

PREDIKSI CURAH HUJAN HARIAN DENGAN METODE *K-NEAREST NEIGHBOR* (KNN) MENGGUNAKAN PYTHON

Afinn Fita Ningsih¹, Riski Aspriyani^{2*}, Mizan Ahmad³

^{1,2,3}Matematika, Universitas Nahdlatul Ulama Al Ghazali, Cilacap
Email: rizky.asp@gmail.com

Abstrak

Iklm merupakan kondisi rata-rata cuaca yang terjadi dalam jangka waktu yang relatif panjang pada suatu wilayah yang luas, sehingga memerlukan proses pemantauan dan proyeksi secara berkelanjutan. Salah satu komponen dalam iklim adalah curah hujan. Curah hujan merupakan salah satu informasi yang penting, khususnya pada sektor pertanian. Informasi curah hujan diperlukan untuk mengantisipasi kemungkinan terjadinya kejadian ekstrem, yang dapat mengakibatkan produksi pertanian mengalami kegagalan. Prediksi terhadap curah hujan harian perlu dilakukan. Metode *K-Nearest Neighbor* (KNN) dipilih karena mampu menangani data curah hujan yang tidak menentu. Metode KNN merupakan metode klasifikasi yang dapat menghasilkan sebuah prediksi, yang didasarkan pada asumsi bahwa suatu objek dengan jarak yang berdekatan satu dengan yang lainnya juga akan cenderung mempunyai karakteristik yang serupa. Penelitian ini menggunakan data harian terkait curah hujan dengan faktor suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata, dan lama penyinaran matahari selama periode 2021-2025. Setiap variabel terdiri atas 1.826 data, yang selanjutnya dibagi menjadi data *training* dan data *testing*, dengan perbandingan 80:20. Hasil evaluasi klasifikasi menunjukkan nilai *accuracy* 71%, *precision* 56%, *recall* 71%, dan *f1-score* 61%. Berdasarkan kriteria performa yang merujuk pada akurasi 71%, dikategorikan dalam rentang performa cukup. Hal ini menunjukkan kemampuan model yang memadai dalam melakukan prediksi.

Kata kunci: Curah Hujan, Prediksi, *K-Nearest Neighbor* (KNN)

Abstract

Climate is the average weather conditions that occur over a relatively long period of time over a large area, thus requiring continuous monitoring and projection. One component of climate is rainfall. Rainfall is crucial information, particularly in the agricultural sector. Rainfall information is needed to anticipate the possibility of extreme events, which can lead to agricultural production failure. Therefore, daily rainfall predictions are necessary. The *K-Nearest Neighbor* (KNN) method was chosen because it is able to handle unpredictable rainfall data. The KNN method is a classification method that can generate predictions based on the assumption that objects with proximity to each other tend to have similar characteristics. This study used daily rainfall data with factors such as average air temperature, average humidity, average air pressure, average wind speed, and solar irradiance during the 2021-2025 period. Each variable consists of 1,826 data points, which then serve as training and test data, with a ratio of 80:20. The classification evaluation results showed an accuracy of 71%, a precision of 56%, a recall of 71%, and an f1-score of 61%. Based on the performance criteria, which refers to accuracy, 71% is categorized as within the adequate performance range. This demonstrates the model's strong predictive capabilities.

Keywords: Rainfall, Prediction, *K-Nearest Neighbor* (KNN)

PENDAHULUAN

Iklim didefinisikan sebagai kondisi rata-rata cuaca yang berlangsung dalam jangka waktu yang relatif panjang pada suatu wilayah yang luas, sehingga memerlukan proses pemantauan serta proyeksi yang dilakukan secara berkelanjutan. Kondisi iklim berpengaruh besar terhadap

kelangsungan hidup manusia (Tjasyono, 2012). Dalam konteks iklim, pola curah hujan tahunan digunakan untuk menentukan klasifikasi iklim suatu daerah. Curah hujan merupakan salah satu parameter atau unsur cuaca yang paling berperan penting. Tingkat curah hujan yang terlalu tinggi ataupun terlalu rendah dapat menimbulkan berbagai dampak, baik positif maupun negatif, terhadap kehidupan manusia. Curah hujan yang tinggi memberikan manfaat dalam penyediaan sumber daya air serta meningkatkan kesuburan, tetapi berpotensi menimbulkan bencana alam berupa banjir dan tanah longsor. Selain itu, curah hujan tinggi juga dapat menghambat berbagai aktivitas manusia di lapangan. Sedangkan curah hujan yang rendah dapat meminimalkan risiko terjadinya banjir dan membantu proses pengeringan hasil pertanian/perkebunan. Namun demikian, dapat berisiko menyebabkan kekeringan, kegagalan panen, serta meningkatnya potensi kebakaran hutan dan lahan. Untuk itu, curah hujan adalah parameter cuaca yang berperan penting dalam memenuhi kebutuhan sumber daya yang ada (Tjasyono, 2012). Terdapat beberapa faktor yang dapat memengaruhi curah hujan yaitu, suhu udara, kelembapan, tekanan udara, kecepatan angin, dan lama penyinaran matahari (Rakhmalia, 2018). Faktor-faktor tersebut tentunya memberikan dampak pada perubahan cuaca yang terjadi.

Kabupaten Cilacap merupakan kabupaten terluas di Jawa Tengah dan terletak di posisi barat daya Provinsi Jawa Tengah. Terletak di antara $7^{\circ}30'$ dan $7^{\circ}45'20''$ lintang selatan dan antara $108^{\circ}4'30''$ dan $109^{\circ}30'30''$ bujur timur dengan luas Kabupaten Cilacap mencapai 6,48% dari luas Provinsi Jawa Tengah dan terbagi menjadi 24 kecamatan (Badan Pusat Statistik Kabupaten Cilacap, 2025). Kabupaten Cilacap juga dikenal sebagai daerah dengan potensi lahan pangan yang cukup luas. Data BPS Provinsi Jawa Tengah menunjukkan bahwa pada tahun 2025 luas lahan sawah di Kabupaten Cilacap mencapai 60,26 hektare atau sekitar 54,14% dari total wilayahnya. Data BPS Provinsi Jawa Tengah menjelaskan bahwa produksi padi sawah di Kabupaten Cilacap mencapai 758.568 ton, meningkat 14,96% dibandingkan dengan produksi padi pada tahun 2024. Untuk itu, informasi curah hujan akan menjadi hal yang penting dalam mengantisipasi kemungkinan terjadinya kejadian ekstrem, yang mengakibatkan produksi pertanian mengalami kegagalan. Dengan demikian, diperlukan data prediksi curah hujan yang dapat memberikan keputusan praktis di lapangan. Sehingga penguatan akurasi prediksi menjadi langkah mutlak untuk melindungi sektor pertanian, menjaga stabilitas sumber daya air, dan memitigasi risiko bencana demi kelangsungan hidup manusia yang lebih aman.

Secara umum, prediksi curah hujan dilakukan dengan memanfaatkan data historis yang kemudian dianalisis menggunakan metode atau pendekatan tertentu. Berbagai metode prediksi digunakan untuk memperkirakan nilai pada data runtun waktu (*time series*). Pola yang tidak menentu yang dimiliki curah hujan juga akan menyebabkan kesulitan dalam memprediksi secara manual (Nanda et al., 2022). Oleh karena itu, pendekatan prediksi yang hanya didasarkan pada metode statistika/konsep-konsep matematika memiliki keterbatasan. Namun demikian, perkembangan

dalam bidang *data mining* memungkinkan sistem komputasi untuk mempelajari serta mengidentifikasi pola-pola data yang kompleks (Lumbanraja et al., 2019). Di sisi lain, pendekatan alternatif yang dapat diterapkan adalah metode *heuristik* modern yang didasarkan pada algoritma dalam *soft computing* (SC), yaitu *K-Nearest Neighbor* (KNN). Metode KNN merupakan bagian dari teknik *data mining* yang mampu mengidentifikasi pola pada data sekuensial, yaitu data yang memiliki karakteristik berdasarkan urutan waktu (Nanda et al., 2022). Dengan kemampuan tersebut, metode KNN dapat dimanfaatkan sebagai salah satu teknik dalam *data mining* untuk melakukan proses prediksi (Alfani W.P.R. et al., 2021).

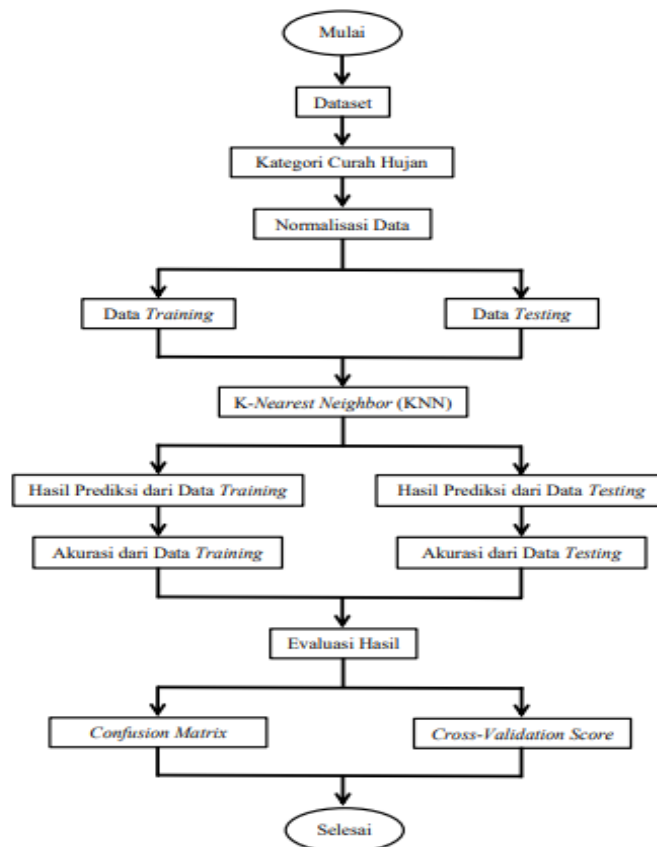
Metode *K-Nearest Neighbor* (KNN) merupakan metode klasifikasi yang dapat menghasilkan sebuah prediksi yang didasarkan pada asumsi bahwa suatu objek dengan jarak yang berdekatan satu dengan yang lainnya juga akan cenderung memiliki karakteristik yang serupa. Dengan demikian, apabila karakteristik suatu objek telah diketahui, maka karakteristik objek lainnya dapat diprediksi berdasarkan kedekatan dengan tetangga terdekatnya (Rachma, 2022). KNN juga termasuk salah satu algoritma dalam *machine learning* yang relatif sederhana dan dapat diterapkan dalam analisis data dengan jumlah dimensi variabel yang cukup banyak (Alkhatib et al., 2013). *Machine learning* sendiri merupakan salah satu pendekatan komputasi cerdas yang memiliki potensi besar dalam bidang klasifikasi dan prediksi (Darmayanti et al., 2021). Dalam metode KNN, nilai *k* menunjukkan jumlah tetangga terdekat yang digunakan sebagai acuan dalam menentukan kategori prediksi pada data pengujian (*testing*) (Sukanto et al., 2020). Pemilihan variabel yang digunakan dalam proses analisis memiliki pengaruh yang signifikan terhadap hasil serta tingkat akurasi klasifikasi yang dihasilkan (Rachma, 2022). Dalam konteks prediksi curah hujan, beberapa variabel yang umum digunakan meliputi suhu udara, kelembapan, tekanan udara, kecepatan angin, dan lama penyinaran matahari. Variabel-variabel tersebut dapat dimanfaatkan untuk mengidentifikasi kecenderungan terjadinya hujan pada periode waktu mendatang (Nanda et al., 2022).

Penelitian ini menggunakan data harian terkait curah hujan dengan faktor suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata, dan lama penyinaran matahari sebagai data uji prediksi menggunakan KNN dengan bantuan Python. Pada penelitian sebelumnya, terdapat hasil prediksi yang menunjukkan nilai *Root Mean Square Error* (RMSE) sebesar 36,3942 yang dipengaruhi oleh keterbatasan jumlah dataset yang digunakan untuk meramalkan curah hujan (Fatimah et al., 2025). Hal ini memberikan informasi bahwa dalam peramalan untuk menentukan perbandingan data *training* dan data *testing* dapat menggunakan perbandingan sebesar 80 : 20, yang menghasilkan suatu prediksi dengan tingkat akurasi data sebesar 89% (Rangkuti et al., 2021). Dengan demikian, dalam penelitian ini untuk menganalisis prediksi curah hujan harian dengan mengevaluasi tingkat akurasi prediksi yang dihasilkan pada metode *K-Nearest Neighbor* berbantuan Python dilakukan menggunakan perbandingan data training dan data testing 80:20.

METODE

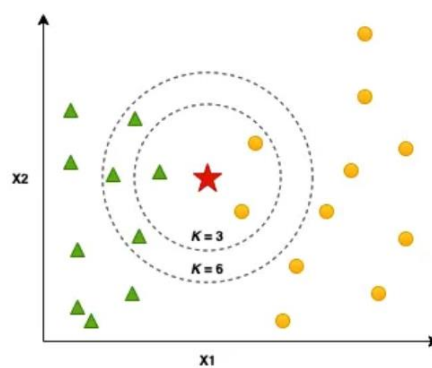
Penelitian ini menggunakan pendekatan kuantitatif dengan sumber data yang digunakan adalah data sekunder. Data sekunder merupakan data yang diperoleh secara tidak langsung dari sumber yang telah tersedia dan sebelumnya dikumpulkan oleh pihak lain, seperti dalam buku, jurnal, maupun laporan. Dalam penelitian ini, data diperoleh dari data Synop ME48, yaitu data harian hasil pengamatan cuaca permukaan (meteorologi) BMKG setiap jam pengamatan. Data yang digunakan berupa data harian pada Januari 2021 hingga Desember 2025 yang meliputi curah hujan, suhu rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata, serta lama penyinaran matahari. Data-data tersebut dimuat dalam dataset yang dibagi menjadi dua kelompok, yaitu data pelatihan (*training*) dan data pengujian (*testing*), dengan perbandingan sebesar 80:20.

Dalam penelitian ini, curah hujan ditetapkan sebagai variabel dependen. Sementara itu, penentuan variabel independen dilakukan dengan mempertimbangkan faktor-faktor meteorologis yang berperan dalam proses pembentukan awan dan terjadinya presipitasi (Rakhmalia, 2018), yaitu suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata, dan lama penyinaran matahari. Prosedur dalam menganalisis data pada penelitian ini dapat ditunjukkan pada gambar berikut ini.



Gambar 1. Flowchart Analisis Data

Metode KNN mengklasifikasikan pengamatan yang tidak diketahui ke kelas mayoritas di antara pengamatan tetangga terdekatnya, yang diukur dengan menggunakan matriks jarak dalam data latih. Algoritma KNN merupakan metode pengklasifikasian objek-objek yang paling dekat dengan objek berdasarkan pembelajaran. Prinsip kerja algoritma KNN dilakukan dengan menghitung jarak antara data yang akan dievaluasi dengan sejumlah k tetangga terdekatnya yang terdapat dalam data pelatihan (*training data*). Selanjutnya, proses prediksi menggunakan metode KNN didasarkan pada pola kedekatan atau kemiripan yang muncul secara berulang dalam data yang diamati dari waktu ke waktu. Untuk mempermudah pemahaman mengenai konsep kerja metode KNN, dapat diperhatikan ilustrasi gambar di bawah ini (Rina, 2023):



Gambar 2. *K-Nearest Neighbor (KNN)*

Gambar 2 menunjukkan sebuah dataset yang terdiri atas dua kelas, yaitu kelas positif yang direpresentasikan oleh simbol segitiga berwarna hijau dan kelas negatif yang direpresentasikan oleh simbol lingkaran berwarna kuning. Pada ilustrasi tersebut, simbol bintang berwarna merah merepresentasikan data uji yang akan ditentukan kelasnya menggunakan algoritma KNN. Selain itu, pada gambar juga ditampilkan penggunaan dua nilai parameter k , yaitu $k = 3$ dan $k = 6$.

Algoritma KNN bekerja dengan cara mengidentifikasi sejumlah tetangga terdekat dari suatu data uji (*testing*), kemudian menentukan kelas data tersebut berdasarkan kelas yang paling dominan di antara tetangga terdekat tersebut. Ketika $k = 3$, jumlah tetangga terdekat yang dipertimbangkan adalah tiga data. Berdasarkan ilustrasi pada gambar, dua dari tiga tetangga terdekat termasuk ke dalam kelas negatif, sedangkan satu tetangga lainnya termasuk dalam kelas positif. Karena kelas yang paling dominan adalah kelas negatif, maka data uji yang direpresentasikan oleh bintang berwarna merah pada nilai $k = 3$ akan diklasifikasikan atau diprediksi sebagai kelas negatif. Prinsip yang sama juga berlaku ketika nilai $k = 6$, yaitu kelas data uji ditentukan berdasarkan mayoritas kelas dari enam kategori terdekat yang teridentifikasi.

Mencari tetangga atau jarak pada algoritma KNN menggunakan jarak Euclidean yang didefinisikan sebagai berikut (Amelia, 2018):

$$D_{i,j} = \sqrt{(X_1^{latih} - X_1^{uji})^2 + \dots + (X_D^{latih} - X_D^{uji})^2}$$

$$= \sqrt{\sum_{i,j=1}^D (X_i^{latih} - X_j^{uji})^2}$$

Keterangan:

$D_{i,j}$: Jarak antar objek (jarak *euclidean*)

X_i^{latih} : Nilai pada data *training*

X_j^{uji} : Nilai pada data *testing*

D : Banyaknya data

i, j : 1, 2, 3, ..., n

Dalam penelitian ini digunakan Algoritma KNN dengan tahapan prosedur sebagai berikut (Amelia, 2018):

1. Tentukan jumlah tetangga terdekat (k) yang akan dipilih.
2. Menghitung jarak antara data pengujian (*testing*) dan seluruh data pelatihan (*training*). Pengukuran jarak yang digunakan dalam penelitian ini adalah jarak *Euclidean*.
3. Mengurutkan data pelatihan (*training*) berdasarkan nilai jarak, mulai dari jarak yang paling dekat hingga yang paling jauh.
4. Memilih sejumlah k data dengan jarak terdekat sebagai tetangga yang akan dijadikan dasar dalam proses klasifikasi/prediksi.
5. Menghitung frekuensi atau jumlah kemunculan masing-masing kategori dari k data yang telah terpilih, untuk menentukan kelas mayoritas sebagai hasil klasifikasi/prediksi.

HASIL DAN PEMBAHASAN

Analisis data pada penelitian ini menggunakan Python. Hasil analisis data yang diperoleh dijelaskan sebagai berikut.

Pengkategorian Data Curah Hujan

Pengkategorian dalam penelitian ini merujuk pada literatur di *Buletin* Edisi Juli 2020 yang diterbitkan oleh Stasiun Meteorologi Umu Mahang Kunda, Sumba Timur.

Tabel 1. Kategori Curah Hujan

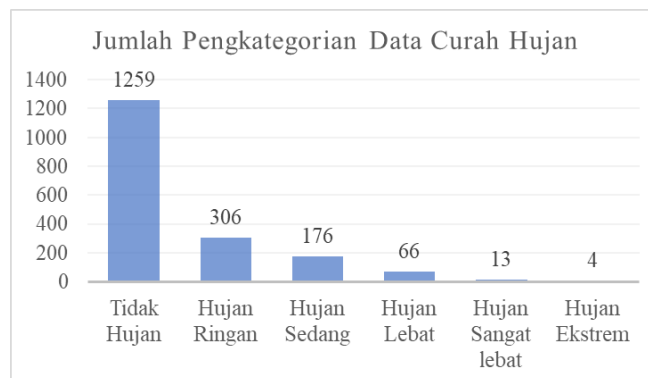
Kategori	Rentang Curah Hujan (mm/hari)
Tidak Hujan	$x < 5$
Hujan Ringan	$5 \leq x < 20$
Hujan Sedang	$20 \leq x < 50$
Hujan Lebat	$50 \leq x < 100$
Hujan Sangat Lebat	$100 \leq x < 150$
Hujan Ekstrem	$x \leq 150$

Berdasarkan klasifikasi yang ditetapkan Stasiun Meteorologi atau Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), curah hujan dibagi dalam enam kategori, yang disajikan pada Tabel 1 (Limahelu et al., 2020). Berdasarkan enam kategori tersebut, Tabel 2 menyajikan hasil pengkategorian yang diterapkan pada data curah hujan dalam penelitian ini sebagai berikut.

Tabel 2. Pengkategorian Data Curah Hujan

No.	Tanggal	Curah Hujan	Kategori
1	01/01/2021	10,1	Hujan Ringan
2	02/01/2021	3,7	Tidak Hujan
3	03/01/2021	0	Tidak Hujan
4	04/01/2021	1,6	Tidak Hujan
5	05/01/2021	8,2	Hujan Ringan
6	06/01/2021	0	Tidak Hujan
7	07/01/2021	0	Tidak Hujan
8	08/01/2021	7,5	Hujan Ringan
9	09/01/2021	4,8	Tidak Hujan
10	10/01/2021	43,9	Hujan Sedang
...
...
...
1822	27/12/2025	32,4	Hujan Sedang
1823	28/12/2025	2,4	Tidak Hujan
1824	29/12/2025	12,4	Hujan Ringan
1825	30/12/2025	0	Tidak Hujan
1826	31/12/2025	19,9	Hujan Ringan

Hasil pengkategorian data curah untuk setiap kategori disajikan dalam bentuk grafik, sebagaimana ditunjukkan sebagai berikut:



Gambar 3. Grafik Jumlah Pengkategorian Data Curah Hujan

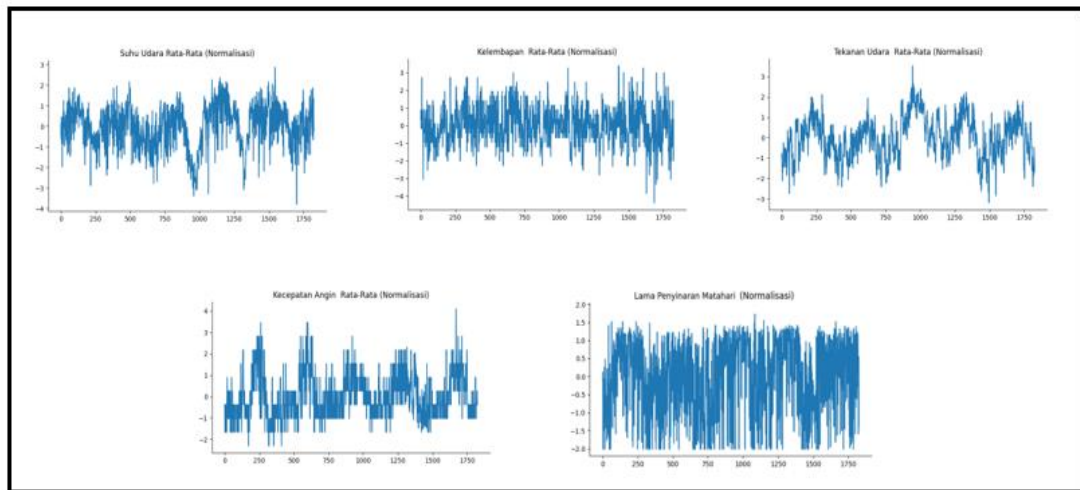
Berdasarkan Gambar 3, diketahui bahwa jumlah pengkategorian terhadap data curah hujan. Curah hujan dengan kategori tidak hujan berjumlah 1.259 data, kategori hujan ringan berjumlah 306, kategori hujan sedang berjumlah 176, kategori hujan lebat berjumlah 66, kategori hujan sangat lebat berjumlah 13, dan kategori hujan ekstrem berjumlah 4.

Normalisasi Data

Proses normalisasi ditetapkan pada data yang memiliki perbedaan yang cukup besar, karena kondisi tersebut berpotensi menghasilkan model dengan tingkat akurasi yang rendah (Sholeh et al., 2022). Normalisasi berfungsi untuk mengubah data sehingga berada pada skala yang lebih seragam

atau mendekati kondisi normal. Hal ini diperlukan karena data yang diperoleh menunjukkan adanya nilai yang sangat besar maupun sangat kecil.

Pada metode KNN, normalisasi data merupakan tahap yang penting karena algoritma tersebut mengandalkan perhitungan jarak dalam menentukan tetangga terdekat. Melalui proses normalisasi, setiap variabel data diperlakukan yang setara dengan melakukan skala nilai ke dalam rentang yang sebanding, sehingga dapat meningkatkan akurasi kinerja model secara keseluruhan. Berikut disajikan grafik yang menunjukkan hasil proses normalisasi data pada setiap variabel X yang berperan sebagai variabel independen:



Gambar 4. Grafik Hasil Normalisasi

Pembagian Data Training dan Data Testing

Data dalam proses klasifikasi dibagi menjadi dua bagian, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*) dengan perbandingan 80 : 20 yang ditunjukkan pada tabel berikut:

Tabel 3. Pembagian Data Training dan Data Testing

Data	Jumlah
<i>Training</i>	1.460
<i>Testing</i>	366
Total	1.826

K-Nearest Neighbor

Berdasarkan prosedur tahapan algoritma *K-Nearest Neighbor* (KNN), proses perhitungan jarak dalam penelitian ini menggunakan metode *Euclidean Distance* dengan nilai parameter k yang ditetapkan adalah 43. Oleh karena itu, model KNN yang ditetapkan dalam penelitian ini dirumuskan sebagai berikut:

```

KNeighborsClassifier
KNeighborsClassifier(metric='euclidean', n_neighbors=43, weights='distance')
    
```

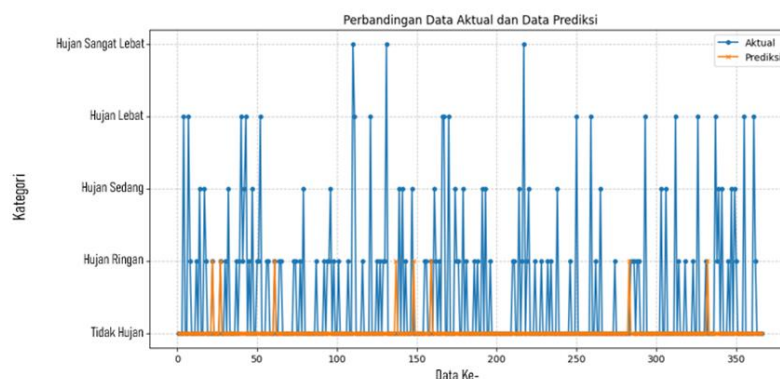
Gambar 5. Sintak *K-Nearest Neighbor* (KNN)

Penentuan nilai k tersebut didasarkan pada rumus $k = \sqrt{n}$ (Sukamto et al., 2020). Meskipun demikian, pemilihan nilai k pada algoritma KNN pada dasarnya tidak memiliki standar baku yang pasti (Alfani W.P.R. et al., 2021). Model KNN menggunakan metode *Euclidean Distance* untuk menghitung jarak antara setiap objek pada data uji (*testing*) dengan data latih (*training*). Adapun rumus perhitungan jarak *Euclidean* tersebut adalah sebagai berikut:

$$D_{i,j} = \sqrt{(X_1^{latih} - X_1^{uji})^2 + \dots + (X_D^{latih} - X_D^{uji})^2}$$

Hasil Prediksi

Hasil prediksi yang diperoleh dari data *training* menunjukkan kesesuaian dengan nilai aktual pada data tersebut. Sementara itu, hasil prediksi dari model KNN pada data *testing* disajikan dalam bentuk grafik yang memperlihatkan perbandingan antara nilai aktual dan nilai prediksi yang dihasilkan oleh metode KNN untuk data *testing*, sebagaimana ditunjukkan berikut ini:



Gambar 6. Grafik Perbandingan Nilai Aktual dan Nilai Hasil Prediksi dari Model KNN untuk Data *Testing*

Berdasarkan hasil prediksi yang diperoleh melalui penerapan model KNN, sebagian besar data pengujian dapat diklasifikasikan atau diprediksi sesuai dengan kategori sebenarnya. Dari total 366 data uji yang digunakan dalam penelitian ini, sebanyak 260 data (71%) dapat diprediksi dengan benar, sedangkan 106 data (29%) lainnya menunjukkan ketidaksesuaian antara hasil prediksi dengan nilai aktualnya. Hasil penelitian ini didukung oleh hasil penelitian yang menjelaskan bahwa penerapan Data Mining untuk menentukan potensi hujan harian dengan menggunakan algoritma KNN mampu menghasilkan prediksi yang cukup baik (Agung et al., 2023; Harun et al., 2020). Begitu pula dengan hasil penelitian yang menyebutkan bahwa model KNN cukup baik dalam menjelaskan hubungan antara variabel X dalam meramalkan curah hujan (Fatimah et al., 2025).

Diagram hasil pada Gambar 6 menunjukkan bahwa data aktual memperlihatkan variasi kategori curah hujan yang cukup beragam, mulai dari kondisi tidak hujan hingga hujan dengan intensitas sangat lebat. Namun demikian, hasil prediksi dari model didominasi oleh kategori “Tidak Hujan” dan sebagian kecil diprediksi sebagai kategori “Hujan Ringan”. Diketahui bahwa nilai data

aktual direpresentasikan dengan warna biru, sedangkan nilai hasil prediksi yang dihasilkan oleh model KNN pada data pengujian (testing) ditunjukkan dengan warna oranye. Hasil prediksi tersebut menghasilkan dua kategori, yaitu kategori tidak hujan dan kategori hujan ringan. Berdasarkan hasil prediksi tersebut, kategori tidak hujan merupakan kategori dengan jumlah prediksi benar yang paling dominan dibandingkan dengan kategori lain. Hasil ini menunjukkan bahwa penerapan *Data Mining* untuk menentukan potensi hujan harian dengan menggunakan algoritma KNN mampu menghasilkan prediksi yang cukup baik (Harun et al., 2020). Prediksi benar yang paling dominan, sejalan dengan karakteristik dataset cuaca di wilayah tropis atau subtropis, di mana frekuensi hari tanpa hujan secara statistik lebih tinggi. Hal ini didukung oleh kebutuhan adanya sistem prediksi cuaca yang lebih akurat dan adaptif untukantisipasi perubahan iklim global dan peningkatan frekuensi kejadian cuaca ekstrem (Wulandari & Aziz, 2025).

Tingkat Akurasi

Perbandingan antara nilai aktual dan nilai hasil prediksi digunakan untuk menghasilkan atau menentukan tingkat akurasi model. Hasil akurasi pada data *training* sebesar 1,00 atau 100% di mana akurasi tersebut dihitung berdasarkan hasil prediksi pada data *training*. Sedangkan akurasi pada data *testing* sebesar 0,71 atau 71% di mana perhitungan akurasi berdasarkan model pada data *testing*. Evaluasi hasil dalam penelitian ini dilakukan berdasarkan *Confusion matrix* dan *Cross-Validation Score* (Rachma, 2022). Hasil *Confusion matrix* ditunjukkan pada gambar 7 sebagai berikut.

	precision	recall	f1-score	support
Hujan Ekstrem	0.00	0.00	0.00	1
Hujan Lebat	0.00	0.00	0.00	12
Hujan Ringan	0.33	0.07	0.12	54
Hujan Sangat Lebat	0.00	0.00	0.00	2
Hujan Sedang	0.00	0.00	0.00	38
Tidak Hujan	0.73	0.99	0.84	259
accuracy			0.71	366
macro avg	0.18	0.18	0.16	366
weighted avg	0.56	0.71	0.61	366

Gambar 7. Hasil Evaluasi Klasifikasi

Hasil *accuracy* diperoleh sebesar 0,71 atau dalam present yaitu sebesar 71%. Hasil dari *precision* yaitu sebesar 0,56 atau dalam present yaitu sebesar 56%. Hasil dari *recall* yaitu sebesar 0,71 atau dalam present yaitu sebesar 71%. Dan hasil dari *f1-score* yaitu sebesar 0,61 atau dalam present yaitu sebesar 61%. Berdasarkan kriteria performa klasifikasi dengan nilai akurasi atau *accuracy* sebesar 71% menghasilkan nilai dengan performa klasifikasi cukup baik. Sedangkan hasil *Cross-Validation Score* melalui penerapan *K-Fold Cross Validation* dengan skema *10-fold cross validation* menunjukkan variasi *fold* pertama sebesar 68%, *fold* kedua sebesar 69%, *fold* ketiga sebesar 68%, *fold* keempat sebesar 65%, *fold* kelima sebesar 68%, *fold* keenam sebesar 68%, *fold* ketujuh sebesar

67%, *fold* kedelapan sebesar 68%, *fold* kesembilan sebesar 67%, dan *fold* kesepuluh sebesar 69%. Secara keseluruhan, rata-rata nilai akurasi yang diperoleh dari sepuluh *fold* pada proses *10-fold cross-validation* tersebut adalah sebesar 68%. Tingkat akurasi ini menunjukkan bahwa model KNN cukup baik dalam melakukan prediksi (Harun et al., 2020; Nanda et al., 2022).

KESIMPULAN

Berdasarkan nilai $k = 43$ dengan data yang digunakan adalah 1826 data dengan perbandingan 80 : 20 diperoleh bahwa 1.460 data untuk pelatihan dan 366 data untuk pengujian. Hasil akurasi penelitian ini dianalisis melalui *Confusion Matrix*, dengan nilai *accuracy* yaitu sebesar 71%, hasil dari *precision* yaitu sebesar 56%, hasil dari *recall* yaitu sebesar 71%, dan hasil dari *f1-score* yaitu sebesar 61%. Berdasarkan kriteria performa klasifikasi yang merujuk pada akurasi 0,71 atau 71% dikategorikan sebagai performa yang cukup baik.

DAFTAR PUSTAKA

- Agung, A. S., Fauzi, A. A., Akram, A., & Risal, N. U. R. (2023). Implementasi Teknik Data Mining terhadap Klasifikasi Data Prediksi Curah Hujan BMKG Di Sulawesi Selatan. *Jurnal Tekno Insentif*, 17(1), 22–32.
- Alfani W.P.R., A., Rozi, F., & Sukmana, F. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(1), 155–160. <https://doi.org/10.29100/jipi.v6i1.1910>
- Alkhatib, K., Najadat, H., Hmeidi, I., & Shatnawi, M. K. A. (2013). Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm. *International Journal of Business, Humanities and Technology*, 3(3), 32–44.
- Amelia, Y. R. (2018). *Penerapan Data Mining untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT. Bintang Multi Sarana Palembang)*. Universitas Islam Negeri Raden Fatah.
- Badan Pusat Statistika Kabupaten Cilacap. (2025). Kabupaten Cilacap dalam Angka (Cilacap Regency in Figures) 2025. *BPS-Statistics of Cilacap Regency Catalogue :1102001.3301*, 39.
- Darmayanti, I., Subarkah, P., Anunggilarsa, L. R., & Suhaman, J. (2021). Prediksi Potensi Siswa Putus Sekolah Akibat Pandemi Covid-19 Menggunakan Algoritme K-Nearest Neighbor. *JST (Jurnal Sains Dan Teknologi)*, 10(2), 230–238. <https://doi.org/10.23887/jstundiksha.v10i2.39151>
- Fatihah, V., Arifin, A. Z., & Putro, S. (2025). *Peramalan Curah Hujan di Kabupaten Tuban Menggunakan Algoritma KNN*. 7(4), 297–306.
- Harun, R., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes. *MISI (Jurnal Manajemen Informatika Dan Sistem Informasi)*, 3(1), 8–15.
- Limahelu, E. L., Herwanto, B., Thenu, Y. M., Umar, N., Putri, A. P. S., Subagya, M., Rachman, A. J., Antariksa, A. Y., & Hakim, L. (2020). Buletin Edisi Juli 2020 Stasiun Meteorologi Umbu Mehang Kunda Sumba Timur. *Stasiun Meteorologi Umbu Mehang Kunda Sumba Timur*.
- Lumbanraja, F. R., Sani, R. S., Kurniawan, D., & Irawati, A. R. (2019). Implementasi Metode Support Vector Machine Dalam Prediksi Persebaran Demam Berdarah Di Kota Bandar Lampung. *Jurnal Komputasi*, 7(2), 63–73. <https://doi.org/10.23960/komputasi.v7i2.2426>
- Nanda, D. M., Pudjiantoro, T. H., & Sabrina, P. N. (2022). Metode K-Nearest Neighbor (KNN) dalam Memprediksi Curah Hujan di Kota Bandung. *Snestik II*, 387–393.
- Rachma, C. A. (2022). *Implementasi Algoritma K-Nearest Neighbor Dalam Penentuan Klasifikasi*

- Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur.* Universitas Islam Negeri Maulana Malik Ibrahim.
- Rakhmalia, R. I. (2018). *Perbandingan Hasil Metode Naïve Bayes Classifier Dan Support Vector Machine Dalam Klasifikasi Curah Hujan (Studi Kasus : Curah Hujan di Jawa Timur Tahun 2013 - 2017).*
- Rangkuti, M. Y. R., Alfansyuri, M. V., & Gunawan, W. (2021). Penerapan Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia. *Jurnal Teknik Dan Sains*, 2(2), 11–16. <https://doi.org/10.36761/hexagon.v2i2.1082>
- Rina. (2023). *Algoritma k-Nearest Neighbor (KNN): Penjelasan dan Implementasi untuk Klasifikasi Kanker.*
- Sholeh, M., Andayati, D., & Rachmawati, R. Y. (2022). Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes. *Jurnal TeIKa*, 12, 77–87.
- Sukamto, S., Adriyani, Y., & Aulia, R. (2020). Prediksi Kelompok UKT Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *JUITA: Jurnal Informatika*, 8(1), 121–130. <https://doi.org/10.30595/juita.v8i1.6267>
- Tjasyono, B. (2012). *Meteorologi Indonesia Volume I Karakteristik dan Sirkulasi Atmosfer: Vol. I* (D. L. S. Suratno, Welly Fitria (ed.); 4th ed.). Badan Meteorologi Klimatologi dan Geofisika.
- Wulandari, H. S., & Aziz, R. Z. A. (2025). Implementasi Model LSTM , CNN + LSTM Hybrid , dan Transformer untuk Prediksi Cuaca Harian Berbasis Data Multivariat. *Building of Informatics, Technology and Science (BITS)*, 7(2), 1005–1016. <https://doi.org/10.47065/bits.v7i2.7655>