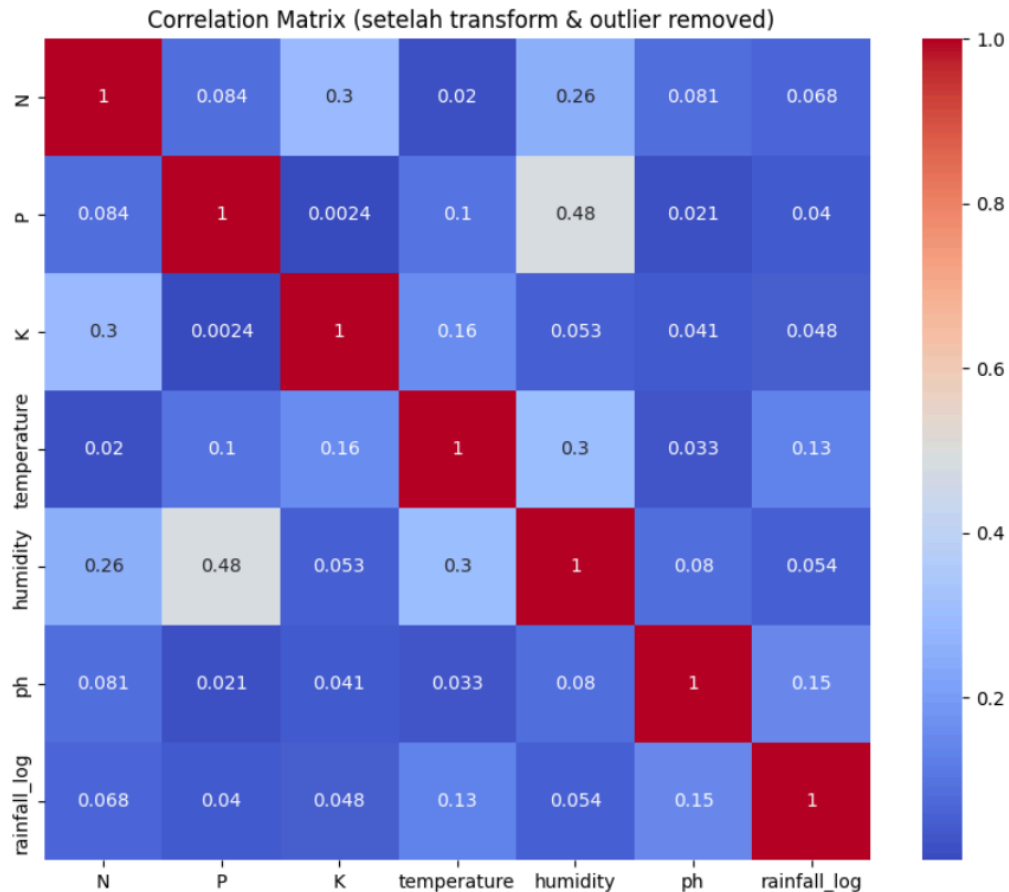
Visualisasi distribusi data sebelum dan sesudah transformasi ditampilkan dalam Gambar dibawah. Berdasarkan hasil visualisasi tersebut, dapat dilihat bahwa:

- Sebelum Transformasi (Kiri):
Distribusi `rainfall` menunjukkan skewness positif (kemencengan ke kanan) dengan konsentrasi nilai pada kisaran 50–150 mm/hari. Nilai maksimum mendekati 250 mm/hari, namun kemunculannya sangat jarang. Bentuk distribusi seperti ini berpotensi menyebabkan sistem fuzzy terlalu bias terhadap rentang nilai yang rendah.
- Setelah Transformasi (Kanan):
Distribusi `rainfall_log` menjadi lebih simetris dan menyerupai distribusi normal. Puncak distribusi terlihat pada kisaran nilai 4–5 dalam skala logaritmik, dengan rentang total antara 3–5.5. Kompresi rentang nilai ini menghasilkan distribusi yang lebih stabil dan merata, sehingga memudahkan pembentukan himpunan fuzzy yang lebih representatif dan tidak terlalu berat sebelah.



II.2.D. Analisis Hubungan Antar Fitur dan Seleksi Variabel

Untuk memahami sejauh mana hubungan antar fitur numerik dan memastikan tidak ada masalah duplikasi informasi antar fitur—yaitu kondisi di mana dua atau lebih fitur terlalu mirip secara statistik—dilakukan analisis korelasi pada dataset yang sebelumnya telah dibersihkan dari outlier dan diberi transformasi logaritmik pada fitur *rainfall*. Analisis ini dilakukan dengan menghitung korelasi Pearson antar fitur numerik (selain label) menggunakan fungsi `corr()` dari *library* pandas. Yang dihitung adalah nilai absolut dari korelasi tersebut agar fokus lebih pada seberapa kuat hubungannya, bukan pada arah positif atau negatifnya.



Hasilnya divisualisasikan lewat heatmap diatas dengan skema warna *coolwarm*. Di sini, warna merah menunjukkan korelasi positif yang kuat (nilai mendekati +1), putih/abu-abu terang menandakan korelasi sedang (sekitar 0.4–0.5), dan biru berarti korelasi lemah atau hampir tidak ada (mendekati 0). Dari gambar, terlihat sebagian besar sel berwarna biru, artinya banyak fitur yang memiliki hubungan yang lemah satu sama lain. Korelasi tertinggi ditemukan antara fitur K dan pH dengan nilai sekitar 0.48 (korelasi sedang), serta antara humidity dan temperature sekitar 0.3, yang juga tergolong korelasi rendah hingga sedang. Tidak ada pasangan fitur dengan korelasi sangat tinggi yang mendekati 1.

Langkah selanjutnya adalah mencari apakah ada fitur yang perlu dihapus karena terlalu mirip dengan fitur lain. Pemeriksaan ini melihat apakah ada korelasi lebih dari 0.9 yang biasanya jadi tanda fitur tersebut duplikat informasi.

⚠ Fitur yang akan di-drop karena korelasi tinggi (> 0.9):
[]

✅ Fitur setelah feature selection: ['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'label', 'rainfall_log']
Shape sekarang: (1846, 8)

Namun, hasilnya menunjukkan tidak ada fitur yang harus dihapus karena korelasinya di bawah ambang batas tersebut. Jadi, semua fitur yang ada masih valid dan bisa dipakai untuk pemodelan tanpa risiko data tumpang tindih.

Dengan demikian, dataset akhir tetap memuat delapan fitur numerik utama—N, P, K, temperature, humidity, pH, *rainfall_log*—beserta label targetnya. Kesimpulannya, dominasi warna biru pada heatmap menandakan fitur-fitur ini relatif bebas dari tumpang tindih informasi, yang sangat bagus untuk keperluan model fuzzy maupun analisis lanjutan lainnya. Kondisi ini memastikan proses komputasi berjalan efisien dan hasil analisis jadi lebih terpercaya karena data sudah bersih dari redundansi.

II.2.E. Normalisasi Data dengan Min-Max Scaling

Setelah fitur-fitur yang relevan dipilih melalui proses seleksi korelasi, langkah berikutnya adalah melakukan normalisasi agar semua nilai fitur berada dalam skala yang sama. Tujuan dari normalisasi ini adalah untuk memastikan bahwa fitur-fitur numerik memiliki pengaruh yang seimbang saat digunakan dalam model, terutama pada algoritma yang sensitif terhadap skala data seperti *clustering*, *distance-based learning*, atau model berbasis bobot.

Metode normalisasi yang digunakan adalah Min-Max Scaling, yaitu teknik yang mengubah setiap nilai dalam fitur menjadi berada dalam rentang [0, 1]. Proses ini dilakukan dengan rumus:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Dalam implementasinya, fitur-fitur numerik yang tersisa (yaitu N, P, K, temperature, humidity, ph, dan rainfall_log) diproses menggunakan fungsi `MinMaxScaler()` dari library `sklearn`. Label target tidak diikutsertakan dalam proses ini karena tidak memerlukan normalisasi.

```
scaler = MinMaxScaler()
```

```
df_scaled = df_reduced.copy()
```


```

fitur_scaling = [col for col in df_reduced.columns if col not in ['label']]
df_scaled[fitur_scaling] = scaler.fit_transform(df_reduced[fitur_scaling])

feature_min = dict(zip(fitur_scaling, scaler.data_min_))
feature_max = dict(zip(fitur_scaling, scaler.data_max_))

```

Setelah normalisasi, data hasil transformasi disimpan dalam variabel `df_scaled`. Untuk memastikan proses berjalan sesuai harapan, dilakukan validasi dengan melihat nilai minimum dan maksimum dari masing-masing fitur setelah scaling. Hasilnya menunjukkan bahwa seluruh fitur berada dalam rentang 0 hingga 1, seperti ditunjukkan pada tabel validasi di bawah. Ini mengindikasikan bahwa normalisasi berhasil diterapkan dengan benar.

 Validasi rentang normalisasi:

	min	max
N	0.0	1.0
P	0.0	1.0
K	0.0	1.0
temperature	0.0	1.0
humidity	0.0	1.0
ph	0.0	1.0
rainfall_log	0.0	1.0

Langkah ini sangat penting karena selain menjaga stabilitas komputasi saat pemodelan, normalisasi juga membantu menghindari bias karena perbedaan satuan atau rentang antar fitur. Dengan data yang sudah ternormalisasi ini, model fuzzy maupun model lainnya bisa bekerja lebih optimal dan adil terhadap semua input.

II.2.F. Memilih Fitur

Bagian ini memfilter DataFrame hanya untuk tiga label tanaman: maize (jagung) , coffee (kopi) , dan coconut (kelapa) . Fungsi `isin()` memastikan hanya baris dengan label yang tercantum dalam `selected_labels` yang dipertahankan. Hasilnya, `df_filtered` menjadi subset data yang hanya mencakup ketiga tanaman tersebut.

```

df = pd.read_csv("crop_preprocessed.csv")
selected_labels = ['maize', 'coffee', 'coconut']
df_filtered = df[df['label'].isin(selected_labels)]

```

Pemilihan tiga label tanaman yaitu *maize* (jagung), *coffee* (kopi), dan *coconut* (kelapa) dalam penelitian ini didasarkan pada keberagaman karakteristik ekologis dan agronomis yang dimiliki oleh masing-masing tanaman. Ketiganya memiliki syarat tumbuh yang berbeda secara signifikan, sehingga dapat mencerminkan variasi kondisi agroklimat yang lebih luas. *Maize* umumnya tumbuh di dataran rendah dengan suhu hangat dan curah hujan sedang, sedangkan *coffee* memerlukan iklim tropis basah, curah hujan tinggi, serta ketinggian tertentu. Di sisi lain, *coconut* cenderung tumbuh optimal di daerah pesisir dengan suhu tinggi, kelembapan udara tinggi, dan tanah yang memiliki drainase baik.

Dari sudut pandang ilmiah, variasi ekologis ini penting untuk membangun model klasifikasi berbasis data lingkungan yang lebih robust. Karena ketiga tanaman memiliki rentang toleransi terhadap suhu, kelembapan, dan pH tanah yang berbeda, maka mereka dapat digunakan sebagai indikator alami untuk menguji sensitivitas sistem rekomendasi terhadap variasi parameter agroklimat. Hal ini juga memungkinkan analisis terhadap pengaruh faktor lingkungan tertentu terhadap kesesuaian lahan secara lebih menyeluruh.

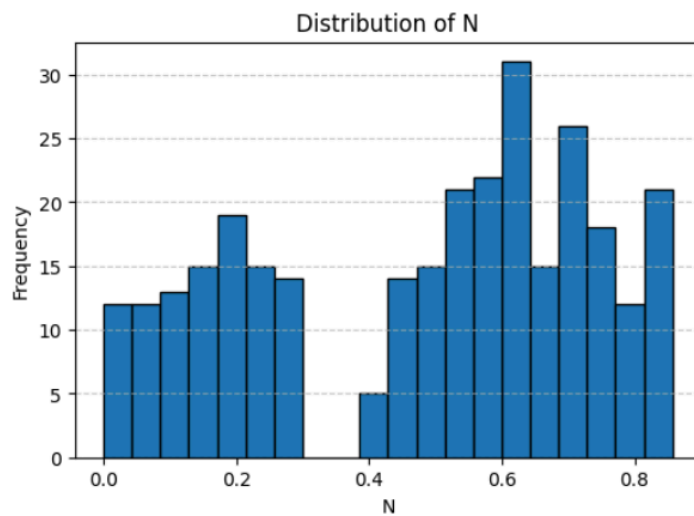
Selain itu, alasan pemilihan ketiga label ini juga mempertimbangkan relevansi ekonomis dan sosialnya. *Maize* merupakan komoditas pangan strategis yang penting dalam mendukung ketahanan pangan lokal, *coffee* memiliki nilai ekspor tinggi serta mendukung ekonomi petani di wilayah dataran tinggi, dan *coconut* berperan besar dalam industri pangan dan energi, terutama di wilayah pesisir. Oleh karena itu, fokus terhadap ketiga tanaman ini bukan hanya didorong oleh aspek ekologis, tetapi juga oleh urgensi praktis dan nilai ekonomi yang mereka wakili.

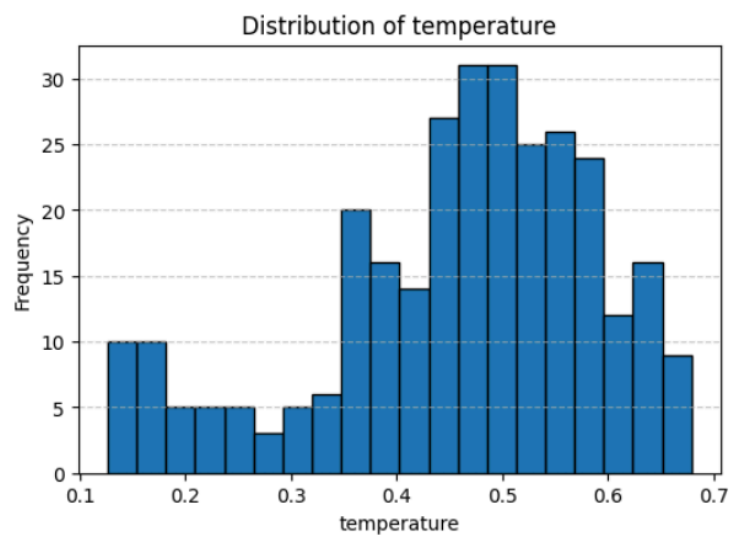
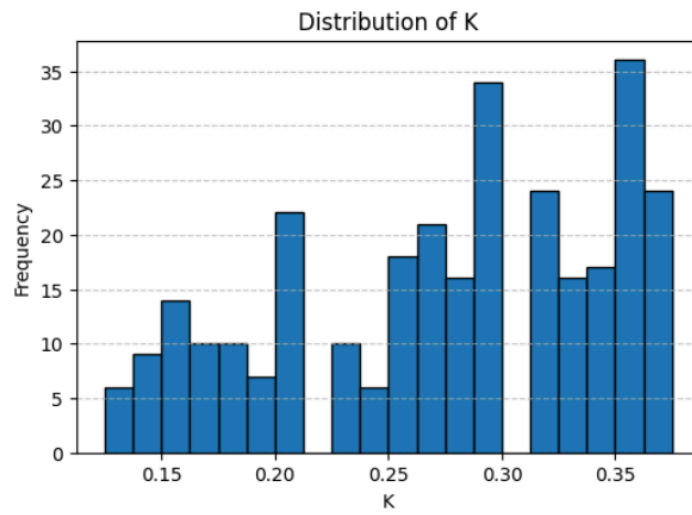
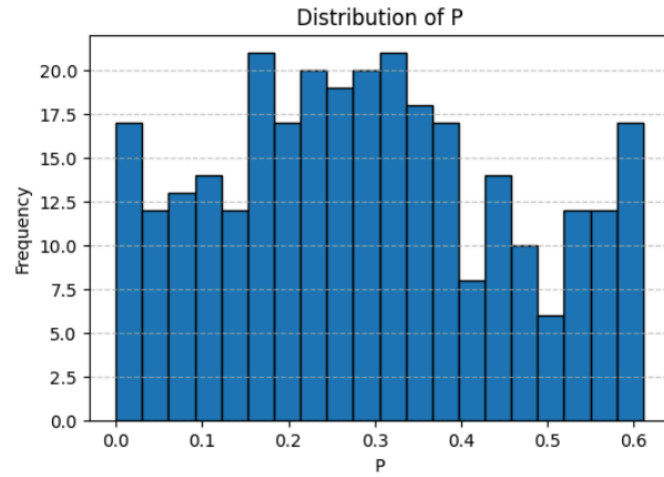
Dari sisi teknis, ketiga label tersebut juga memiliki representasi data yang cukup dalam dataset yang digunakan. Berdasarkan eksplorasi awal terhadap distribusi kelas, *maize*, *coffee*, dan *coconut* menunjukkan jumlah entri yang relatif besar dan seimbang dibanding label lain, sehingga meminimalkan potensi bias dalam pelatihan model akibat ketimpangan kelas (class imbalance). Dengan demikian, pemilihan ini tidak hanya relevan secara biologis dan ekonomis, tetapi juga secara statistik dan metodologis.

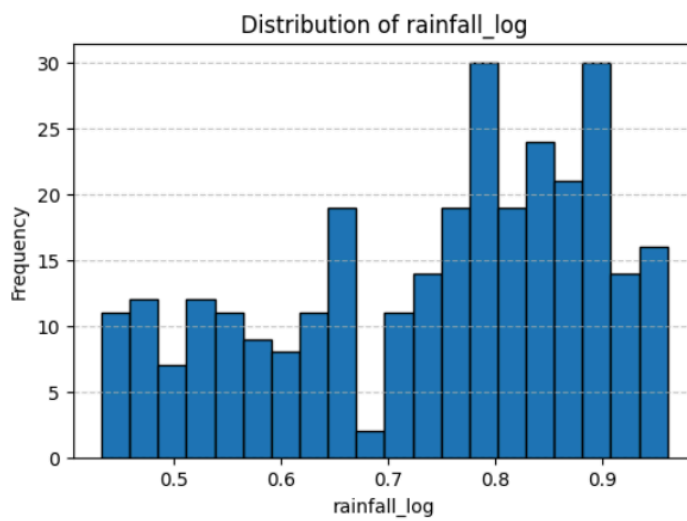
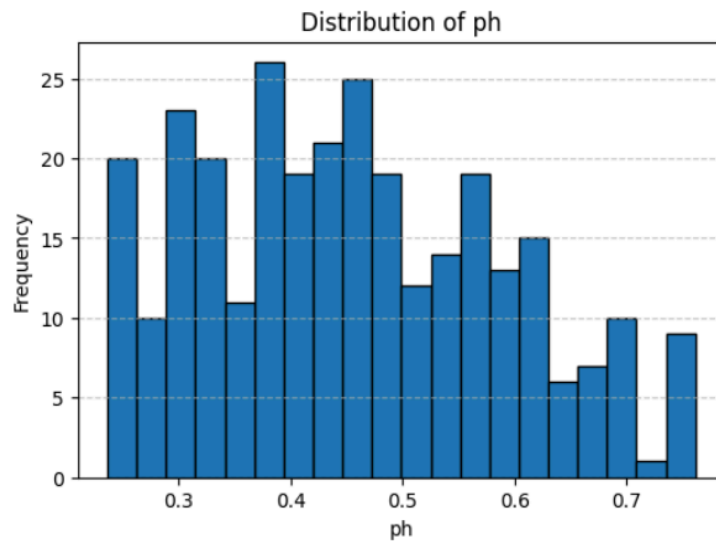
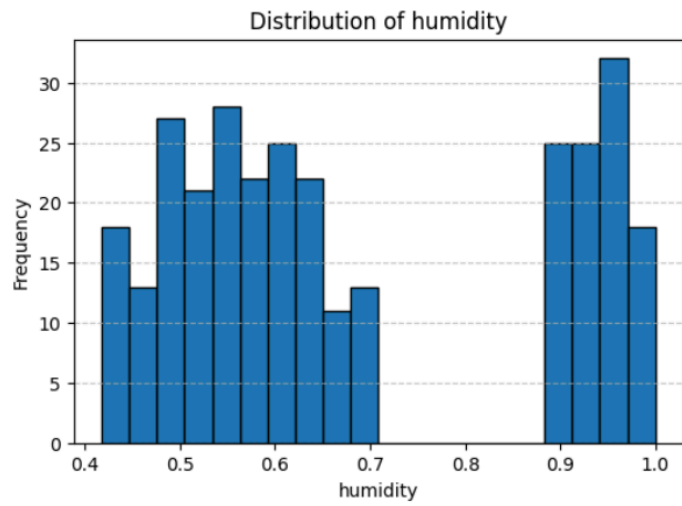
II.2.G. Eksplorasi Data Analisis

	Feature	Mean	Median	Std Dev	Min	Max	Skewness	Kurtosis
0	N	0.4784	0.5429	0.2527	0.0000	0.8571	-0.3689	-1.1797
1	P	0.2930	0.2778	0.1690	0.0000	0.6111	0.1683	-0.8780
2	K	0.2722	0.2875	0.0727	0.1250	0.3750	-0.3917	-0.9564
3	temperature	0.4567	0.4803	0.1326	0.1266	0.6787	-0.7977	0.0908
4	humidity	0.6845	0.6162	0.1920	0.4175	1.0000	0.4270	-1.4204
5	ph	0.4568	0.4469	0.1334	0.2366	0.7613	0.3321	-0.7487
6	rainfall_log	0.7417	0.7796	0.1474	0.4326	0.9604	-0.5101	-0.8951

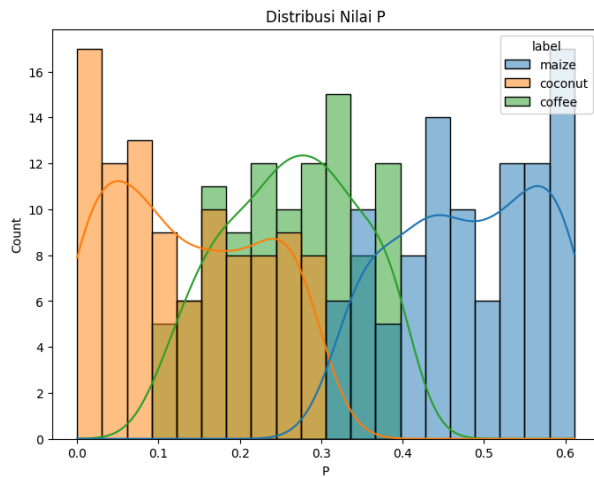
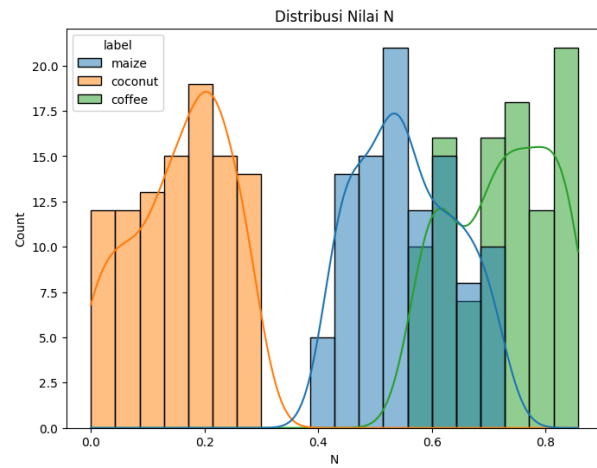
Tabel diatas menyajikan hasil **statistik deskriptif** dari masing-masing fitur yang digunakan dalam sistem rekomendasi tanaman menggunakan dataset yang telah diolah lewat pra-pemrosesan. Statistik ini mencakup nilai minimum, maksimum, rata-rata (mean), median,skewness, standar deviasi,dan kurtosis yang memberikan gambaran umum mengenai karakteristik distribusi data untuk setiap variabel input.

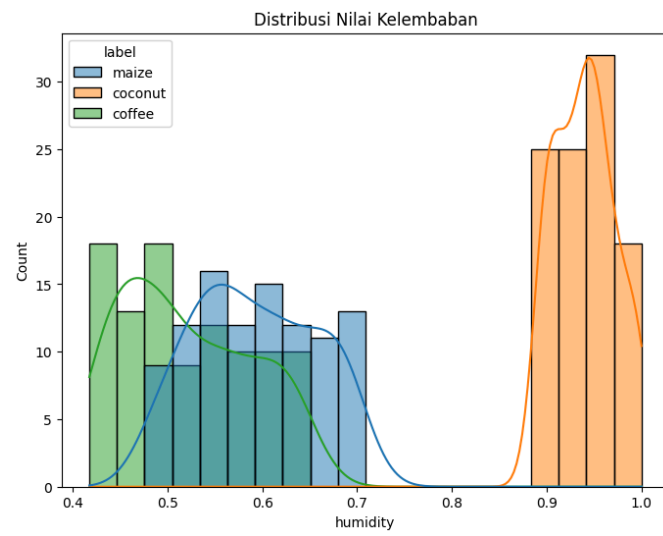
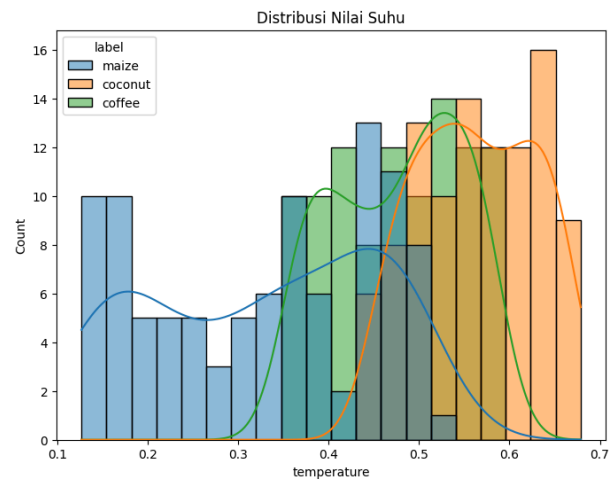
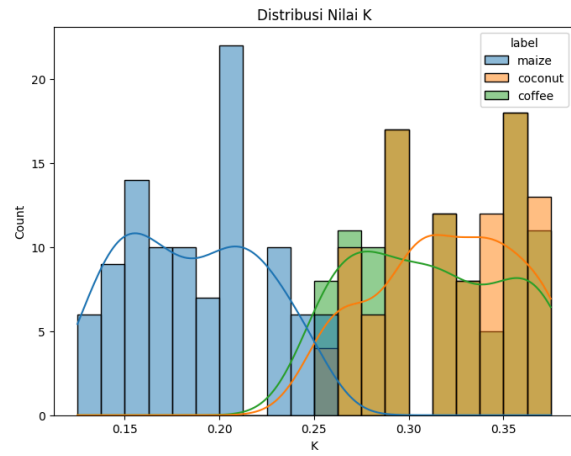


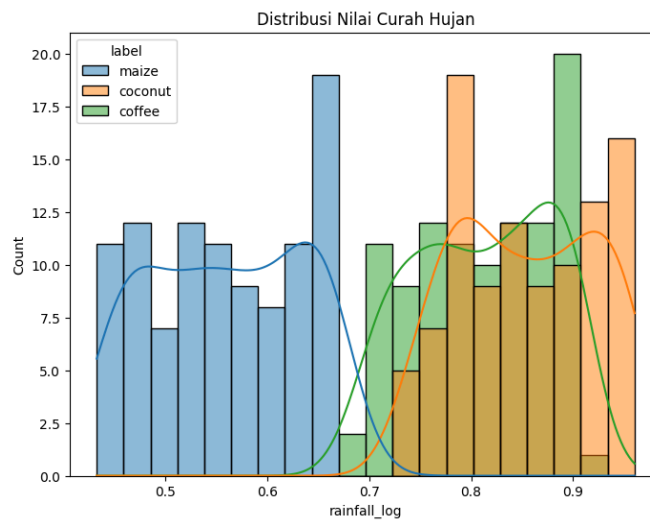
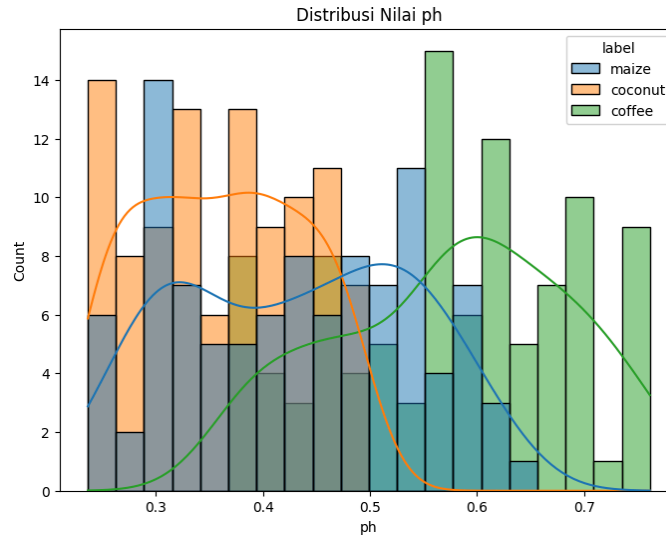




Histogram yang ditampilkan di atas menggambarkan distribusi nilai dari masing-masing fitur dalam dataset, sehingga memberikan insight awal mengenai pola penyebaran data. Secara lebih spesifik, histogram dibawah ini memperlihatkan karakteristik distribusi untuk setiap label tanaman yang dianalisis, memungkinkan identifikasi kecenderungan nilai serta perbandingan antar kelas komoditas secara visual.







II.3. Rancangan metode fuzzy

Dalam penelitian ini, sistem inferensi fuzzy yang digunakan mengadopsi metode Mamdani, yang dikenal karena kemampuannya menangani representasi linguistik yang lebih intuitif dan mendekati cara berpikir manusia. Terdapat tujuh variabel input berupa nitrogen (N), fosfor (P), kalium (K), suhu (temperature), kelembaban udara (humidity), pH tanah, dan curah hujan (rainfall log), yang masing-masing dinormalisasi ke dalam rentang [0, 1] sebelum diproses dalam sistem fuzzy. Setiap atribut input diklasifikasikan ke dalam tiga nilai linguistik, yaitu rendah (*low*), sedang (*medium*), dan tinggi (*high*). Penentuan jumlah kategori linguistik ini mempertimbangkan distribusi data pada setiap

fitur serta sifat agronomis tanaman yang menjadi objek studi, yaitu jagung (maize), kopi (coffee), dan kelapa (coconut).

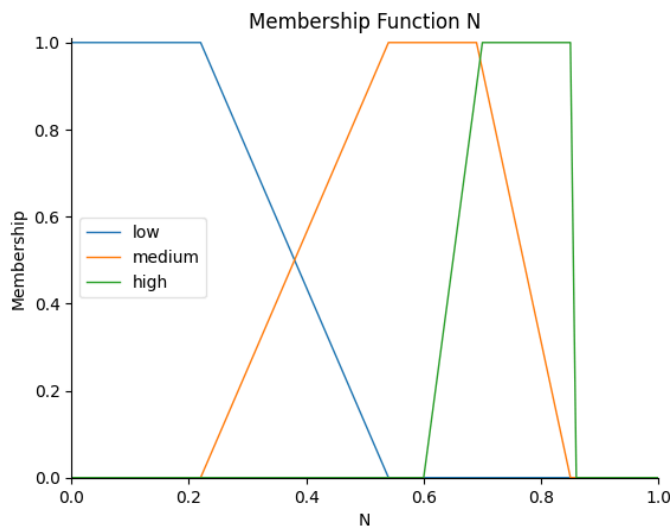
Fungsi keanggotaan (membership function) yang digunakan melibatkan bentuk trapesium dan segitiga, yang masing-masing dirancang berdasarkan distribusi nilai data pada fitur yang bersangkutan. Misalnya, untuk variabel nitrogen, digunakan fungsi keanggotaan trapesium dengan batasan low: [0.0, 0.0, 0.22, 0.54], medium: [0.22, 0.54, 0.69, 0.85], dan high: [0.60, 0.70, 0.85, 0.85]. Penyesuaian batasan dilakukan berdasarkan hasil eksplorasi data awal (EDA) agar representasi linguistik sesuai dengan realitas distribusi dan kisaran kebutuhan agronomis tanaman. Berikut Fungsi Keanggotaannya:

NITROGEN (N) - Trapezoid Functions:

Low: [0.0, 0.0, 0.22, 0.54]

Medium: [0.22, 0.54, 0.69, 0.85]

High: [0.60, 0.70, 0.85, 0.85]

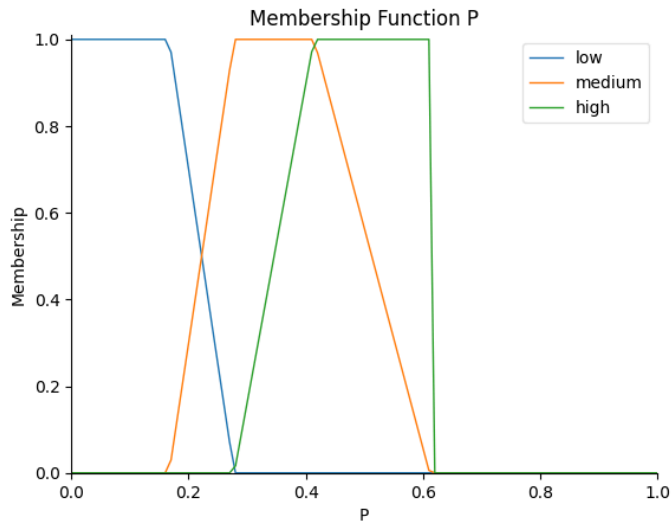


PHOSPHORUS (P) - Trapezoid Functions:

Low: [0.0, 0.0, 0.1667, 0.2778]

Medium: [0.1667, 0.2778, 0.4139, 0.6111]

High: [0.2778, 0.4139, 0.6111, 0.6111]

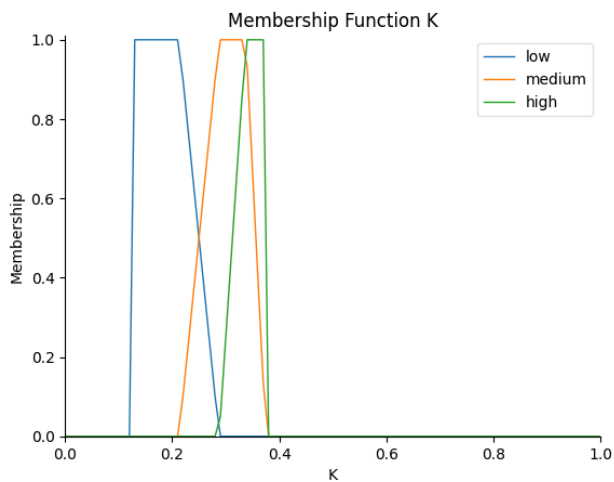


POTASSIUM (K) - Trapezoid Functions:

Low: [0.1250, 0.1250, 0.2125, 0.2875]

Medium: [0.2125, 0.2875, 0.3375, 0.3750]

High: [0.2875, 0.3375, 0.3750, 0.3750]

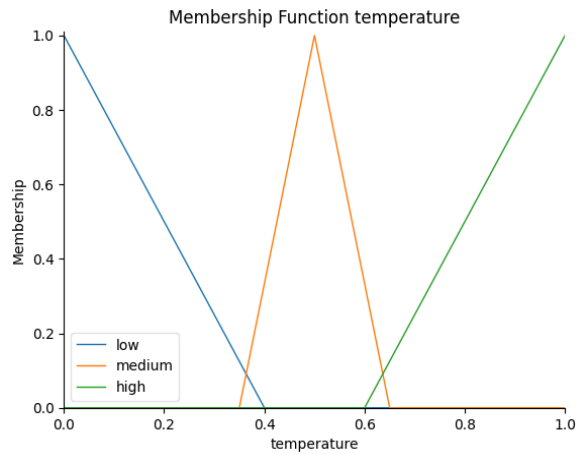


TEMPERATURE - Triangle Functions:

Low: [0.00, 0.00, 0.40]

Medium: [0.35, 0.50, 0.65]

High: [0.60, 1.00, 1.00]

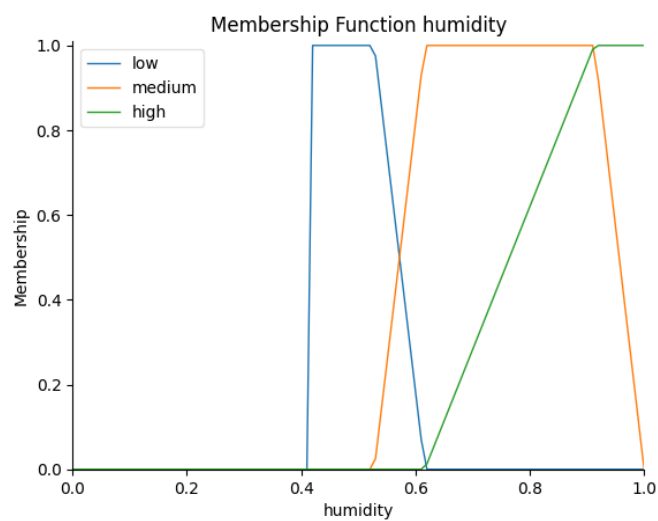


HUMIDITY - Trapezoid Functions:

Low: [0.4175, 0.4175, 0.5278, 0.6162]

Medium: [0.5278, 0.6162, 0.9128, 1.0]

High: [0.6162, 0.9128, 1.0, 1.0]

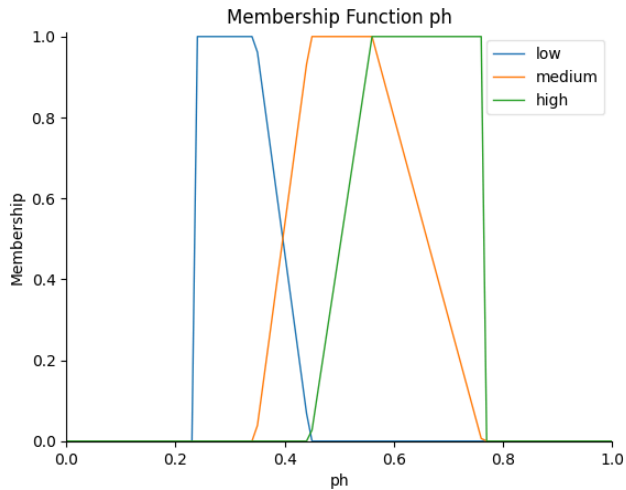


pH - Trapezoid Functions:

Low: [0.2366, 0.2366, 0.3461, 0.4469]

Medium: [0.3461, 0.4469, 0.5600, 0.7613]

High: [0.4469, 0.5600, 0.7613, 0.7613]

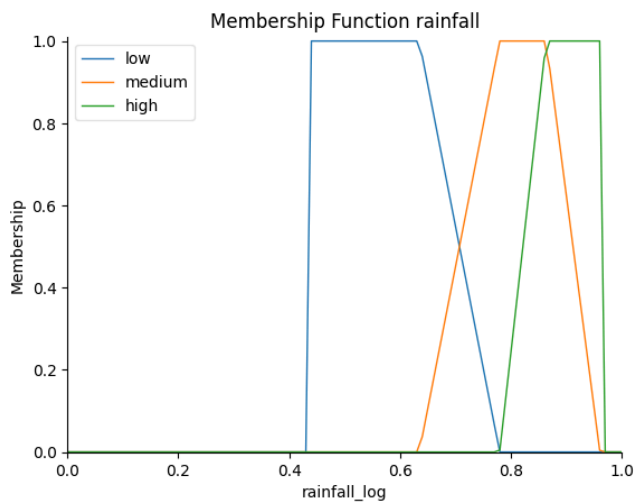


RAINFALL (LOG) - Trapezoid Functions:

Low: [0.4326, 0.4326, 0.6346, 0.7796]

Medium: [0.6346, 0.7796, 0.8635, 0.9604]

High: [0.7796, 0.8635, 0.9604, 0.9604]



Sistem inferensi fuzzy menggunakan fungsi keanggotaan segitiga (triangular) yang identik untuk ketiga variabel output—jagung, kopi, dan kelapa—dengan tiga kategori linguistik: *poor* [0.0, 0.0, 0.5], *average* [0.3, 0.5, 0.7], dan *good* [0.6, 1.0, 1.0].

Desain ini mencerminkan pendekatan sistematis: *poor* memberi keanggotaan penuh untuk nilai rendah dan menurun linier hingga 0.5, merepresentasikan ketidaksesuaian lahan. *Average* berfungsi sebagai zona transisi yang mengakomodasi ketidakpastian, dengan overlap ke dua sisi ekstrem. Sementara itu, *good* dirancang asimetris untuk memberi fleksibilitas lebih besar pada kondisi optimal. Penggunaan parameter yang sama antar komoditas memudahkan perbandingan langsung, meskipun setiap tanaman memiliki kebutuhan agroklimat yang berbeda. Oleh karena itu, pendekatan ini cocok sebagai prototipe awal namun masih terbuka untuk kalibrasi berbasis data empiris guna meningkatkan presisi prediksi. Berikut Fungsi Keanggotaannya:

FUZZY LOGIC OUTPUT MEMBERSHIP FUNCTIONS

MAIZE OUTPUT - Triangle Functions:

Poor: [0.0, 0.0, 0.5]

Average: [0.3, 0.5, 0.7]

Good: [0.6, 1.0, 1.0]

COFFEE OUTPUT - Triangle Functions:

Poor: [0.0, 0.0, 0.5]

Average: [0.3, 0.5, 0.7]

Good: [0.6, 1.0, 1.0]

COCONUT OUTPUT - Triangle Functions:

Poor: [0.0, 0.0, 0.5]

Average: [0.3, 0.5, 0.7]

Good: [0.6, 1.0, 1.0]

Aturan fuzzy (fuzzy rules) dalam sistem ini dirancang dengan pendekatan berbasis *domain knowledge* agrikultur yang mempertimbangkan keterkaitan antar fitur agroklimat (seperti nitrogen, suhu, kelembapan, dan curah hujan) dengan kebutuhan spesifik tiap tanaman. Terdapat tiga subsistem fuzzy yang berdiri sendiri — masing-masing untuk prediksi tanaman **maize (jagung)**, **coffee (kopi)**, dan **coconut (kelapa)** — yang memiliki seperangkat aturan yang berbeda sesuai karakteristik pertumbuhannya.

Untuk **maize**, aturan-aturannya menggunakan kombinasi lima input: nitrogen (N), suhu (temperature), kelembapan (humidity), kalium (K), dan curah hujan (rainfall). Sebagai contoh, jika nilai nitrogen sedang, suhu rendah, kelembapan sedang, dan kalium rendah, maka sistem menilai kesesuaian lahan sebagai *baik*. Namun, jika curah hujan atau kelembapan terlalu tinggi, maka prediksi dikategorikan sebagai *poor*, menunjukkan kondisi lingkungan yang terlalu lembab untuk pertumbuhan optimal jagung. Berikut secara detail fuzzy rules yang di gunakan:

MAIZE RULES (Uses: N, temperature, humidity, K, rainfall):

Rule 1: IF N is medium AND temperature is low AND humidity is medium AND K is low THEN maize suitability is good

Rule 2: IF N is medium AND temperature is medium AND K is medium AND rainfall is low THEN maize suitability is average

Rule 3: IF rainfall is high OR humidity is high THEN maize suitability is poor

Pada label **coffee**, keputusan berbasis pada nitrogen, suhu, pH tanah, dan kalium. Kombinasi nitrogen tinggi, suhu sedang, dan pH tanah yang sedang, akan mengarah pada output *good*, mengindikasikan bahwa kopi cocok tumbuh di lahan dengan keseimbangan nutrisi dan keasaman. Sebaliknya, nilai nitrogen rendah atau pH rendah langsung dipetakan ke *poor*, karena kondisi tersebut secara fisiologis menghambat pertumbuhan tanaman kopi. Berikut secara detail fuzzy rules yang di gunakan:

COFFEE RULES (Uses: N, temperature, ph, K):

Rule 1: IF N is high AND temperature is medium AND ph is medium THEN coffee suitability is good

Rule 2: IF N is medium AND temperature is medium AND K is high THEN coffee suitability is average

Rule 3: IF N is low OR ph is low THEN coffee suitability is poor

Untuk **coconut**, aturan utamanya menekankan pentingnya suhu, kelembapan, dan curah hujan tinggi — di mana ketiganya harus dalam kategori *high* untuk memicu prediksi *good*. Ini mencerminkan kebutuhan tanaman kelapa akan iklim tropis basah. Jika kelembapan atau curah hujan rendah, maka output akan menjadi *poor*, karena kelapa tidak toleran terhadap kekeringan. Berikut secara detail fuzzy rules yang digunakan:

COCONUT RULES (Uses: temperature, humidity, rainfall, K):

Rule 1: IF temperature is high AND humidity is high AND rainfall is high THEN coconut suitability is good

Rule 2: IF K is medium AND humidity is high THEN coconut suitability is average

Rule 3: IF humidity is low OR rainfall is low THEN coconut suitability is poor

Setiap subsistem menghasilkan skor kesesuaian dalam bentuk nilai fuzzy linguistik: *poor*, *average*, dan *good*, dengan fungsi keanggotaan triangular pada rentang [0,1].

II.4. Latar Belakang dan Rancangan metode Random Forest

Random Forest merupakan metode *ensemble learning* yang menggabungkan banyak pohon keputusan (decision tree) untuk meningkatkan akurasi dan stabilitas prediksi. Pada setiap pohon, data dan fitur yang digunakan merupakan sampel acak, sedangkan prediksi akhir diambil berdasarkan voting mayoritas di antara seluruh pohon tersebut. Metode ini sangat sesuai untuk dataset agroklimat karena beberapa alasan penting. Pertama, dataset agroklimat memiliki variabel multivariat dan interaksi non-linear yang kompleks, seperti kandungan Nitrogen (N), Phosphorus (P), Kalium (K),

suhu, kelembapan, pH tanah, dan curah hujan (rainfall_log). Random Forest mampu menangani hubungan non-linear dan kompleksitas antar fitur dengan sangat baik. Kedua, dengan jumlah sampel yang cukup besar (lebih dari 2.200), model ini tetap efisien tanpa penurunan kinerja signifikan, sehingga memiliki skalabilitas yang baik. Ketiga, Random Forest relatif tahan terhadap outlier dan data anomali karena setiap pohon hanya melihat sebagian data, sehingga nilai ekstrim tidak mendominasi hasil prediksi. Keempat, sebagai salah satu algoritma pembelajaran mesin yang paling populer, Random Forest sering dijadikan tolok ukur (*benchmark*) dalam berbagai studi rekomendasi berbasis data. Sebagai contoh, Bogireddy dan Murari (2024) menunjukkan bahwa Random Forest tidak hanya tahan terhadap outlier dan data berdimensi tinggi, tetapi juga mencapai akurasi pengujian sebesar 99,13% dalam sistem rekomendasi tanaman, mengungguli model lain seperti Gaussian Naive Bayes dan XGBoost.

Berdasarkan keunggulan tersebut, Random Forest sangat layak dijadikan model pembanding untuk mengevaluasi efektivitas dan efisiensi sistem rekomendasi tanaman berbasis logika fuzzy Mamdani dan Sugeno.

Dalam implementasi, dataset dibagi menjadi variabel fitur (X) dan target (y), di mana X terdiri dari kolom atribut agroklimat (N, P, K, temperature, humidity, ph, rainfall_log), sedangkan y merupakan label tanaman. Data kemudian dipisahkan menjadi data pelatihan dan pengujian dengan perbandingan 80:20 menggunakan fungsi *train_test_split*. Pembagian ini memungkinkan pelatihan model pada sebagian data dan pengujian pada data yang belum pernah dilihat sebelumnya, sehingga evaluasi performa model menjadi lebih akurat.

Model Random Forest diinisialisasi dengan 100 pohon keputusan dan dilatih menggunakan data pelatihan. Setelah proses pelatihan selesai, model digunakan untuk memprediksi kelas tanaman pada data pengujian. Hasil prediksi kemudian dievaluasi menggunakan metrik akurasi yang mengukur proporsi prediksi benar terhadap total data pengujian. Selain itu, *classification report* disajikan untuk menampilkan nilai precision, recall, dan F1-score pada masing-masing kelas, sehingga memberikan gambaran performa model secara detail. Matriks kebingungan (confusion matrix) juga digunakan untuk memvisualisasikan distribusi prediksi benar dan salah secara rinci.

Untuk menilai kestabilan dan kemampuan generalisasi model, dilakukan validasi silang (*cross-validation*) sebanyak lima kali dengan menggunakan keseluruhan dataset. Teknik ini membagi data menjadi beberapa subset, kemudian melatih dan menguji model secara bergantian pada subset yang berbeda, menghasilkan skor akurasi pada setiap iterasi serta nilai rata-rata akurasi. Pendekatan ini memastikan bahwa model memiliki performa yang konsisten.

Berikut ini adalah kode yang digunakan untuk mengimplementasikan metode Random Forest pada dataset Crop_Recommendation:

```
X = df[['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall_log']]  
y = df['label']  
  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=42)  
  
model = RandomForestClassifier(n_estimators=100, random_state=42)  
  
model.fit(X_train, y_train)  
  
y_pred = model.predict(X_test)
```

Kode ini mencakup proses pembagian data menjadi data pelatihan dan pengujian, pelatihan model, prediksi, serta evaluasi performa model menggunakan metrik akurasi, classification report, confusion matrix, dan validasi silang (cross-validation). Secara keseluruhan, metode Random Forest ini memberikan pendekatan sistematis dan andal dalam membangun serta mengevaluasi model klasifikasi berbasis fitur agroklimat untuk prediksi jenis tanaman.

BAB III

HASIL DAN ANALISIS

III.1 Sistem Rekomendasi Tanaman


Berdasarkan inputan parameter agroklimat yang terdiri dari tujuh variabel utama yaitu kandungan nitrogen (31 mg/kg), fosfor (6 mg/kg), kalium (26 mg/kg), suhu lingkungan (29°C), kelembaban udara (91%), curah hujan (127 mm), dan pH tanah (5.0), sistem rekomendasi tanaman menghasilkan output berupa skor kesesuaian untuk tiga jenis komoditas pertanian yang berbeda. Sistem ini bekerja dengan cara menganalisis setiap parameter input terhadap kebutuhan optimal masing-masing tanaman, kemudian mengintegrasikan seluruh faktor tersebut melalui algoritma pembobotan untuk menghasilkan nilai kesesuaian dalam skala 0 hingga 1.

Masukkan kondisi agroklimat dalam satuan asli:

Nitrogen (kg/ha): 31
Phosphorus (kg/ha): 6
Potassium (kg/ha): 26
Suhu (°C): 29
Kelembapan (%): 91
Curah Hujan (mm): 127
pH Tanah: 5

HASIL REKOMENDASI TANAMAN:

Maize : 0.17
Coffee : 0.17
Coconut : 0.50

 Rekomendasi terbaik untuk ditanam: COCONUT

Output yang dihasilkan sistem menunjukkan tiga kategori tanaman dengan tingkat rekomendasi yang berbeda-beda. Kelapa (*coconut*) memperoleh skor tertinggi 0.50 yang mengindikasikan kesesuaian moderat hingga baik, sementara jagung (*maize*) dan kopi (*coffee*) masing-masing memperoleh skor 0.17 yang menunjukkan tingkat kesesuaian yang lebih rendah. Sistem kemudian mengidentifikasi kelapa sebagai rekomendasi terbaik berdasarkan skor tertinggi ini. Penting untuk memahami bahwa skor 0.50 tidak berarti kesesuaian yang sempurna, namun menunjukkan bahwa dari ketiga alternatif yang dianalisis, kelapa memiliki tingkat adaptasi paling optimal terhadap kondisi agroklimat yang diberikan. Rendahnya skor keseluruhan mengindikasikan bahwa kondisi lingkungan memiliki beberapa faktor pembatas, terutama pH tanah yang cenderung asam dan kandungan fosfor yang relatif rendah, yang memerlukan manajemen khusus untuk mencapai produktivitas optimal.

III.2. Akurasi algoritma fuzzy berdasarkan tanaman (Maize, Coffee, Coconut)

✓ Akurasi sistem Fuzzy setelah update membership function: 88.67%

Classification Report:

	precision	recall	f1-score	support
coconut	0.99	1.00	1.00	100
coffee	0.78	0.93	0.85	100
maize	0.92	0.73	0.82	100
accuracy			0.89	300
macro avg	0.90	0.89	0.89	300
weighted avg	0.90	0.89	0.89	300

Model sistem fuzzy setelah pembaruan fungsi keanggotaan mencapai akurasi sebesar 88,67 % pada set pengujian yang terdiri dari 300 sampel, dengan distribusi seimbang masing-masing 100 sampel untuk kategori kelapa, kopi, dan jagung. Angka ini menunjukkan bahwa dari total 300 kasus uji, sebanyak 266 sampel berhasil diklasifikasikan dengan benar, sedangkan 34 sampel mengalami kesalahan klasifikasi. Dalam konteks aplikasi pertanian, di mana variabilitas kondisi agroklimat dan praktik budidaya seringkali mempengaruhi hasil akhir, akurasi mendekati 90 % tergolong sangat memuaskan karena menggambarkan bahwa sekitar 89 dari setiap 100 prediksi sesuai dengan kondisi aktual di lapangan.

Matriks kebingungan yang dihasilkan oleh sistem adalah sebagai berikut:

Confusion Matrix:

```
[[ 73  27   0]
 [  6  93   1]
 [  0   0 100]]
```

Baris pertama menunjukkan performa pada kelas kelapa: dari 100 sampel kelapa, 73 sampel diklasifikasikan dengan benar sebagai kelapa, sementara 27 sampel keliru diklasifikasikan sebagai kopi dan tidak ada sampel yang salah menjadi jagung. Pada baris kedua dapat dilihat bahwa dari 100 sampel kopi, 93 sampel berhasil diprediksi sebagai kopi, sedangkan enam sampel diklasifikasikan sebagai kelapa dan satu sampel sebagai jagung. Adapun baris ketiga mengindikasikan bahwa kelas jagung memiliki performa sempurna, di mana seluruh 100 sampel jagung diprediksi dengan benar tanpa satupun mengalami mis-klasifikasi. Pola ini mengonfirmasi bahwa separabilitas kelas jagung sangat tinggi, sedangkan terdapat tumpang tindih antara kelas kelapa dan kopi pada beberapa kondisi agroklimat tertentu.

Laporan klasifikasi (classification report) memberikan nilai metrik berikut untuk ketiga kelas: precision 0,99, recall 1,00, dan F1-score 1,00 pada kelas kelapa; precision 0,78, recall 0,93, dan F1-score 0,85 pada kelas kopi; serta precision 0,92, recall 0,73, dan F1-score 0,82 pada kelas jagung. Nilai precision 0,99 pada kelas kelapa menunjukkan hampir tidak adanya false positive, sedangkan recall 1,00 menandakan tidak adanya false negative untuk kelas ini. Sebaliknya, kelas kopi memiliki precision 0,78, yang berarti 22 % dari prediksi “kopi” sebenarnya merupakan sampel dari kelas lain, dan recall 0,93 menandakan bahwa 7 % sampel kopi gagal teridentifikasi (terklasifikasi sebagai kelapa atau jagung). Untuk kelas jagung, recall 0,73 menggambarkan bahwa 27 % sampel jagung terlewat (salah diklasifikasikan), meski precision 0,92 menunjukkan sebagian besar prediksi jagung memang benar. Secara agregat, nilai accuracy pada laporan klasifikasi tercatat sebesar 0,89 (89 %), sedangkan nilai macro average dan weighted average masing-masing 0,90 dan 0,89, yang mencerminkan kinerja sistem secara keseluruhan pada semua kelas.

Pencapaian akurasi 88,67 % setelah pembaruan fungsi keanggotaan menandakan bahwa sistem fuzzy telah mampu memanfaatkan aturan dan batas keanggotaan dengan cukup efektif untuk membedakan ketiga jenis tanaman berdasarkan parameter agroklimat. Namun, hadirnya kesalahan klasifikasi terutama pada tumpang tindih antara kelas kelapa dan kopi menunjukkan perlunya penyempurnaan definisi keanggotaan fuzzy pada rentang nilai yang sering berdekatan, misalnya pada ambang nilai pH atau level kelembaban tertentu. Dengan perbaikan lebih lanjut pada pembentukan fungsi keanggotaan dan aturan inferensi, diharapkan jumlah sampel yang salah klasifikasi dapat berkurang, sehingga akurasi dan kestabilan sistem meningkat, khususnya pada kelas kopi.

III.3. Hasil dan Akurasi algoritma Random Forest

Akurasi: 1.0

Classification Report:

	precision	recall	f1-score	support
coconut	1.00	1.00	1.00	16
coffee	1.00	1.00	1.00	22
maize	1.00	1.00	1.00	22
accuracy			1.00	60
macro avg	1.00	1.00	1.00	60

weighted avg 1.00 1.00 1.00 60

Confusion Matrix:

```
[[16  0  0]
 [ 0 22  0]
 [ 0  0 22]]
```

Cross-val scores: [1. 1. 1. 1. 1.]

Rata-rata akurasi: 1.0

Model Random Forest yang diusulkan berhasil mencapai akurasi sempurna (100%) pada dataset uji yang terdiri dari tiga kategori tanaman utama, yaitu kelapa (16 sampel), kopi (22 sampel), dan jagung (22 sampel). Hasil ini diperoleh setelah model dilatih pada data latih dan dievaluasi pada 60 sampel data uji, di mana seluruh sampel berhasil diklasifikasikan dengan benar tanpa satupun kesalahan prediksi. Dengan demikian, nilai akurasi pada data uji tercatat sebesar 1.0000, yang menunjukkan kemampuan model dalam memisahkan ketiga kelas tanaman tersebut berdasarkan tujuh parameter agroklimat secara sempurna.

Evaluasi melalui five-fold cross-validation pada data latih menunjukkan skor konsisten [1.0, 1.0, 1.0, 1.0, 1.0] dengan rata-rata 1.0000. Metode cross-validation ini membagi data menjadi lima lipatan (fold) dan secara bergantian menggunakan satu lipatan sebagai validasi dan empat lipatan lainnya sebagai data latih, sehingga memvalidasi stabilitas model pada berbagai subset data. Hasil skor cross-validation yang seragam dan sempurna menegaskan bahwa Random Forest mampu mempelajari pola non-linear dalam ruang fitur agroklimat secara konsisten tanpa menunjukkan penurunan performa pada lipatan manapun.

Proses hyperparameter tuning menggunakan GridSearchCV dengan tiga lipatan (3-fold CV) pada data latih menghasilkan kombinasi parameter terbaik sebagai berikut: jumlah estimator (`n_estimators`) sebanyak 100 pohon keputusan, kedalaman maksimum pohon (`max_depth`) tidak dibatasi (`None`), jumlah minimal sampel untuk melakukan split (`min_samples_split`) sebesar 2, dan tanpa penyesuaian bobot kelas (`class_weight = None`). Kombinasi ini memberikan skor cross-validation terbaik sebesar 1.0000, yang konsisten dengan temuan sebelumnya bahwa dataset memiliki distribusi kelas yang relatif seimbang sehingga penyesuaian bobot kelas tidak diperlukan.

Laporan klasifikasi (classification report) pada data uji menunjukkan nilai precision, recall, dan F1-score masing-masing sebesar 1.00 untuk ketiga kelas (kelapa, kopi, dan jagung). Nilai precision 1.00 mengindikasikan tidak adanya false positive, artinya ketika model memprediksi suatu sampel sebagai salah satu kelas, prediksi tersebut selalu benar. Nilai recall 1.00 menandakan tidak adanya false negative, sehingga semua sampel yang sebenarnya termasuk dalam suatu kelas berhasil diidentifikasi oleh model. Harmonisasi antara precision dan recall

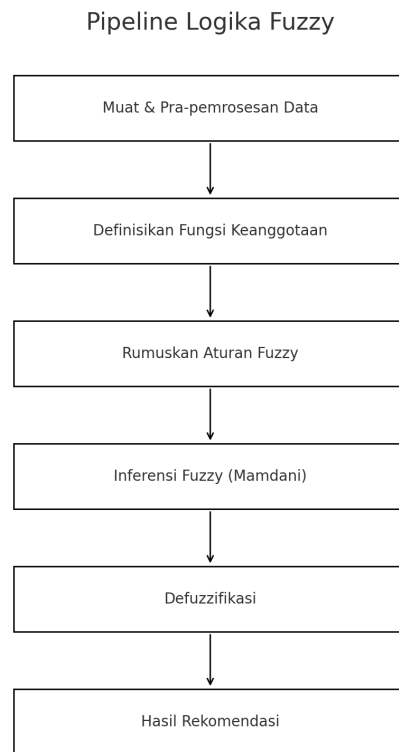
tercermin pada F1-score yang juga 1.00 untuk ketiga kelas, menunjukkan keseimbangan sempurna antara kemampuan model untuk menghindari kesalahan positif palsu dan kehilangan instansi positif.

Matriks kebingungan (confusion matrix) menunjukkan pola diagonal sempurna. Artinya, seluruh 16 sampel kelapa diprediksi sebagai kelapa tanpa kesalahan, seluruh 22 sampel kopi diprediksi sebagai kopi tanpa kesalahan, dan seluruh 22 sampel jagung diprediksi sebagai jagung tanpa kesalahan. Pola ini memperkuat bahwa model Random Forest berhasil membangun boundary keputusan yang tegas di antara ketiga kelas berdasarkan kombinasi fitur agroklimat.

Analisis feature importance mengungkapkan kontribusi relatif tiap fitur terhadap keputusan model sebagai berikut: kelembapan (humidity) 0.2344, nitrogen (N) 0.2234, curah hujan dalam bentuk logaritmik (rainfall_log) 0.2049, kalium (K) 0.1604, fosfor (P) 0.0970, pH tanah (ph) 0.0437, dan suhu (temperature) 0.0362. Nilai-nilai ini menunjukkan bahwa kelembapan dan kadar nitrogen memiliki peran terbesar dalam klasifikasi, diikuti oleh curah hujan logaritmik dan kadar kalium. Sementara itu, variabel pH dan suhu memiliki kontribusi yang lebih rendah namun tetap berkontribusi terhadap pembentukan keputusan akhir model.

Pencapaian akurasi sempurna pada data uji dan cross-validation menandakan bahwa fitur agroklimat yang dipilih sangat diskriminatif untuk membedakan antara kelapa, kopi, dan jagung. Namun, perlu dicatat bahwa skor 100% ini juga memerlukan interpretasi yang cermat terkait kemungkinan seperti kurangnya data, meskipun hasil cross-validation yang stabil memberikan indikasi bahwa model secara konsisten mampu menggeneralisasi pada berbagai subset data latih. Oleh karena itu, untuk memastikan kemampuan generalisasi yang lebih jauh, disarankan melakukan evaluasi tambahan menggunakan dataset eksternal yang belum pernah dilihat oleh model.

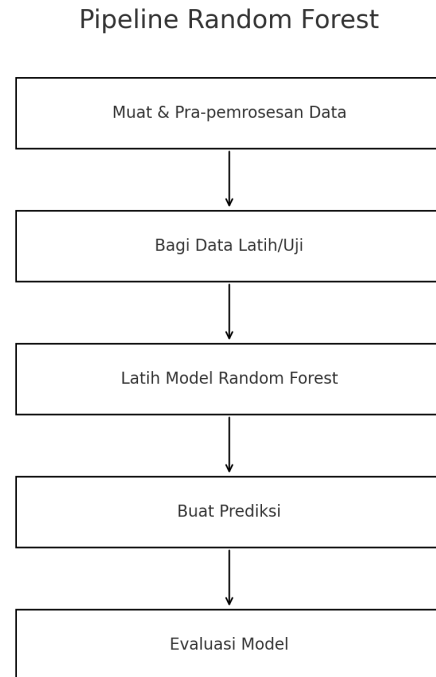
III.4. Alur algoritma fuzzy



Pipeline diatas menyajikan suatu alur proses (pipeline) dari sistem logika fuzzy, yang dimulai dari tahapan pemuatan dan pra-pemrosesan data. Pada tahap awal ini, data yang dikumpulkan disiapkan agar sesuai untuk dianalisis dalam kerangka logika fuzzy. Tahap berikutnya adalah definisi fungsi keanggotaan, di mana setiap variabel input diubah menjadi representasi fuzzy menggunakan fungsi-fungsi seperti segitiga, trapezoid, atau Gaussian. Representasi ini memungkinkan sistem untuk menangani data numerik dalam bentuk linguistik yang lebih mendekati penalaran manusia.

Selanjutnya, dirumuskan seperangkat aturan fuzzy yang merepresentasikan pengetahuan atau logika pengambilan keputusan dalam bentuk aturan IF-THEN. Tahapan inferensi fuzzy, khususnya dengan metode Mamdani, digunakan untuk mengevaluasi aturan-aturan tersebut dan menghasilkan output fuzzy berdasarkan kombinasi masukan dan aturan yang berlaku. Output fuzzy ini kemudian dikonversi kembali ke dalam nilai tegas melalui proses defuzzifikasi, menghasilkan nilai numerik yang dapat digunakan untuk pengambilan keputusan akhir atau hasil rekomendasi. Proses ini secara keseluruhan mencerminkan pendekatan sistematis dalam pengolahan ketidakpastian dan kompleksitas data menggunakan logika fuzzy.

III.5. Alur Random forest



Pipeline tersebut mengilustrasikan tahapan operasional dalam implementasi algoritma Random Forest, suatu metode ensemble learning yang berbasis pohon keputusan. Proses dimulai dari pemuatan dan pra-pemrosesan data, di mana data mentah dibersihkan dan ditransformasikan agar siap digunakan dalam pelatihan model. Ini mencakup penanganan nilai hilang, encoding fitur kategorikal, serta normalisasi jika diperlukan. Setelah data siap, langkah penting berikutnya adalah membagi dataset menjadi subset pelatihan dan pengujian, yang bertujuan untuk memastikan evaluasi performa model dilakukan secara objektif dengan data yang belum pernah dilihat sebelumnya.

Tahapan selanjutnya adalah pelatihan model Random Forest, di mana sejumlah pohon keputusan dibentuk menggunakan subset acak dari data dan fitur. Proses ini memperkuat generalisasi model dengan mengurangi varians, yang merupakan keunggulan utama dari metode ini dibandingkan pohon tunggal. Setelah pelatihan selesai, model digunakan untuk melakukan prediksi terhadap data uji. Tahap akhir dalam pipeline ini adalah evaluasi model, yang melibatkan pengukuran kinerja menggunakan metrik seperti akurasi, precision, recall, F1-score, atau AUC tergantung pada jenis permasalahan. Evaluasi ini merupakan aspek krusial bagi peneliti untuk

menilai efektivitas dan generalisasi model sebelum digunakan dalam konteks dunia nyata atau penelitian lanjutan.

III.6. Analisis Komparatif Performa Logika Fuzzy dan Algoritma Random Forest dalam Klasifikasi Tanaman

Pada evaluasi kinerja kedua model—sistem fuzzy dengan pembaruan fungsi keanggotaan dan model Random Forest—terdapat perbedaan substansial pada akurasi, stabilitas, dan karakteristik penerapannya. Sistem fuzzy, setelah dilakukan pembaruan pada fungsi keanggotaan, menunjukkan akurasi 88,67 % pada himpunan uji yang terdiri atas 300 sampel (masing-masing 100 sampel untuk kelas kelapa, kopi, dan jagung). Pada sisi lain, model Random Forest mencapai akurasi 100 % pada 60 sampel uji yang terbagi menjadi 16 sampel kelapa, 22 sampel kopi, dan 22 sampel jagung.

Hasil klasifikasi sistem fuzzy merefleksikan bahwa dari total 300 sampel, sebanyak 266 sampel diklasifikasikan dengan benar, sementara 34 sampel mengalami kesalahan klasifikasi. Matriks kebingungan menunjukkan bahwa kelas jagung berhasil diprediksi sepenuhnya (100 % benar), sedangkan terdapat tumpang tindih antara kelas kelapa dan kopi, dengan 27 sampel kelapa yang keliru teridentifikasi sebagai kopi dan 7 sampel kopi yang keliru teridentifikasi sebagai kelapa atau jagung. Classification report memperlihatkan nilai precision kelas kelapa sebesar 0,99 dan recall 1,00, nilai precision kelas kopi sebesar 0,78 dan recall 0,93, serta nilai precision kelas jagung sebesar 0,92 dan recall 0,73. Nilai-nilai tersebut menandakan bahwa sistem fuzzy hampir tidak menghasilkan false positive untuk kelas kelapa, tetapi masih terdapat proporsi false negative yang memengaruhi kelas kopi dan jagung.

Secara konseptual, sistem fuzzy unggul dalam menangani ketidakpastian dan tumpang tindih nilai fitur agroklimat melalui penggunaan fungsi keanggotaan linguistik dan aturan inferensi berbasis himpunan fuzzy. Pendekatan ini lebih fleksibel terhadap variasi kondisi agroklimat yang tidak dapat dipisahkan secara kaku, misalnya rentang pH atau kelembaban yang tumpang tindih antara kelapa dan kopi. Namun demikian, ketergantungan pada definisi fungsi keanggotaan yang presisi menyebabkan model ini masih rentan terhadap kesalahan klasifikasi ketika nilai-nilai fitur agroklimat berada pada ambang batas yang berdekatan. Dengan kata lain, sistem fuzzy mampu mengekspresikan gradien keanggotaan secara halus, tetapi memerlukan penyempurnaan lebih lanjut untuk meminimalkan ambiguitas antara kelas yang memiliki karakteristik agroklimat serupa.

Sebaliknya, model Random Forest menunjukkan performa sempurna dengan akurasi 1,00 pada data uji dan skor cross-validation lima lipatan juga sebesar 1,00. Matriks kebingungan yang sepenuhnya diagonal menunjukkan tidak ada sampel yang salah klasifikasi untuk semua tiga

kelas tanaman. Evaluasi five-fold cross-validation pada data latih memberikan rata-rata skor 1,00, yang memperlihatkan stabilitas dan konsistensi model dalam mempelajari pola non-linear pada kombinasi fitur agroklimat. Lebih lanjut, analisis feature importance mengungkapkan bahwa variabel kelembaban (0,2344), nitrogen (0,2234), dan curah hujan berbentuk logaritmik (0,2049) memiliki kontribusi terbesar dalam proses pengambilan keputusan.

Keunggulan utama Random Forest terletak pada kemampuannya membangun boundary keputusan non-linier yang kuat, sehingga mengeliminasi tumpang tindih kelas secara lebih tegas dibandingkan pendekatan fuzzy. Selain itu, hasil cross-validation yang seragam mendukung kesimpulan bahwa model ini mampu menggeneralisasi pola dalam dataset latih secara konsisten. Akan tetapi, akurasi sempurna pada data uji dan hasil cross-validation yang identik juga menimbulkan kekhawatiran potensi kurangnya data, terutama karena dataset uji relatif kecil (60 sampel) dan terbagi seimbang. Tanpa evaluasi tambahan terhadap dataset eksternal yang memiliki distribusi fitur agroklimat berbeda, sulit memastikan bahwa model Random Forest tersebut akan mempertahankan kinerja sempurna saat dihadapkan pada data baru.

Berdasarkan perbandingan di atas, pemilihan model sangat bergantung pada konteks aplikasi dan karakteristik data yang digunakan. Apabila prioritas utama adalah akurasi klasifikasi pada dataset yang relatif seragam dan seimbang, serta terdapat jaminan bahwa distribusi fitur agroklimat pada data produksi tidak jauh berbeda dari data pengujian, maka Random Forest merupakan pilihan yang paling cocok karena kemampuannya memisahkan kelas secara deterministik tanpa kesalahan. Sebaliknya, apabila dataset yang akan digunakan mencakup tentang kondisi agroklimat yang lebih beragam dan memerlukan penanganan ketidakpastian nilai fitur secara eksplisit, maka sistem fuzzy lebih sesuai karena mampu mengekspresikan derajat keanggotaan secara halus meski dengan akurasi yang masih di bawah 90 %. Untuk kasus “Crops Recommendation” berbasis logika fuzzy Mamdani, di mana fitur agroklimat seringkali saling tumpang tindih dan memerlukan interpretasi aturan berbasis linguistik, sistem fuzzy akan lebih adaptif terhadap variasi kondisi lapangan. Namun jika validasi eksternal dan keseragaman dataset dapat dipastikan, Random Forest akan memberikan hasil klasifikasi yang lebih andal.

BAB IV

PENUTUP

Penelitian ini telah berhasil membangun dan mengevaluasi sistem rekomendasi tanaman berbasis logika fuzzy Mamdani dan algoritma Random Forest dengan menggunakan dataset agroklimat dari Kaggle. Melalui tahapan pra-pemrosesan data, perancangan fungsi keanggotaan, formulasi fuzzy rules, hingga implementasi model machine learning, sistem ini mampu memberikan rekomendasi tanaman (maize, coffee, coconut) yang sesuai dengan kondisi lingkungan tertentu.

Hasil pengujian menunjukkan bahwa model Random Forest memiliki tingkat akurasi sempurna sebesar 100% dalam skenario pengujian terbatas, sedangkan sistem logika fuzzy Mamdani mencapai akurasi 88,67% dengan performa yang sangat baik pada kelas tertentu seperti coconut. Analisis komparatif menunjukkan bahwa Random Forest unggul dari segi akurasi dan kestabilan prediksi, sementara logika fuzzy unggul dari segi interpretabilitas dan kemampuan menangani ketidakpastian.

Dengan demikian, pendekatan gabungan atau hybrid antara kedua metode dapat menjadi arah pengembangan yang potensial untuk implementasi sistem rekomendasi tanaman yang tidak hanya akurat, tetapi juga transparan dan mudah dipahami oleh pengguna akhir seperti petani. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan pertanian presisi di Indonesia, serta menjadi referensi dalam penerapan teknologi kecerdasan buatan di bidang agrikultur.

DAFTAR PUSTAKA

- [1] I. W. Widayat, A. A. Arsyad, A. J. Mantau, Y. Adhitya, dan M. Köppen, “Fuzzy Methods in Smart Farming: A Systematic Review,” *Informatica (Vilnius)*, vol. 35, no. 1, pp. 106–113, 2024.
- [2] G. Banerjee, U. Sarkar, dan I. Ghosh, “A Fuzzy Logic-Based Crop Recommendation System,” dalam *Proc. Int. Conf. Frontiers in Computing and Systems (COMSYS2020)*, 2020, pp. 57–69.

- [3] K. Borse dan P. G. Agnihotri, "Prediction of Crop Yields Based on Fuzzy Rule-Based System (FRBS) Using the Takagi Sugeno-Kang Approach," dalam *Proc. Int. Conf. Intelligent Computing and Optimization (ICICO)*, 2023.
- [4] T. P. Ezhilarasi dan K. S. Rekha, "Crop Recommendation System for Precision Agriculture Using Fuzzy Clustering Based Ant Colony Optimization," dalam *Lecture Notes in Electrical Engineering*, vol. 925, pp. 261–274, 2022.
- [5] G. Singh dan S. Sharma, "Enhancing Precision Agriculture through Cloud-Based Transformative Crop Recommendation Model," *Scientific Reports*, vol. 15, art. no. 8560, 2025.
- [6] H. Afzal, M. Amjad, A. Raza, dkk., "Incorporating Soil Information with Machine Learning for Crop Recommendation to Improve Agricultural Output," *Scientific Reports*, vol. 15, art. no. 8560, 2025.
- [7] S. R. Bogireddy dan H. Murari, "Enhancing Crop Yield Prediction through Random Forest Classifier: A Comprehensive Approach," dalam *Proc. IEEE Int. Conf. Smart Electronics and Communication (ICOSEC)*, 2024.
- [8] R. P. Sharma, D. Ramesh, P. Pal, S. Tripathi, dan C. Kumar, "IoT-Enabled IEEE 802.15.4 WSN Monitoring Infrastructure-Driven Fuzzy-Logic-Based Crop Pest Prediction," *IEEE Internet of Things Journal*, Feb. 2022.
- [9] R. Aarthi dan D. Sivakumar, "Modeling the Hierarchical Fuzzy System for Suitable Crop Recommendation," dipublikasikan pada 2020.
- [10] A. Burhanuddin, "Analisis Komparatif Inferensi Fuzzy Tsukamoto, Mamdani dan Sugeno Terhadap Produktivitas Padi di Indonesia," 2023.
- [11] C. O. Martinez-Ojeda, T. M. Amado, dan J. C. Dela Cruz, "In Field Proximal Soil Sensing for Real Time Crop Recommendation Using Fuzzy Logic Model," dalam *Proc. 2019 Int. Symp. on Multimedia and Communication Technology (ISMATC)*, pp. 1–6, Aug. 2019.
- [12] M. K. Sinha dan R. K. Tiwary, "Utilizing Fuzzy Logic in Precision Agriculture: Techniques for Disease Detection and Management," *Journal of Statistics and Mathematical Engineering*, 2024.
- [13] M. Addisu, A. Jifar, H. Tessema, T. Alemu, dan A. Tadesse, "Exploitation of Ethiopian Coffee Land-race Potentials: Market Demand Oriented Specialty Variety Development," *Science Research*, vol. 12, 2024.

- [14] D. Senaratne dan N. Senaratne, "Modelling Agro-Forestry Systems for Improved Economic Performance: A Comparative Economic Analysis," 2013.
- [15] R. Margate, J. Maravilla, R. M. Ebuña, dan M. N. Eroy, "Response of Coconut and Coffee to Fertilizers Applied on Either or Both Crops in an Intercropping System," *CORD*, vol. 10, pp. 34, 1994.
- [16] S. Braconnier, "Maize-Coconut Intercropping: Effect of Radiation and Root Competition on Maize," 1998.
- [17] R. K. Dewi, I. Suliansyah, A. Anwar, dan A. Syarif, "Yield Test of Shade-Tolerant Maize Using the Intercropping Method of Maize-Soybean Under Coconut Trees," *International Journal of Science and Society*, 2024.
- [18] K. Hebbar, D. Balasimha, dan G. Thomas, "Plantation Crops Response to Climate Change: Coconut Perspective," 2013.
- [19] A. Melke dan M. Fetene, "Eco-physiological Basis of Drought Stress in Coffee (*Coffea arabica*, L.) in Ethiopia," *Theoretical and Experimental Plant Physiology*, vol. 26, pp. 225–239, 2014.