

Klasifikasi Cuaca Menggunakan Algoritma Fuzzy Mamdani dan CART

I Dewa Ayu Agung Rai Ratna Karang^{a1}, Ida Ayu Gde Suwiprabayanti Putra^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana, Bali
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹dewayu.ratna0824@gmail.com
²iagsuwiprabayantiputra@unud.ac.id

Abstract

Weather prediction plays a vital role in sectors such as agriculture, transportation, and disaster mitigation. Extreme weather conditions can lead to unpredictable deviations that may cause significant harm to society. This study aims to predict weather conditions using the Fuzzy Mamdani algorithm and CART (Classification and Regression Tree). A total of 1,461 daily historical weather records from Seattle, United States, were obtained from the Kaggle website "Weather Prediction." The fuzzy system was applied to convert numerical weather parameters—precipitation, maximum temperature, minimum temperature, and wind speed—into representative scores based on expert-defined rules. These fuzzy scores were then used as additional features in training the CART model to enhance weather classification accuracy. The dataset was split into 80% training data (1,168 records) and 20% testing data (293 records). Evaluation results show that the integration of Fuzzy and CART achieved an accuracy of 81.23% on the testing set, with high precision, recall, and F1 score for dominant categories such as sun and rain. This study demonstrates that the combination of fuzzy logic and decision trees is effective for weather classification based on historical data.

Keywords: CART, Decision Tree, Fuzzy Logic, Fuzzy Mamdani, Historical Data, Weather Classification, Weather Prediction

1. Pendahuluan

Cuaca merupakan kondisi fisik atmosfer yang berkaitan dengan kondisi di bumi. Cuaca memiliki peran yang penting dalam keseharian makhluk hidup di bumi. Perubahan cuaca yang ekstrem dapat menyebabkan kerugian bagi masyarakat. Cuaca yang cenderung berubah-ubah sehingga memungkinkan terjadinya penyimpangan terhadap suatu hal yang tidak dapat dihindari. Penyimpangan tersebut dapat dilihat dari peristiwa turunnya hujan yang terus menerus selama beberapa hari yang dapat menimbulkan bencana, seperti bencana banjir [3]. Akibatnya, banyak masyarakat yang dirugikan karena bencana banjir tersebut. Hal ini dapat menyebabkan bencana banjir dapat melumpuhkan berbagai kegiatan, seperti kegiatan perekonomian, terhambatnya transportasi, merusak infrastruktur kota, dan menimbulkan dampak negatif lainnya [3]. Curah hujan mempunyai dampak yang signifikan terhadap pertanian, hidrologi, iklim, dan kesejahteraan manusia [9]. Keadaan cuaca dipengaruhi oleh beberapa unsur cuaca yang saling berkaitan satu dengan yang lain yaitu suhu udara, kelembaban udara, kecepatan angin, tekanan udara dan intensitas penyinaran matahari [10].

Pada penelitian [4] Algoritma *Fuzzy Mamdani* digunakan untuk memprediksi cuaca di wilayah kerja BMKG Lansiana. Parameter *input* cuaca yang digunakan ada tiga yaitu suhu, kelembapan, dan tekanan udara, dengan label *output* cuaca ada 4 yaitu tidak hujan, hujan ringan, hujan sedang, dan hujan lebat. Berdasarkan hasilnya dapat disimpulkan bahwa penerapan logika *Fuzzy* untuk sistem peramalan cuaca dapat melakukan klasifikasi cuaca dengan baik dengan tingkat akurasi sistem 61,062 % untuk data cuaca per hari pada tahun 2013-2017. Tingkat akurasi terendah terdapat pada tahun 2015 sebesar 54,26% dan tingkat akurasi tertinggi pada tahun 2017 sebesar 65,20%.

Pada penelitian [5] menggabungkan algoritma *C4.5* dengan *Fuzzy Sugeno* untuk mengoptimalkan rule base dalam sistem *Fuzzy*. Sistem ini menggunakan tiga parameter *input* yaitu suhu, kelembapan, dan curah hujan dengan lima label *output* cerah/berawan, hujan ringan, hujan sedang, hujan lebat, dan hujan sangat lebat. Hasil dari penelitian ini menunjukkan bahwa penggabungan algoritma *C4.5* dengan *Fuzzy Sugeno* dapat meningkatkan efisiensi proses klasifikasi *Fuzzy*. Beberapa cara yang telah dilakukan pada penelitian sebelumnya untuk mengurangi beban komputasi, diantaranya dengan mengurangi sejumlah aturan dalam logika *Fuzzy*. Mengurangi sejumlah aturan akan berdampak pada tingkat akurasi *Fuzzy* yang berkurang. Hasil perbandingan algoritma *Fuzzy Sugeno* yang diintegrasikan dengan algoritma *C4.5* mendapatkan hasil akurasi sebesar 88,57 %. Pengurangan jumlah aturan yang awalnya 288 rule menjadi 57 rule menyebabkan beban komputasi berkurang, hal tersebut berdampak pada berkurangnya tingkat akurasi.

Pada penelitian [9] menggunakan penerapan metode logika *Fuzzy Mamdani* yang dapat digunakan untuk meramalkan curah hujan di Kota Kendari. Sistem ini menggunakan tiga parameter *input* yaitu suhu, kelembapan, dan tekanan udara. Label *output* yaitu tidak hujan, hujan ringan, hujan sedang, dan hujan lebat. Sistem diuji menggunakan data historis BMKG dengan akurasi yang cukup baik, yaitu sebesar 89,30%. Aturan *Fuzzy* yang paling akurat adalah aturan *Fuzzy* ke-14 yaitu pada suhu hangat, kelembapan lembab, dan tekanan normal. Aturan *Fuzzy* yang paling tidak akurat adalah aturan *Fuzzy* ke-22 yaitu pada suhu panas, kelembapan lembab, dan tekanan ringan.

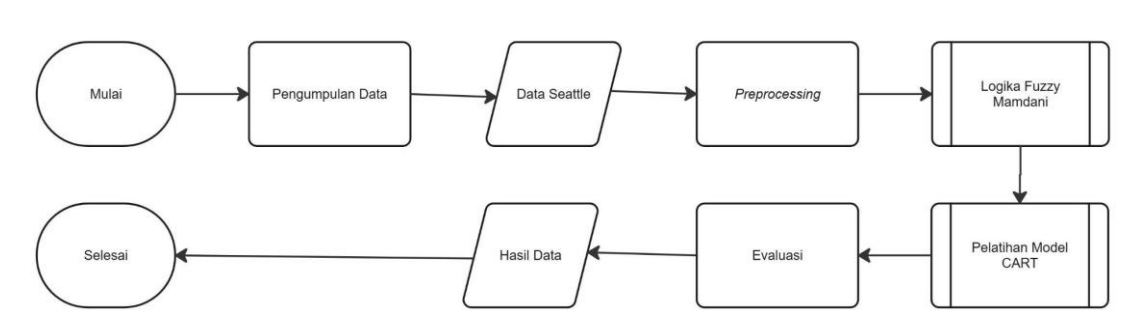
Pada penelitian [13] menggunakan kombinasi metode *Fuzzy C-Means* untuk membuat fungsi keanggotaan serta *Fuzzy Sugeno* untuk memproses inferensi. Dalam memprediksi cuaca menggunakan data cuaca BMKG Stasiun Geofisika Bandung yang dibagi menjadi data cuaca tahun 2015 sampai 2019. Parameter *input* yang digunakan ada tiga yaitu suhu, kelembapan, dan kecepatan angin, serta label *output* ada 5 berdasarkan standar BMKG yaitu berawan, hujan ringan, hujan sedang, hujan lebat, dan hujan sangat lebat. Penelitian ini menghasilkan akurasi yang baik yaitu sebesar 71,10 % dengan jumlah data prediksi benar sebesar 251 dari total 353 data cuaca.

Sebagian besar dari sistem klasifikasi cuaca yang ada mengandalkan algoritma klasifikasi berbasis data numerik dengan *machine learning* seperti *Decision Tree*, namun belum menggabungkan nilai interpretatif dari logika *Fuzzy* dalam proses prediksi. Penggunaan algoritma *CART* untuk memprediksi cuaca, tidak menggabungkan logika *Fuzzy* atau representasi linguistik. Selain itu informasi numerik pada data cuaca belum sepenuhnya dimanfaatkan dalam bentuk representasi linguistik yang dapat meningkatkan performa klasifikasi. Pada penelitian [1], [7], dan [8] menunjukkan dominasi terhadap penggunaan algoritma *Decision Tree/CART* dalam sistem klasifikasi cuaca tanpa menggunakan metode *Fuzzy Logic*, sedangkan pada penelitian [2], [5], dan [13] menunjukkan penggunaan metode *Fuzzy Logic*, tetapi tidak menggunakan pendekatan statistik atau dengan representasi linguistik dari data numerik.

Berdasarkan penelitian terdahulu dan perbandingan model-model yang telah dikembangkan sebelumnya, menunjukkan bahwa penggabungan metode *Fuzzy Logic* dengan algoritma *machine learning* seperti *Decision Tree* dapat meningkatkan performa klasifikasi [5]. Algoritma *Fuzzy Mamdani* efektif dalam penyederhanaan parameter numerik menjadi nilai linguistik berdasarkan aturan pakar [4][9]. Maka penulis membuat penelitian dengan judul Klasifikasi Cuaca Menggunakan Algoritma *Fuzzy Mamdani* dan *CART*. Judul penelitian tersebut dipilih berdasarkan hasil perbandingan model prediksi cuaca pada penelitian sebelumnya. Untuk mengatasi keterbatasan *Fuzzy Logic*, dilakukan penggabungan algoritma *CART* yang kemudian digunakan sebagai parameter tambahan dalam pelatihan model *CART* untuk meningkatkan kinerja prediksi. Pendekatan ini diharapkan mampu meningkatkan akurasi klasifikasi cuaca secara signifikan. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Fuzzy Mamdani* dan *CART* untuk sistem klasifikasi cuaca berdasarkan data historis cuaca harian dari Seattle dengan hasil performa melalui evaluasi *precision*, *recall*, *F1-score*.

2. Metode Penelitian

Penelitian ini mengimplementasikan algoritma *Fuzzy Mamdani* dan *CART* untuk sistem klasifikasi cuaca dengan menggunakan input berupa data historis cuaca di kota Seattle, Amerika Serikat. *Output* yang akan dihasilkan adalah hasil evaluasi dari algoritma *Fuzzy Mamdani* dan *CART*. Alur penelitian yang digunakan terdiri dari 4 tahapan yang dapat dilihat pada gambar 1, dimulai dengan pengumpulan data, melakukan *preprocessing* data, melakukan pendekatan metode *Fuzzy*, melakukan pelatihan data model dengan menggunakan algoritma *CART*, dan Evaluasi.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

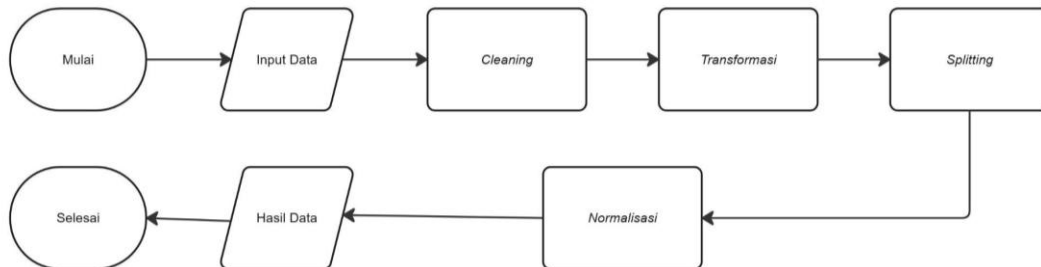
Data yang digunakan dalam penelitian ini berupa data sekunder yang diperoleh dari data historis cuaca di kota Seattle, Amerika Serikat, yang dikumpulkan selama hampir 4 tahun (2012-2015). Dataset ini diperoleh dari situs Kaggle di <https://www.kaggle.com/> dengan format file CSV (.csv). Dataset ini terakhir diperbarui pada tahun 2022. Dataset tersebut terdiri dari 6 kolom dan lebih dari 1.400 entri data, menjadikannya sumber data untuk pengujian klasifikasi cuaca. Variabel yang ada pada data cuaca yaitu *date*, *precipitation*, *temp_max*, *temp_min*, *wind*, dan *weather*. Berikut ini adalah tabel 1 yang menunjukkan contoh data cuaca pada bulan Desember 2015 didapatkan langsung dari situs data terbuka Kaggle:

Tabel 1. Data Cuaca pada Bulan Desember 2015

<i>date</i>	<i>precipitation</i>	<i>temp_max</i>	<i>temp_min</i>	<i>wind</i>	<i>weather</i>
12/25/2015	5.8	5	2.2	1.5	<i>rain</i>
12/26/2015	0	4.4	0	2.5	<i>sun</i>
12/27/2015	8.6	4.4	1.7	2.9	<i>rain</i>
12/28/2015	1.5	5	1.7	1.3	<i>rain</i>
12/29/2015	0	7.2	0.6	2.6	<i>fog</i>
12/30/2015	0	5.6	-1	3.4	<i>sun</i>
12/31/2015	0	5.6	-2.1	3.5	<i>sun</i>

2.2. Preprocessing

Sebelum data cuaca Seattle dari Kaggle digunakan dalam prediksi cuaca, tahap *preprocessing* data perlu dilakukan untuk memastikan bahwa data yang digunakan bersih, terstruktur, dan sesuai format. Berikut merupakan alur *preprocessing* yang bisa dilihat pada gambar 2.



Gambar 2. Alur *Preprocessing*

a. Cleaning

Tahap pertama adalah dengan membersihkan data dari nilai yang hilang (*missing values*) untuk menghindari kesalahan dalam proses analisis data dan pelatihan model. Oleh karena itu, semua baris data yang memiliki nilai kosong pada salah satu atribut akan dihapus (*row-wise detection*). Hasil pembersihan data dapat dilihat pada gambar 3.

```
Memulai prapemrosesan data...

Dataset berhasil dimuat.
Kolom-kolom dalam dataset: ['date', 'precipitation', 'temp_max', 'temp_min', 'wind', 'weather']
5 baris pertama:
   date  precipitation  temp_max  temp_min  wind  weather
0  1/1/2012           0.0       12.8       5.0   4.7  drizzle
1  1/2/2012          10.9       10.6       2.8   4.5   rain
2  1/3/2012           0.8       11.7       7.2   2.3   rain
3  1/4/2012          20.3       12.2       5.6   4.7   rain
4  1/5/2012           1.3        8.9       2.8   6.1   rain

Kolom 'date' berhasil dikonversi.

Jumlah data sebelum pembersihan: 1461
Jumlah data setelah pembersihan: 1461
   date  precipitation  temp_max  temp_min  wind  weather
0 2012-01-01           0.0       12.8       5.0   4.7  drizzle
1 2012-01-02          10.9       10.6       2.8   4.5   rain
2 2012-01-03           0.8       11.7       7.2   2.3   rain
3 2012-01-04          20.3       12.2       5.6   4.7   rain
4 2012-01-05           1.3        8.9       2.8   6.1   rain
```

Gambar 3. Hasil *Cleaning* Data

b. Transformation

Tahap *transformation* yaitu dengan melakukan perubahan pada label cuaca yang masih berupa string menjadi integer dengan label encoding untuk klasifikasi. Setiap label cuaca diberikan representasi angka yang unik. Pengkodean ini sangat penting agar algoritma *machine learning* dapat memproses data dalam bentuk numerik. Adapun hasil *transformation label encoder* terlihat pada gambar 4.

```
Label cuaca berhasil di-encode.
Mapping label:
'drizzle' : 0
'fog' : 1
'rain' : 2
'snow' : 3
'sun' : 4

```

	weather	weather_encoded
0	drizzle	0
1	rain	2
7	sun	4
13	snow	3
192	fog	1

Gambar 4. Hasil *Label Encoder*

c. *Splitting*

Tahap *splitting* yaitu dengan membagi data menjadi data *training* untuk melatih algoritma dan data *testing* untuk menguji. Dari 1.461 data akan dibagi dengan persentase data *training* sebesar 80% (sebanyak 1.168 data) dan data *testing* sebesar 20% (sebanyak 293 data). Dapat dilihat pada gambar 5 yang menunjukkan hasil pembagian data berupa data pengujian (x_{train}).

```
Data berhasil di-split:
- Data training: 1168 baris
- Data testing : 293 baris

Data training (fitur - X_train):
precipitation temp_max temp_min wind
1066 0.0 5.6 -3.2 5.7
638 18.5 13.9 10.0 6.3
799 18.8 12.2 6.1 2.2
380 0.0 6.7 -0.6 2.3
303 34.5 15.0 12.2 2.8

Label training (target - y_train):
1066 1
638 2
799 2
380 4
303 2
Name: weather_encoded, dtype: int64

Data testing (fitur - X_test):
precipitation temp_max temp_min wind
892 0.0 23.9 11.1 2.7
1105 5.8 7.8 6.1 0.5
413 0.0 9.4 4.4 3.4
522 0.0 26.7 12.2 2.5
1036 1.8 13.3 7.2 2.9

Label testing (target - y_test):
892 4
1105 2
413 2
522 4
```

Gambar 5. Hasil *Splitting* Data

d. *Normalisasi*

Tahap normalisasi dilakukan dengan menggunakan metode *Standard Scaler* untuk menyamakan skala nilai dari setiap fitur numerik dalam dataset sehingga memiliki nilai rata-rata 0 dan standar deviasi. Tanpa normalisasi, fitur-fitur dengan rentang skala yang lebih tinggi dapat mendominasi proses pembelajaran. Dengan metode ini, semua fitur numerik akan berada dalam rentang skala yang konsisten agar dapat menghasilkan nilai akurasi yang maksimal. Hasil normalisasi dapat dilihat pada gambar 6.

```
Normalisasi selesai menggunakan StandardScaler.
Contoh hasil normalisasi (5 data pertama dari training):
precipitation temp_max temp_min wind
0 -0.455630 -1.483372 -2.293321 1.715314
1 2.271360 -0.345925 0.349065 2.135391
2 2.315582 -0.578896 -0.431640 -0.735135
3 -0.455630 -1.332626 -1.772851 -0.665122
4 4.629838 -0.195179 0.789463 -0.315058
```

Gambar 6. Hasil Normalisasi Data

2.3. Fuzzification

Dalam penelitian ini, sistem inferensi proses pembuatan himpunan *Fuzzy* dikembangkan dengan pendekatan *Fuzzy Mamdani*, yang memerlukan pembentukan himpunan *Fuzzy* untuk setiap variabel *input* dan *output*. Tujuan dari proses ini adalah untuk mengkonversikan nilai-nilai kuantitatif dari data cuaca menjadi representasi linguistik yang dapat diproses secara logis, seperti “rendah”, “sedang”, atau “tinggi”. Setiap nilai linguistik direpresentasikan menggunakan fungsi yang banyak digunakan dalam sistem *Fuzzy* yaitu fungsi keanggotaan berbentuk segitiga (*Triangular Membership Function*). Secara sistematis, fungsi keanggotaan segitiga dirumuskan sebagai berikut:

$$\mu(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ \frac{c-x}{c-b}, & b < x < c \\ 0, & x \geq c \end{cases} \quad (1)$$

Keterangan:

x = Nilai *input*,
 $\mu(x)$ = Derajat keanggotaan
 a , b , dan c = Parameter fungsi segitiga yang menentukan bentuk kurva

a. Variabel *Input*

Adapun empat variabel input atau parameter utama yang digunakan dalam sistem *Fuzzy* adalah: precipitation, temp_max, temp_min, dan wind. Setiap variabel diberikan tiga nilai linguistik dengan rentang dan fungsi keanggotaan yang ditentukan berdasarkan karakteristik distribusi data historis.

Tabel 2. Himpunan Variabel Input *Fuzzy*

Variabel	Himpunan Fuzzy	Jangkauan
precipitation	Rendah	[0, 0, 2]
precipitation	Sedang	[1, 5, 10]
precipitation	Tinggi	[8, 15, 25]
temp_max	Dingin	[-5, 0, 10]
temp_max	Normal	[8, 15, 25]
temp_max	Panas	[20, 30, 40]
temp_min	Sangat Dingin	[-10, 0, 5]

Variabel	Himpunan Fuzzy	Jangkauan
temp_min	Dingin	[2, 8, 15]
temp_min	Hangat	[10, 20, 30]
wind	Lemah	[0, 0, 2]
wind	Sedang	[1, 3, 5]
wind	Kencang	[4, 7, 10]

Pada tabel 2 menunjukkan bahwa variabel curah hujan dikelompokkan menjadi tiga himpunan Fuzzy, yaitu “rendah”, “sedang”, dan “tinggi” berdasarkan intensitas curah hujan harian. Suhu maksimum dilabelkan ke dalam tiga himpunan Fuzzy untuk menggambarkan kondisi udara harian. Suhu minimum diberikan tiga label linguistik Fuzzy untuk menangkap perbedaan ekstrem pada suhu rendah. Klasifikasi kecepatan angin berdasarkan intensitas hembusan harian.

b. Variabel Output

Variabel *output* atau label klasifikasi pada penelitian ini yaitu kondisi cuaca yang diprediksi (*weather*) dengan lima label utama: *sun*, *drizzle*, *rain*, *fog*, dan *snow*. Kelima label tersebut diubah menjadi skala numerik *Fuzzy* pada domain [0,100] dan dibagi ke dalam lima himpunan *Fuzzy* menggunakan fungsi keanggotaan segitiga.

Tabel 3. Himpunan Variabel *Output Fuzzy*

Variabel	Himpunan Fuzzy	Jangkauan
<i>weather</i>	<i>Sun</i>	[0, 0, 20]
<i>weather</i>	<i>Drizzle</i>	[15, 30, 45]
<i>weather</i>	<i>Rain</i>	[40, 57, 70]
<i>weather</i>	<i>Fog</i>	[60, 75, 90]
<i>weather</i>	<i>Snow</i>	[85, 100, 100]

Tabel 3 menjelaskan bahwa keluaran dari sistem *Fuzzy* berupa skor numerik yang direpresentasikan ke dalam lima label cuaca.

2.4. Fuzzy Inferensi System

Kondisi inti dari sistem inferensi *Fuzzy* adalah aturan-aturan *Fuzzy* yang dirancang untuk merepresentasikan hubungan logis antara variabel *input* dan *output*. Dalam penelitian ini, aturan-aturan *Fuzzy* disusun berdasarkan korelasi antara curah hujan (*precipitation*), suhu maksimum (*temp_max*), suhu minimum (*temp_min*), kecepatan angin (*wind*), dan kondisi cuaca (*weather*). Aturan-aturan tersebut disusun berdasarkan pola data historis dan pendekatan heuristik, serta diturunkan dari pengamatan logika iklim yang umum terjadi di wilayah seperti Seattle. Setelah aturan aktif, maka dilakukan proses agregasi. Pada algoritma *Fuzzy Mamdani* keluaran dari masing-masing aturan dipotong berdasarkan nilai α . Kemudian dilakukan proses defuzzifikasi untuk menghasilkan *output* crisp. Penelitian ini mengimplementasikan sepuluh aturan *Fuzzy*, yaitu sebagai berikut:

- [Rule 1] IF (*precipitation* IS rendah) AND (*temp_max* IS panas) AND (*wind* IS lemah) THEN (*weather* IS sun)
- [Rule 2] IF (*precipitation* IS rendah) AND (*temp_min* IS hangat) THEN (*weather* IS sun)
- [Rule 3] IF (*precipitation* IS sedang) AND (*temp_min* IS dingin) THEN (*weather* IS drizzle)

- [Rule 4] IF (*precipitation* IS sedang) AND (*wind* IS sedang) THEN (*weather* IS drizzle)
- [Rule 5] IF (*precipitation* IS tinggi) AND (*temp_min* IS dingin) AND (*wind* IS sedang) THEN (*weather* IS rain)
- [Rule 6] IF (*precipitation* IS tinggi) AND (*temp_max* IS normal) THEN (*weather* IS rain)
- [Rule 7] IF (*precipitation* IS sedang) AND (*temp_min* IS dingin) AND (*temp_max* IS dingin) AND (*wind* IS lemah) THEN (*weather* IS fog)
- [Rule 8] IF (*precipitation* IS tinggi) AND (*wind* IS lemah) AND (*temp_max* IS dingin) THEN (*weather* IS fog)
- [Rule 9] IF (*precipitation* IS sedang) AND (*temp_min* IS sangat_dingin) AND (*wind* IS sedang) THEN (*weather* IS snow)
- [Rule 10] IF (*precipitation* IS tinggi) AND (*temp_min* IS sangat_dingin) THEN (*weather* IS snow)
- [Rule 11] IF (*precipitation* IS rendah) THEN (*weather* IS sun)
- [Rule 12] IF (*precipitation* IS sedang) THEN (*weather* IS drizzle)
- [Rule 13] IF (*precipitation* IS tinggi) THEN (*weather* IS rain)

2.5. Agregasi

Pada tahap ini sistem menentukan atau menilai apakah terdapat aturan *Fuzzy* yang aktif. Agregasi dicerminkan dari pemeriksaan apakah skor hasil defuzzifikasi lebih dari nol. Hal ini berarti setidaknya ada satu aturan *Fuzzy* yang aktif dalam proses *Fuzzy Inferensi System*. Jika tidak terdapat aturan yang aktif, hasil Inferensi dianggap tidak valid terhadap aturan yang telah ditetapkan untuk menghasilkan klasifikasi.

2.6. Defuzzification

Setelah tahap *Fuzzy Inferensi System*, dilakukan *defuzzification* untuk mengubah hasil *output Fuzzy* menjadi numerik yang merepresentasikan kondisi cuaca. *Defuzzification* menghasilkan nilai skor yang digunakan sebagai representasi kondisi cuaca yang kemudian diklasifikasi ke dalam label seperti *sun*, *drizzle*, *rain*, *fog*, atau *snow* berdasarkan rentang nilai tertentu.

2.7. Integrasi dengan Model CART

Setelah proses inferensi *Fuzzy* menghasilkan skor numerik yang merepresentasikan tingkat keyakinan terhadap kondisi cuaca, langkah selanjutnya adalah mengintegrasikan skor tersebut ke dalam proses klasifikasi dengan menggunakan algoritma *CART* (*Classification and Regression Tree*). *CART* adalah salah satu metode pembelajaran terawasi berbasis pohon keputusan yang efektif untuk tugas klasifikasi karena kemampuannya dalam menangani data numerik, menghasilkan model yang mudah diinterpretasikan, dan tidak memerlukan asumsi distribusi data.

a. Encoding dan Pemisahan Data

Tahap awal integrasi melibatkan proses *encoding* pada label cuaca ke bentuk numerik menggunakan bentuk numerik melalui proses label *encoding* agar diproses oleh algoritma *CART*. Kemudian data dibagi menjadi 2 yaitu data *training* dan data *testing* menggunakan metode *train_test_split*. Fitur-fitur input yang digunakan untuk melatih model *CART* mencakup: curah hujan (*precipitation*), suhu maksimum (*temp_max*), suhu minimum (*temp_min*), kecepatan angin (*wind*), dan kondisi cuaca (*weather*), hasil inferensi dari sistem *Fuzzy* (Skor *Fuzzy*). Sementara itu, label target yang berisi nilai label seperti *sun*, *drizzle*, *rain*, *fog*, dan *snow* (*weather*) yang telah dikonversi menjadi bentuk numerik.

b. Training Model CART

Model dilatih *Decision Tree* untuk membuat aturan berdasarkan kombinasi label cuaca dan skor *Fuzzy*. Parameter *max_depth* untuk membatasi kedalaman *Tree* dan mencegah terjadinya *overfitting*. Model *CART* dilatih dengan data *testing* (*x_train* dan *y_train*) menggunakan *DecisionTreeClassifier*. Model ini akan membangun pohon keputusan yang

secara bertahap memisahkan data berdasarkan fitur-fitur numerik dengan memilih batas ambang terbaik untuk mengurangi ketidakmurnian.

2.8. Evaluasi Model

Evaluasi model bertujuan untuk menilai tingkat akurasi dan efektivitas sistem prediksi cuaca yang dikembangkan melalui integrasi inferensi *Fuzzy* dan algoritma klasifikasi *CART* (*Classification and Regression Tree*). Pada penelitian ini, pengujian dilakukan dengan membandingkan data hasil klasifikasi cuaca dengan data cuaca sebenarnya dalam mendapatkan tingkat akurasi penelitian. Akurasi dihitung sebagai rasio antara jumlah prediksi yang benar dan total jumlah data. Hasil ini menunjukkan kinerja sistem dalam mengonversi *input* cuaca menjadi label yang sesuai.

3. Hasil dan Diskusi

3.1. Hasil Eksperimen

Setelah melalui tahapan *preprocessing*, data kemudian dilatih sekaligus diuji menggunakan algoritma *CART*. Untuk model ini digunakan parameter tambahan skor *Fuzzy* untuk melatih sekaligus menguji data. Hasil yang ditunjukkan berupa akurasi sebesar 81.23% dengan nilai *precision*, *recall*, dan *F1-score* masing masing. Perhitungan *precision*, *recall*, dan *F1-score* pada masing-masing label menunjukkan bahwa label "sun" dan "rain" memiliki *F1-score* tinggi yaitu 0,8214 dan 0,9398. Evaluasi ini mengindikasikan bahwa model sangat andal pada dua label dominan. Namun, pada label "fog" dan "drizzle" menunjukkan performa rendah yaitu 0,1765 dan 0,1250, evaluasi ini menunjukkan bahwa model perlu ditingkatkan dalam menangani data minoritas. Hasil Nilai-nilai *precision*, *recall*, *F1-score* dapat dilihat pada tabel 4.

Tabel 4. Nilai *precision*, *recall*, *F1-score* Fuzzy Mamdani dan *CART*

Label	<i>precision</i>	<i>recall</i>	<i>F1-score</i>
Drizzle	0.2000	0.0909	0.1250
Fog	0.2143	0.1500	0.1765
Rain	0.9750	0.9070	0.9398
Snow	1.0000	0.4000	0.5714
Sun	0.7566	0.8984	0.8214

Sistem klasifikasi cuaca mencapai akurasi keseluruhan sebesar 81.91% dengan memprediksi data dengan benar sebanyak 240 dari 293 data uji. Hasil pengujian ini menunjukkan sistem memiliki performa yang cukup baik dalam memprediksi kondisi cuaca. Nilai akurasi keseluruhan dapat dilihat pada tabel 5.

Tabel 5. Nilai Akurasi Keseluruhan

Akurasi Keseluruhan	
Total Data Uji	293
Prediksi Benar	238
Prediksi Salah	55
Akurasi Keseluruhan	81.23%

4. Kesimpulan

Penelitian ini berhasil mengklasifikasi cuaca dengan menggunakan algoritma *Fuzzy Mamdani* dan *CART (Classification and Regression Tree)*. Sistem ini menggunakan data numerik (*precipitation*, *temp_max*, *temp_min*, *wind*) dalam proses inferensi *Fuzzy* untuk menghasilkan skor *Fuzzy* berdasarkan aturan pakar yang mencerminkan kondisi cuaca. Skor *Fuzzy* tersebut kemudian digabungkan dengan data asli sebagai fitur tambahan dalam proses pelatihan model *CART*. Berdasarkan pengujian terhadap 20% data uji, sistem menunjukkan bahwa model ini mampu mencapai tingkat akurasi sebesar 81.23%, dengan *precision*, *recall*, dan *F1-score* yang tinggi terutama pada label cuaca yang dominan seperti *sun* dan *rain*. Hasil penelitian ini menunjukkan bahwa penggunaan algoritma *Fuzzy Mamdani* dan *CART* dapat menjadi pendekatan yang efektif dalam klasifikasi cuaca berbasis data historis. Untuk pengembangan lebih lanjut, sistem dapat ditingkatkan dengan menambahkan parameter cuaca lainnya, menggunakan penerapan algoritma klasifikasi yang berbeda, serta menguji performa dengan data wilayah yang berbeda.

Daftar Pustaka

- [1] A. Purnamawati, M. N. Winnarto, dan M. Mailasari, "Analisis CART (Classification and Regression Trees) untuk Prediksi Pengguna Sepeda Berdasarkan Cuaca," *J. Teknoinfo*, vol. 16, no. 1, p. 14, 2022.
- [2] E. S. Puspita and L. Yulianti, "Perancangan Sistem Peramalan Cuaca Berbasis Logika Fuzzy," *J. Media Infotama*, vol. 12, no. 1, 2016.
- [3] F. K. Ghassani *et al.*, "Penerapan Logika Fuzzy Untuk Memprediksi Tingkat Curah Hujan di Kota Kendari," *J. of Computers and Digital Business*, vol. 3, no. 2, pp. 76–85, 2024.
- [4] I. H. Wele, N. D. Rumlaklak, and M. Boru, "Sistem Peramalan Cuaca dengan Fuzzy Mamdani (Studi Kasus: Badan Meteorologi, Klimatologi, dan Geofisika Lasiana)," *J. Komputer dan Informatika*, vol. 8, no. 2, pp. 163–169, 2020, doi: 10.35508/jicon.v8i2.2883.
- [5] J. Manurung *et al.*, "Analisis Algoritma C4.5 dan Fuzzy Sugeno untuk Optimasi Rule Base Fuzzy," *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 5, no. 2, pp. 166–171, 2022.
- [6] J. Saini, M. Dutta, and G. Marques, "Fuzzy Inference System Tree with Particle Swarm Optimization and Genetic Algorithm: A Novel Approach for PM10 Forecasting," *Expert Systems with Applications*, vol. 183, p. 115376, 2021.
- [7] K. Halim, D. E. Herwindiati, and T. Sutrisno, "Penerapan Metode Decision Tree untuk Prakiraan Cuaca Kota Bekasi," *J. Ilmu Komputer dan Sistem Informasi*, vol. 11, no. 2, pp. 1–5, 2023.
- [8] M. A. Hasanah *et al.*, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, 2021.
- [9] N. R. H. Karismadi and R. A. Saputra, "Implementasi Logika Fuzzy Mamdani dalam Prediksi Curah Hujan di Kota Kendari," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 1138–1145, 2024.
- [10] N. Septiyani and A. Agoestanto, "Penerapan Logika Fuzzy Mamdani pada Prakiraan Cuaca Harian di Kabupaten Cilacap," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 6, 2023.
- [11] N. Uzoukwu and A. Purqon, "Fuzzy Logic Model for Predicting the Heat Index," *arXiv preprint*, arXiv:2210.16051, 2022.
- [12] R. Kumar, "Decision Tree for the Weather Forecasting," *International Journal of Computer Applications*, vol. 76, no. 2, 2013.
- [13] S. I. Hadi, E. Ermatita, and N. Chamidah, "Penerapan Fuzzy C-Means dan Fuzzy Sugeno dalam Memprediksi Cuaca," *Informatik: Jurnal Ilmu Komputer*, vol. 18, no. 1, pp. 11–22, 2022.
- [14] S. M. K. Labibah, S. Prasetya, and I. Nuriskasari, "Prediksi Cuaca Sebagai Indikator Sumber EBT Menggunakan Estimasi Fuzzy Logic," *Prosiding Seminar Nasional Teknik Mesin*, no. 1, 2022.
- [15] Y. B. Pramono, E. Susanto, and B. Setiadi, "Implementasi Wireless Sensor Network (WSN) untuk Sistem Perkiraan Cuaca dengan Menggunakan Logika Fuzzy," *eProceedings of Engineering*, vol. 2, no. 1, 2015.

- [16] Y. H. M. Hutabarat, "Pengembangan Sistem Informasi Prakiraan Cuaca Berbasis Dampak Menggunakan Model Prakiraan Cuaca Numerik untuk Wilayah Jakarta," *Jurnal Widya Climago*, vol. 2, no. 2, 2020.

Halaman ini sengaja dibiarkan kosong