

Tubes DKA

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

Oleh :


SAMUEL YOHANES MICHAEL K. - 103012300247

FADHILA M. WIRYASUKMA - 103012300303



Pendahuluan

Pertanian sangat penting bagi ketahanan pangan dan ekonomi nasional. Keberhasilan budidaya tanaman tergantung pada kondisi agroklimat seperti nutrisi tanah, suhu, kelembapan, pH, dan curah hujan yang bervariasi antar lokasi dan musim, sehingga pemilihan tanaman yang tepat menjadi tantangan utama. Pertanian presisi memanfaatkan teknologi dan data, termasuk sensor dan AI, untuk meningkatkan akurasi rekomendasi tanaman berdasarkan kondisi lingkungan real-time. Namun, data agroklimat yang tidak pasti dan ambigu sulit diolah dengan logika konvensional, sehingga logika fuzzy digunakan untuk memodelkan nilai “abu-abu” dengan metode Mamdani dan Sugeno. Penelitian ini mengembangkan sistem rekomendasi tanaman berbasis logika fuzzy Mamdani dan Random Forest, yang terbukti memberikan rekomendasi lebih akurat dan adaptif guna mendukung pertanian presisi.



Rumusan Masalah

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?

Tujuan

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.

Sumber dari dataset yang digunakan

Penelitian ini menggunakan dataset [Crop Recommendation](#) dari Kaggle yang berisi 2.200 data dengan 22 jenis tanaman. Dataset memiliki tujuh variabel input yang menggambarkan kondisi agroklimat untuk menentukan tanaman yang sesuai. Fokus penelitian adalah pada tiga tanaman: jagung, kopi, dan kelapa.

Variabel input tersebut meliputi:

- 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)

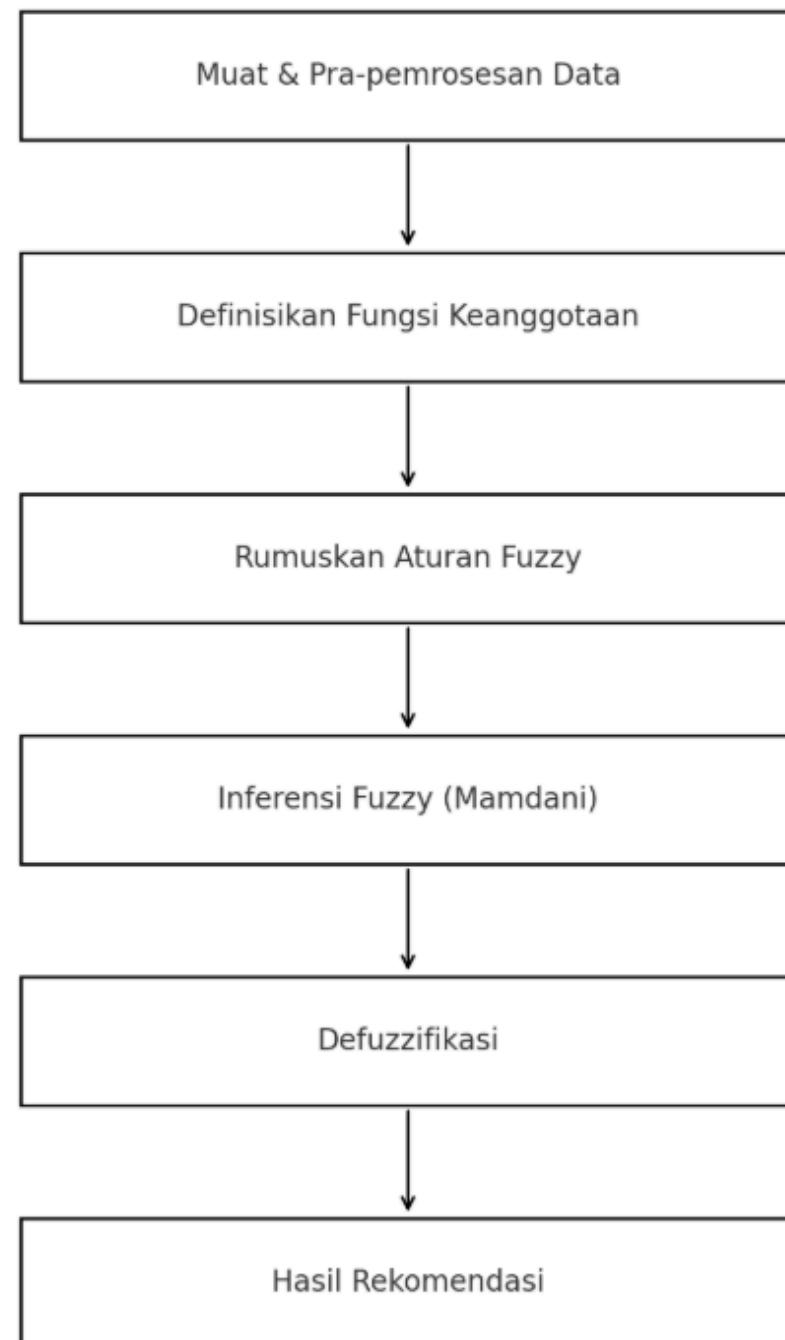
Paparan statistik dari dataset

Feature	Mean	Median	Std Dev	Min	Max	Skewness	Kurtosis
N	505.518	370.000	369.173	0	1.400.000	5.094	-10.586
P	533.627	510.000	329.859	50.000	1.450.000	10.101	8.556
K	481.491	320.000	506.479	50.000	2.050.000	23.735	44.365
temperature	256.162	255.987	50.637	88.257	436.755	1.848	12.270
humidity	714.818	804.731	222.638	142.580	999.819	-10.910	2.987
ph	64.695	64.250	7.739	35.048	99.351	2.837	16.491
rainfall	1.034.637	948.676	549.584	202.113	2.985.601	9.651	6.030

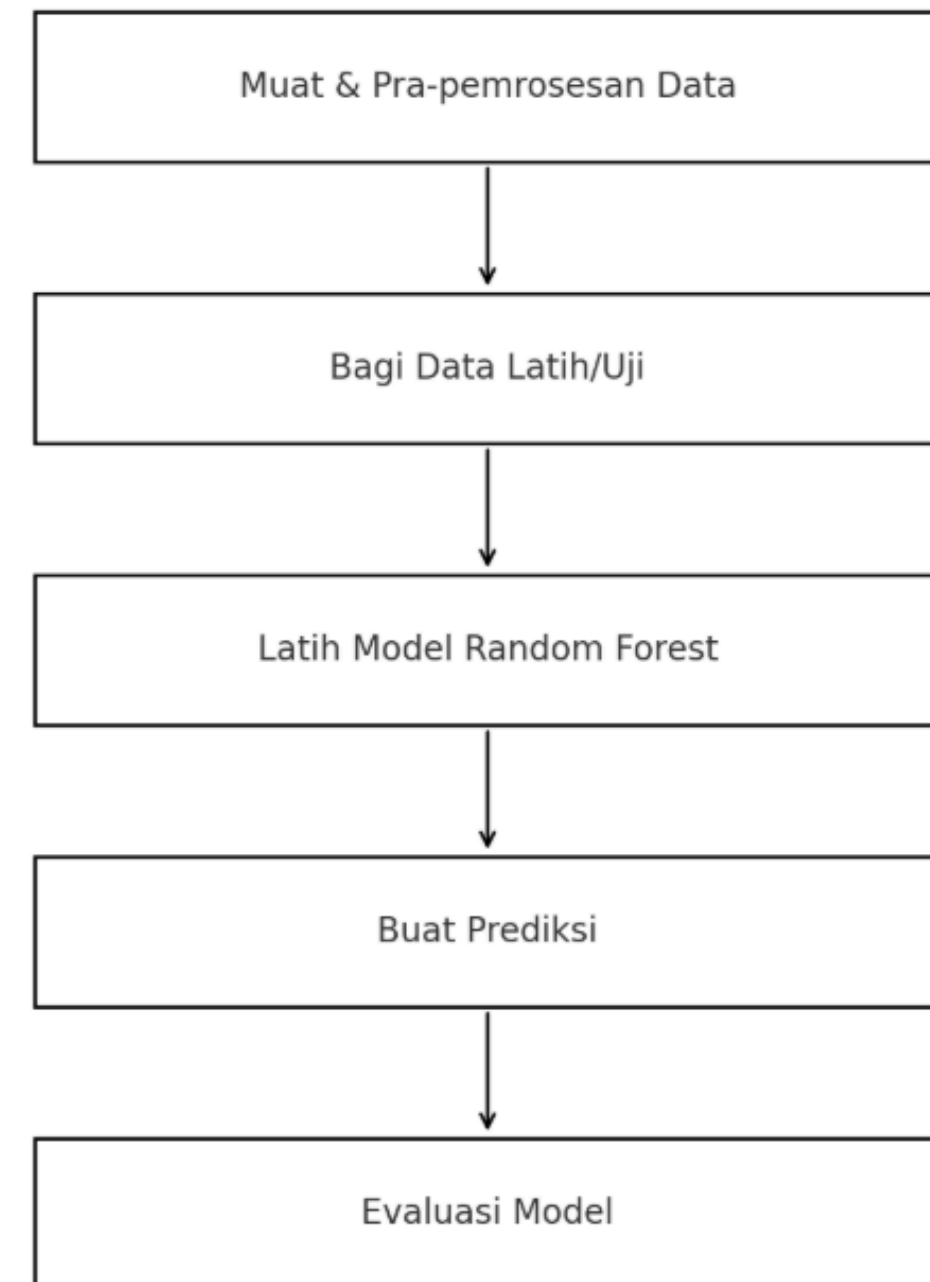
Tabel statistik deskriptif menunjukkan nilai minimum, maksimum, rata-rata, median, standar deviasi, skewness, dan kurtosis dari setiap variabel input, yang memberikan gambaran karakteristik dan distribusi data yang digunakan dalam sistem rekomendasi tanaman.

Pipeline Fuzzy dan Random Forrest

Pipeline Logika Fuzzy



Pipeline Random Forest



Pre-processing Dataset

1. Mengatasi missing value

Kode untuk mendeteksi missing value :

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

Kode untuk mengisi missing value :

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

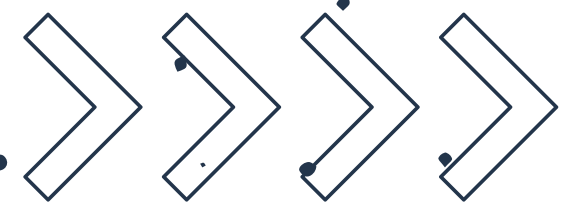
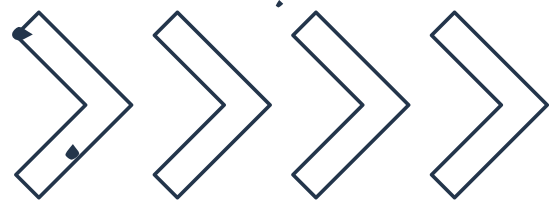
2. Deteksi & Drop Outlier

Untuk mendeteksi keberadaan outlier pada dataset, digunakan metode Interquartile ($1.5 \times \text{IQR}$)

Didapat Hasil :

- *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*

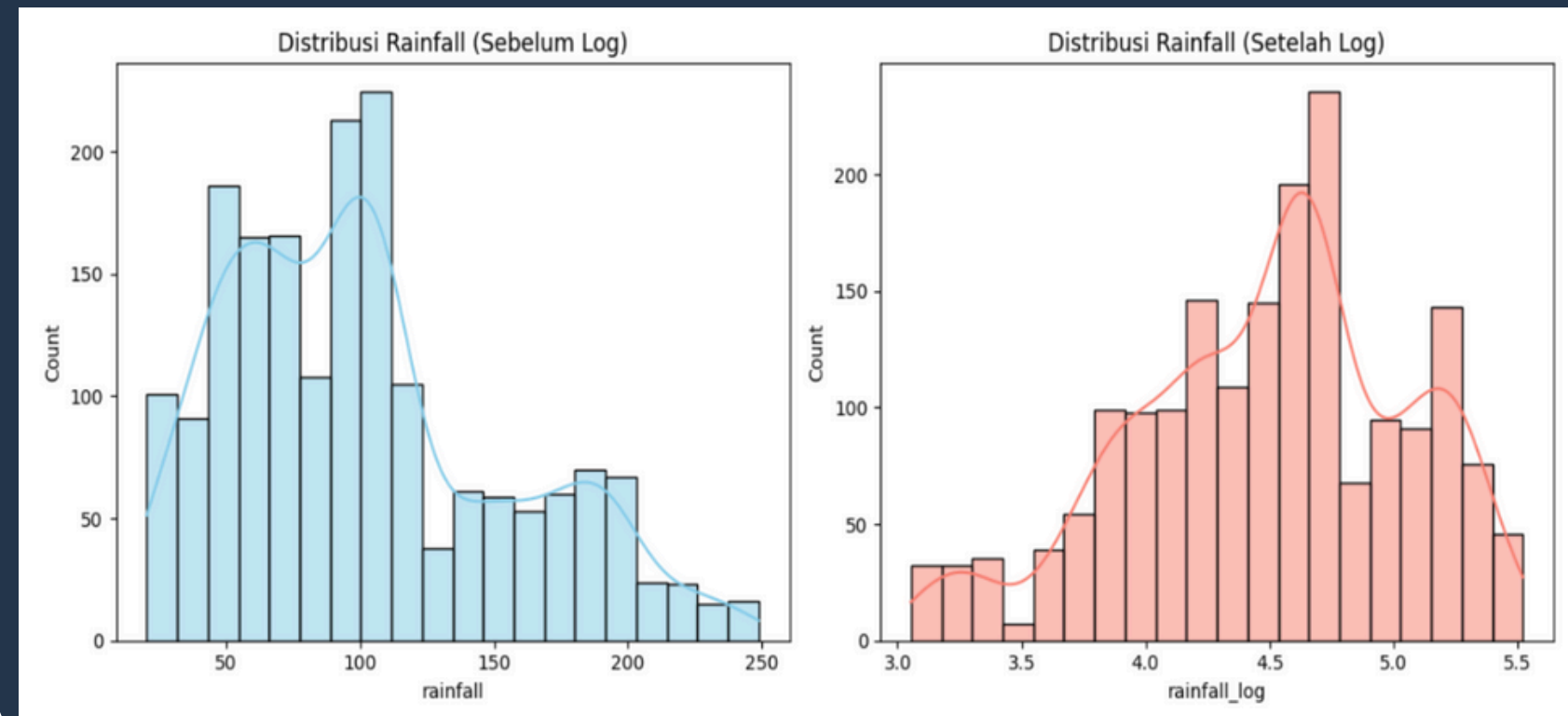
Data setelah drop outlier, shape sekarang: (1846, 8)



Pre-processing Dataset

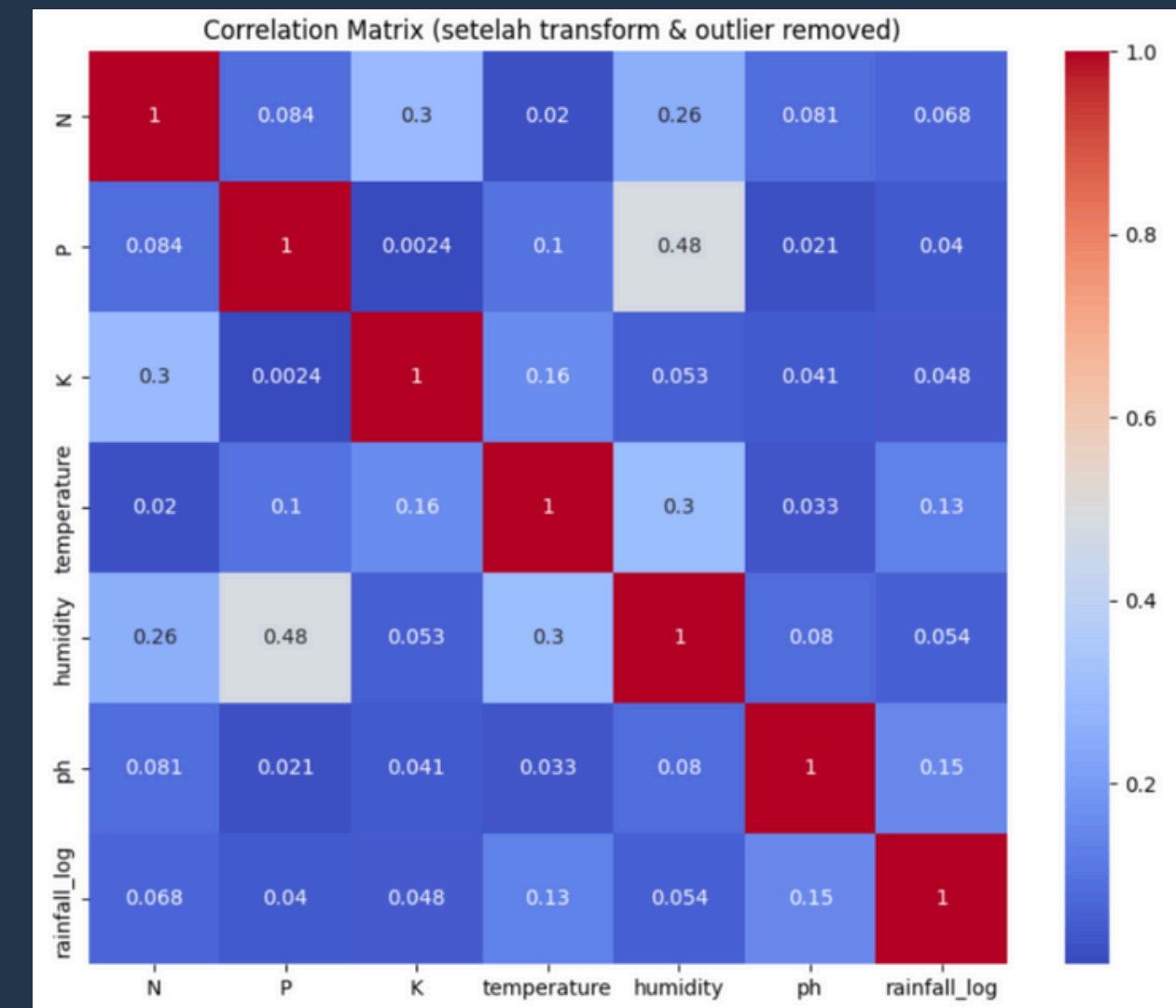
3. Mengatasi skewness pada fitur Rainfall

Transformasi dilakukan menggunakan fungsi `np.log1p(x)`, yaitu logaritma natural dari nilai $x + 1$.



4. Analisis Hubungan Antar Fitur dan Seleksi Variabel

Analisis ini dilakukan dengan menghitung korelasi Pearson antar fitur numerik (selain label) menggunakan fungsi `corr()` dari library `pandas`.



Pre-processing Dataset

5. Normalisasi Data dengan Min-Max Scaling

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:

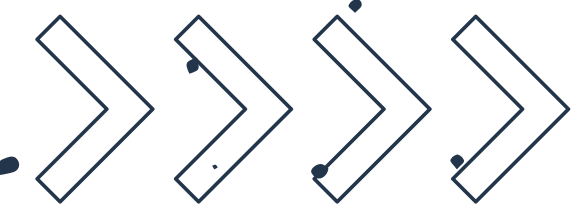
$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Hasil :

```
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
```

6. Memilih label

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.



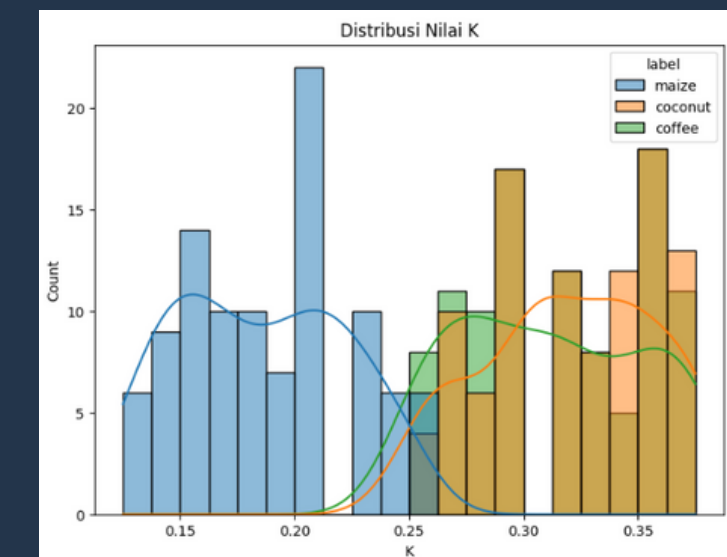
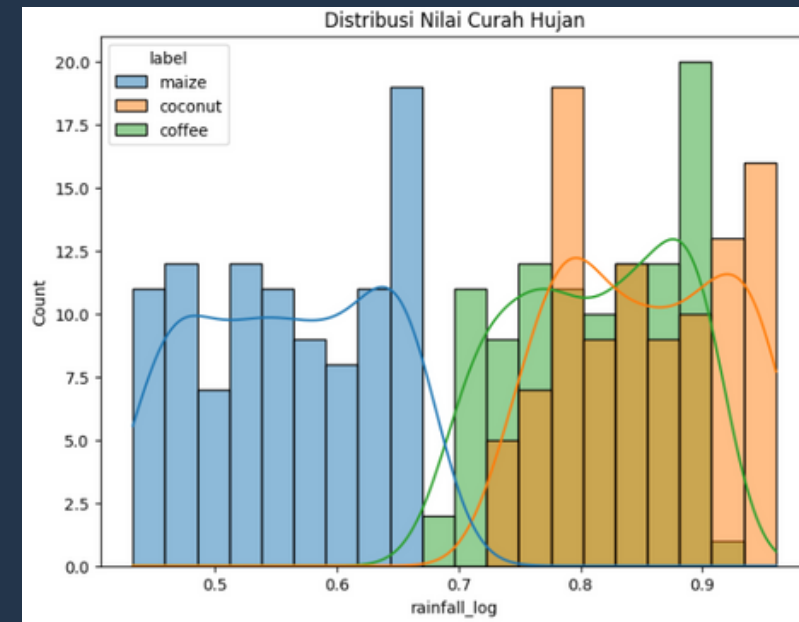
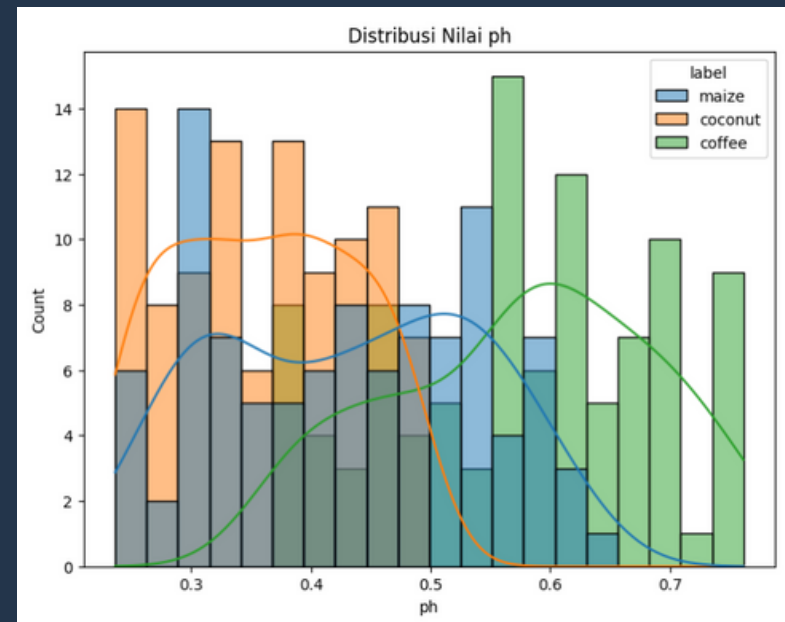
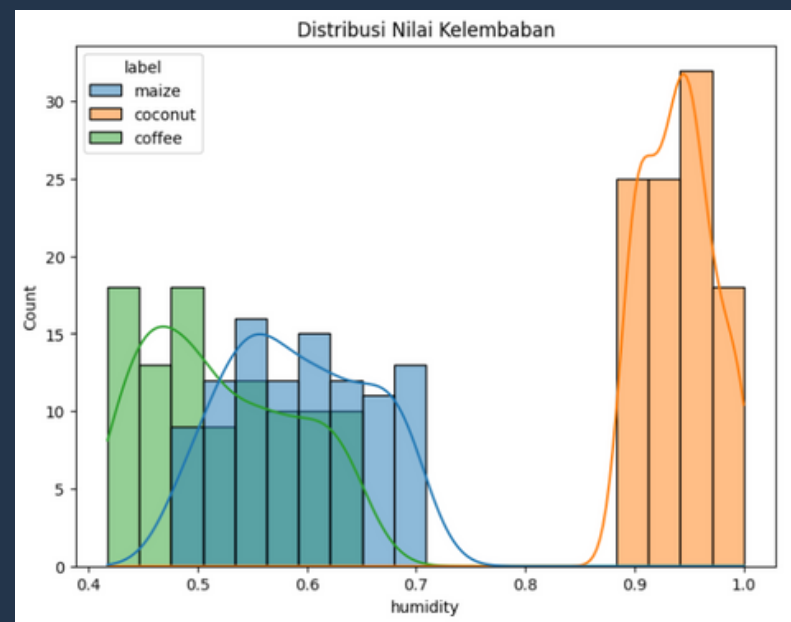
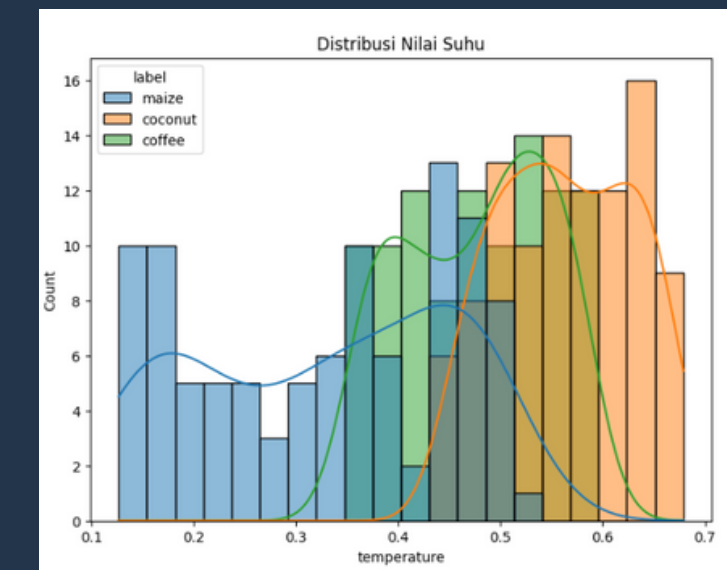
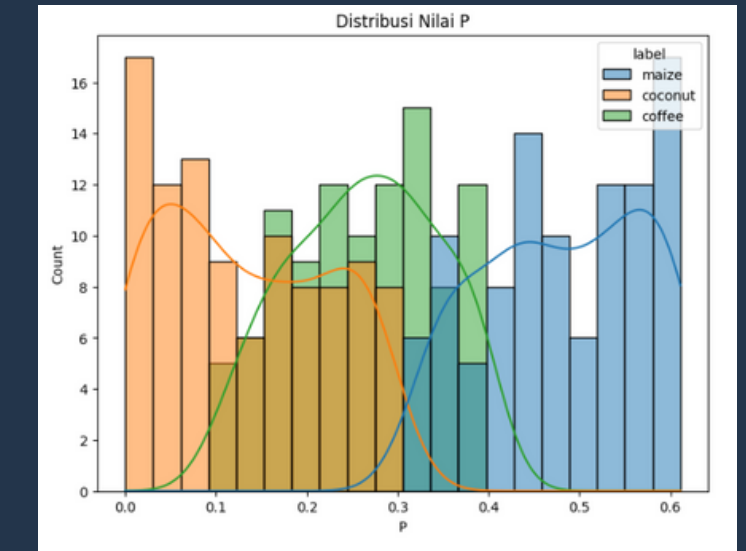
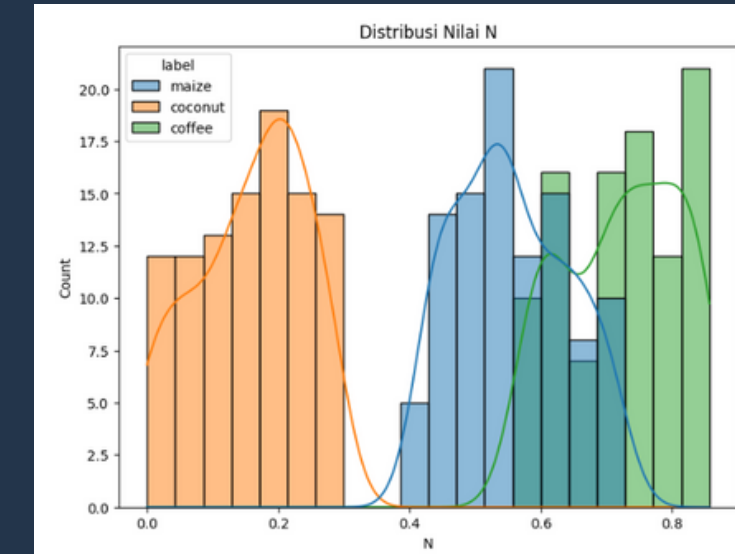
Pre-processing Dataset

7. Eksplorasi data analisis

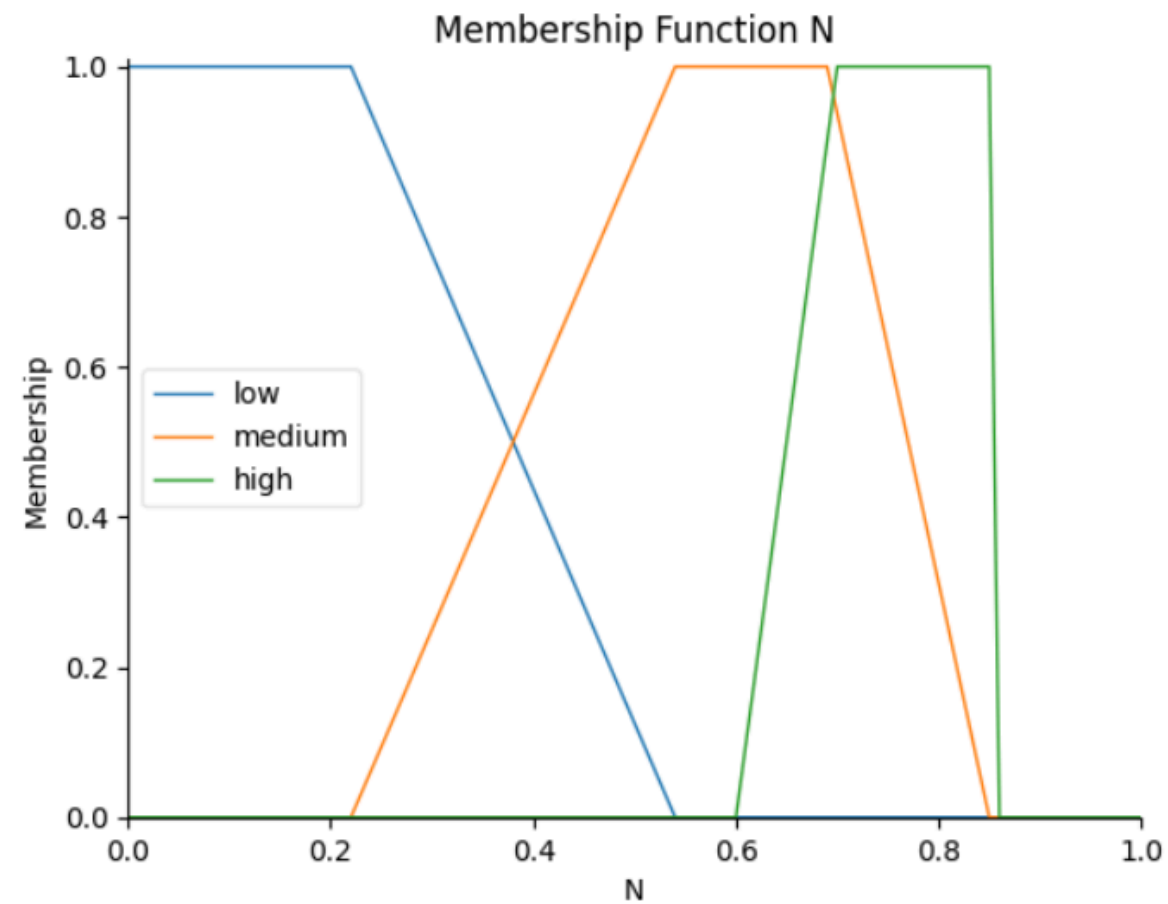
Statistik Deskriptif Setelah Pre-Processing :

	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

Distribusi Setiap fitur pada Tanaman yang dianalisis :



Rancangan Membership Function

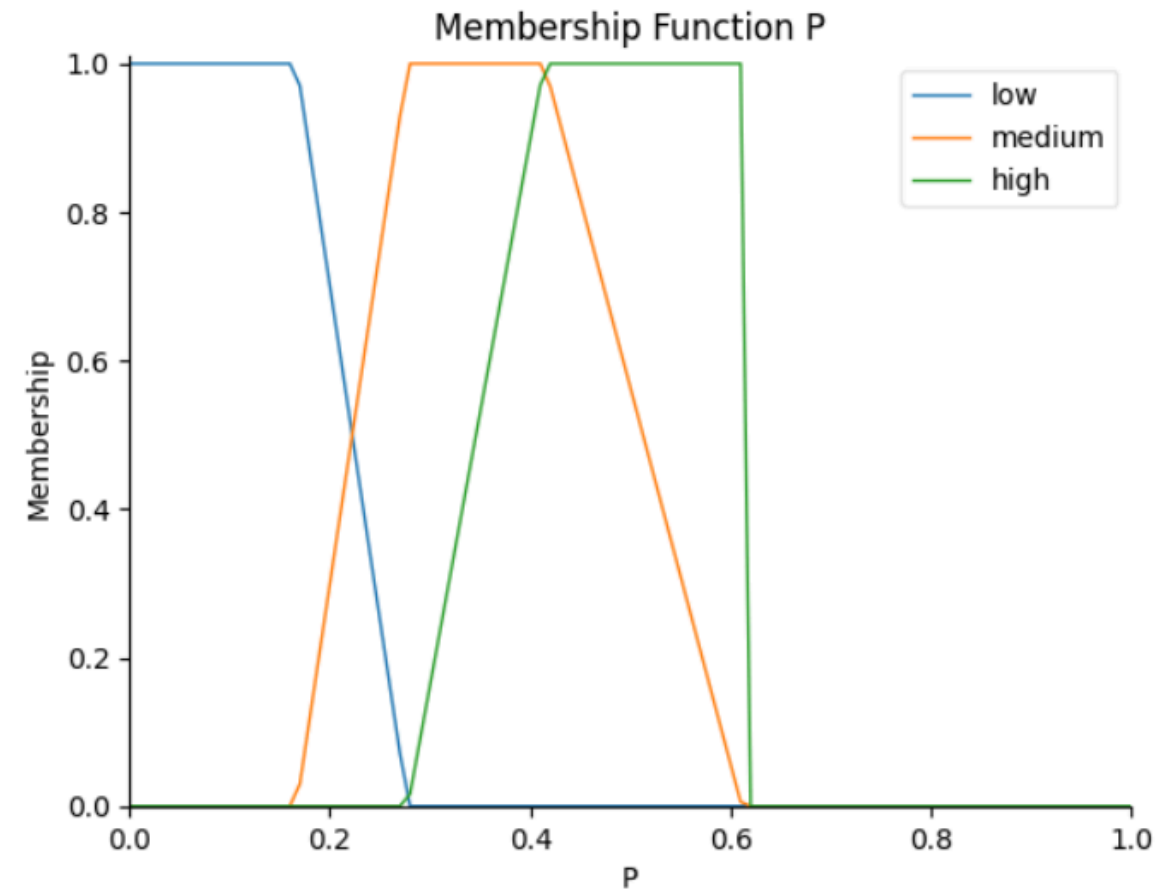


NITROGEN (N) - Trapezoid Functions:

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]



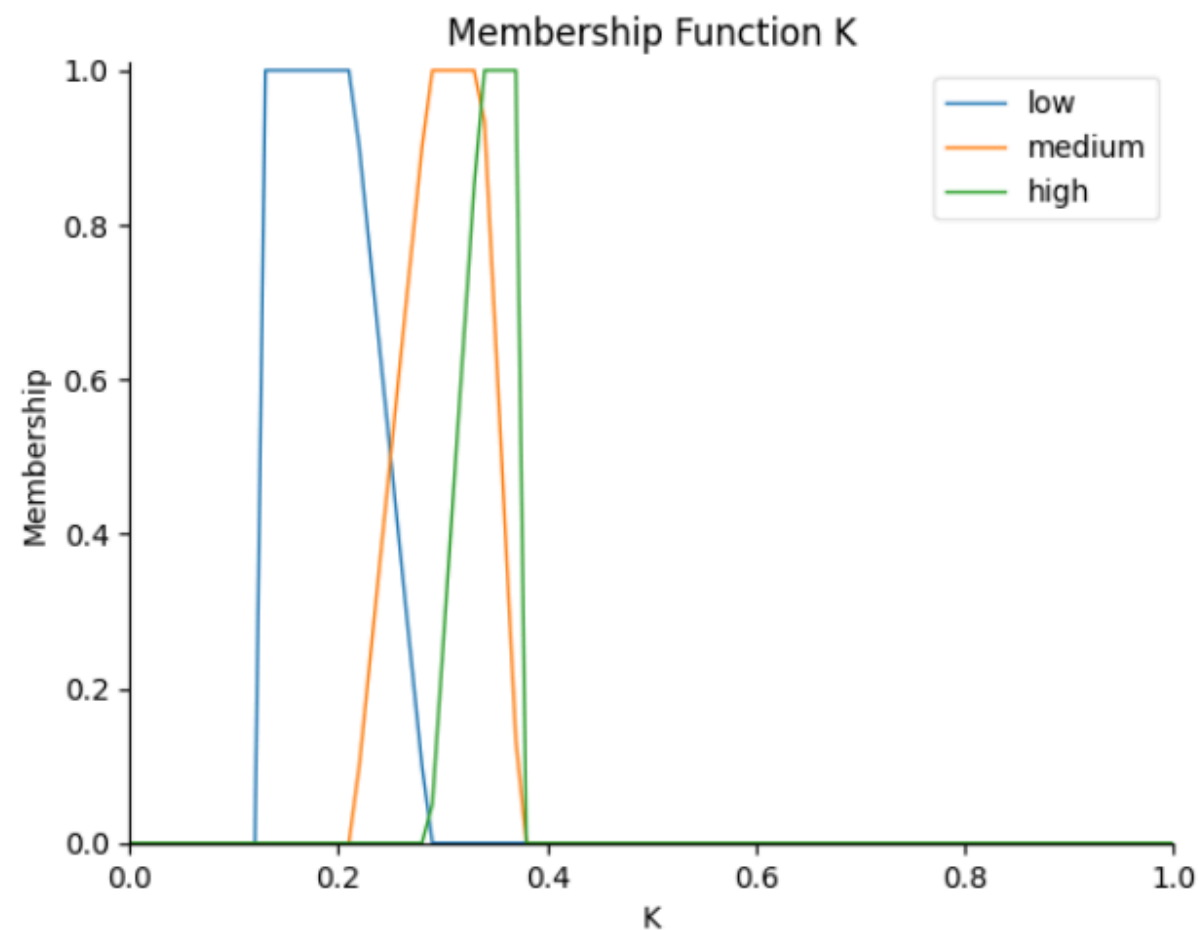
PHOSPHORUS (P) - Trapezoid Functions:

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]

Rancangan Metode Fuzzy

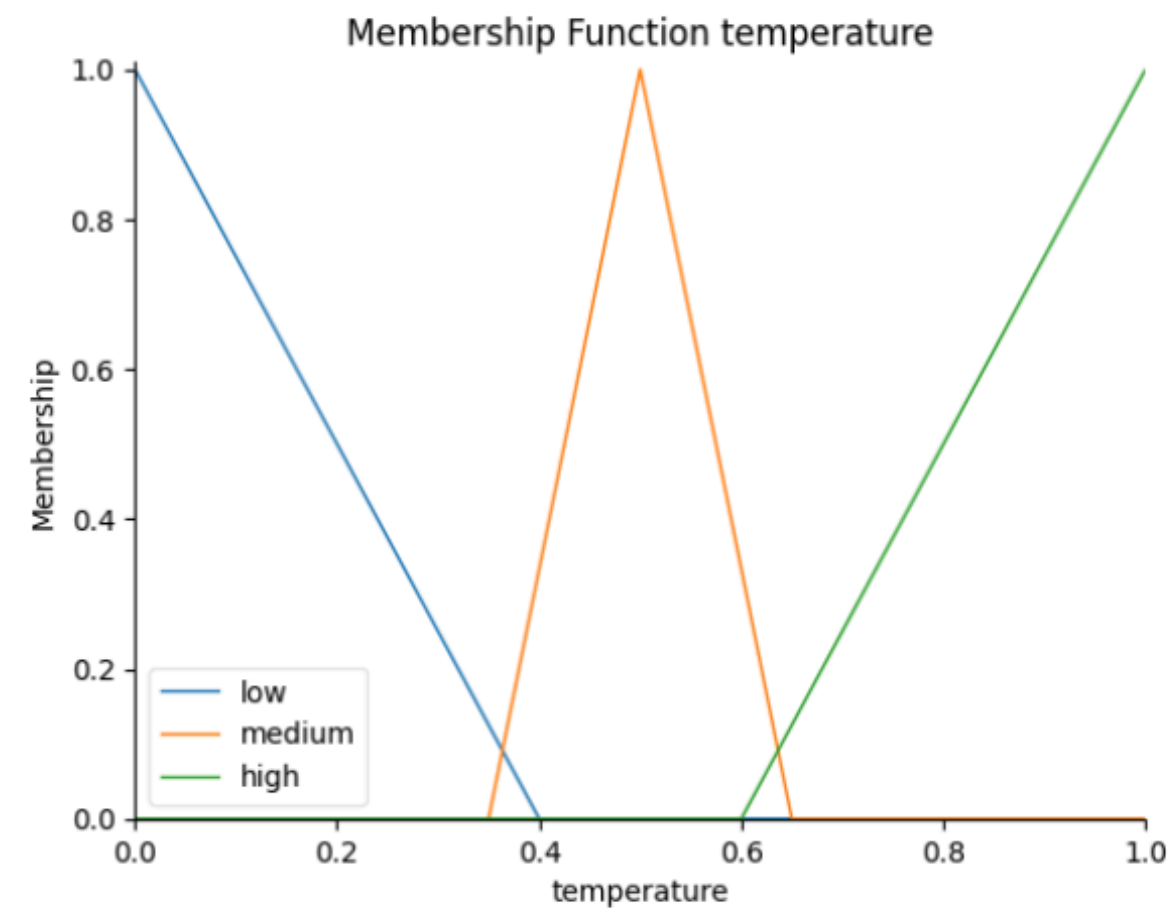


POTASSIUM (K) - Trapezoid Functions:

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]



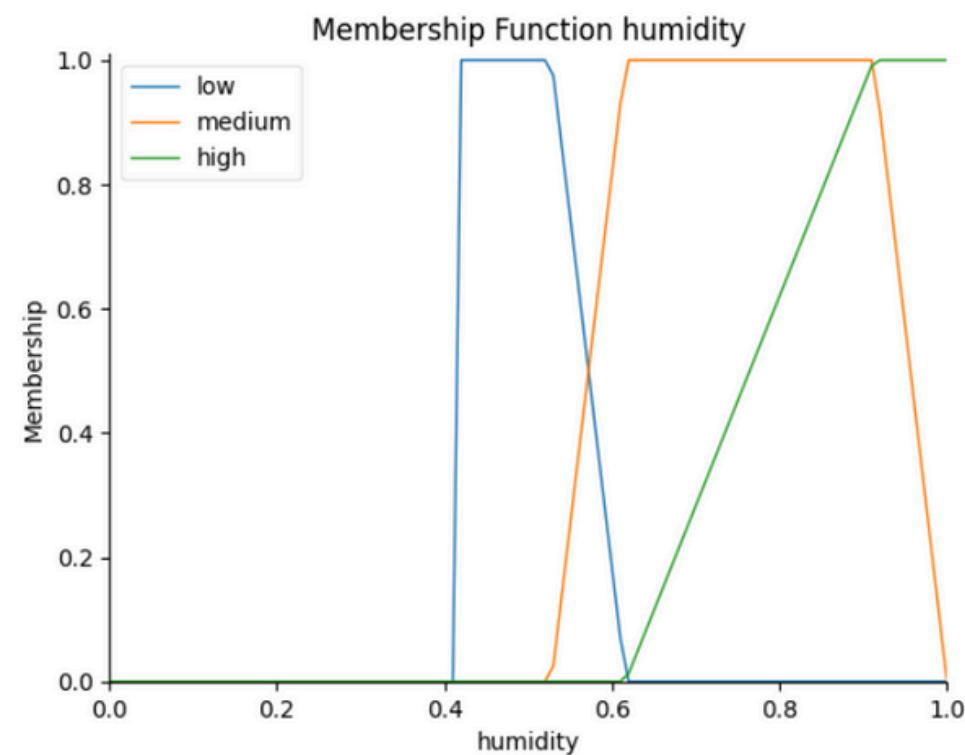
TEMPERATURE - Triangle Functions:

Low: [0.00, 0.00, 0.40]

Medium: [0.35, 0.50, 0.65]

High: [0.60, 1.00, 1.00]

Rancangan Metode Fuzzy

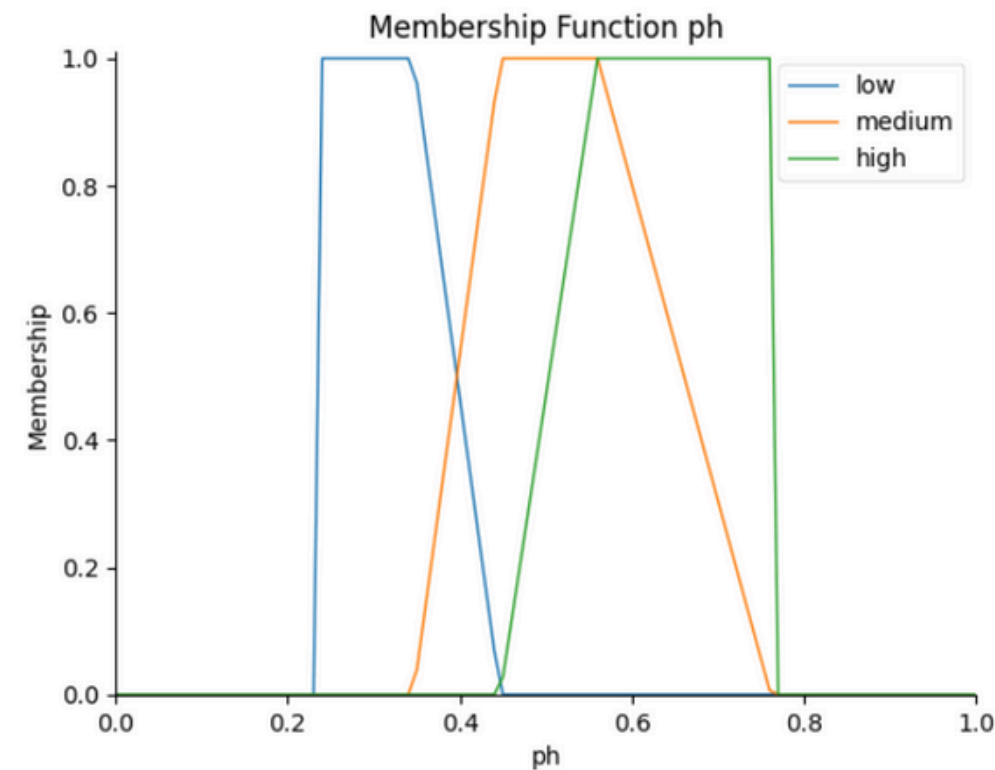


HUMIDITY - Trapezoid Functions:

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]

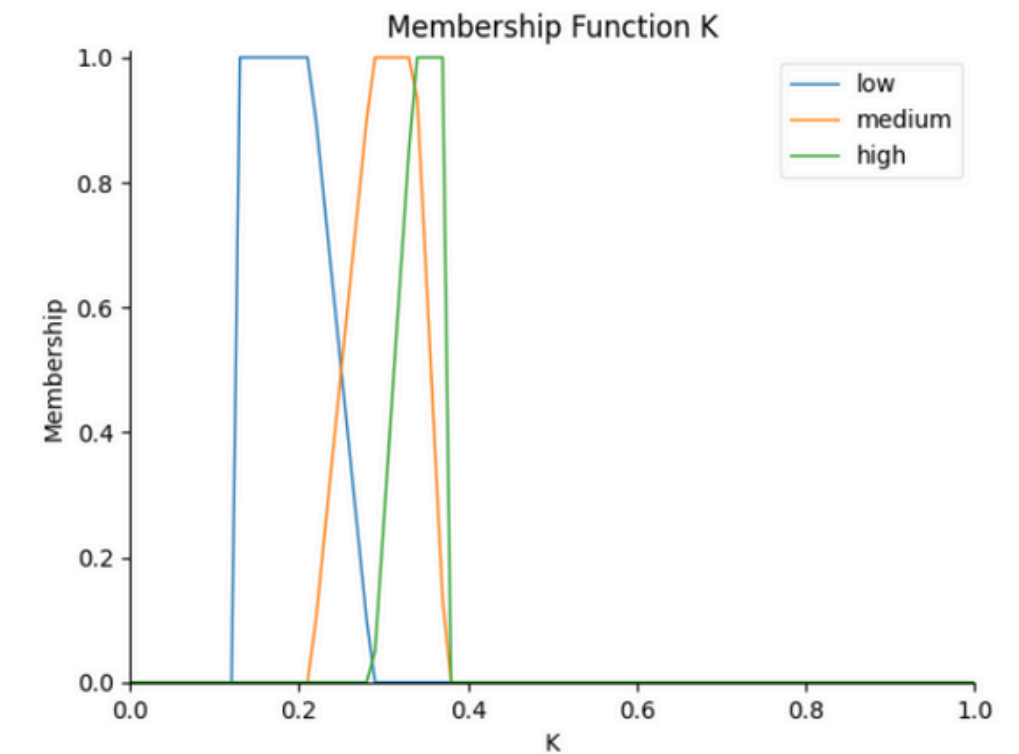


pH - Trapezoid Functions:

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]

Rancangan Metode Fuzzy

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]

MAIZE RULES

- 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

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]

COFFEE RULES

- 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

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]

COCONUT RULES

- 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

Hasil dan Analisis

ALGORITMA FUZZY

✓ 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

Confusion Matrix:

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

Akurasi 88.67% menunjukkan bahwa sistem fuzzy mampu memberikan prediksi yang benar pada 88.67% dari total kasus yang diuji. Ini adalah tingkat akurasi yang sangat baik untuk sistem rekomendasi berbasis fuzzy logic, terutama dalam domain pertanian dimana variabilitas kondisi lingkungan sangat tinggi. Angka ini mengindikasikan bahwa dari setiap 100 prediksi yang dibuat sistem, sekitar 89 prediksi akan sesuai dengan hasil yang diharapkan atau kondisi aktual di lapangan.

Hasil dan Analisis

ALGORITMA RANDOM FORREST

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

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.



Perbandingan

Metrik	Sistem Fuzzy	Random Forest
Akurasi	0,8867 (266/300)	1,00 (60/60)
Presisi (Coconut)	0,99	1
Recall (Coconut)	1	1
F1-Score (Coconut)	1	1
Presisi (Coffee)	0,78	1
Recall (Coffee)	0,93	1
F1-Score (Coffee)	0,85	1
Presisi (Maize)	0,92	1
Recall (Maize)	0,73	1
F1-Score (Maize)	0,82	1
Macro Avg (P / R / F1)	0,90 / 0,89 / 0,89	1,00 / 1,00 / 1,00
Weighted Avg (P / R / F1)	0,90 / 0,89 / 0,89	1,00 / 1,00 / 1,00

Kesimpulan

Random Forest mencapai akurasi 100 % pada 60 sampel, sedangkan Fuzzy Logic mencapai 88,67 % pada 300 sampel. Random Forest unggul dalam menangani pola non-linear, robust terhadap outlier, tetapi berisiko kurang interpretatif. Fuzzy Logic mampu menangani ketidakpastian dan lebih mudah dijelaskan dalam konteks pertanian, namun memerlukan tuning ekstensif dan kurang optimal untuk pola non-linear kompleks. Secara keseluruhan, Random Forest lebih cocok untuk data agroklimat numerik, sementara hybrid Random Forest-Fuzzy Logic dapat menggabungkan akurasi dan interpretabilitas.



Terima Kasih

