

# KLASIFIKASI CURAH HUJAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (KNN) DI SULAWESI TENGAH

Moh.Fajrin Sigit Aldy\*<sup>1)</sup>, Dwi Shinta Angreni<sup>2)</sup>, Mohammad Yazdi Pusadan<sup>3)</sup>, Wirdayanti<sup>4)</sup>

1. Sistem Informasi, Fakultas Teknik, Universitas Tadulako, Indonesia
2. Sistem Informasi, Fakultas Teknik, Universitas Tadulako, Indonesia
3. Teknik Informatika, Fakultas Teknik, Universitas Tadulako, Indonesia
4. Teknik Informatika, Fakultas Teknik, Universitas Tadulako, Indonesia

## Article Info

**Kata Kunci:** Curah Hujan; Klasifikasi; *K-Nearest Neighbor*; Prediksi; Sulawesi Tengah

**Keywords:** *Central Sulawesi*; *Classification*; *K-Nearest Neighbor*; *Prediction*; *Rainfall*;

## Article history:

Received 2 September 2024

Revised 11 Oktober 2024

Accepted 4 November 2024

Available online 1 December 2024

## DOI :

<https://doi.org/10.29100/jipi.v9i4.5643>

\* Corresponding author.

Corresponding Author

E-mail address:

mrialdy31@gmail.com

## ABSTRAK

Provinsi Sulawesi Tengah memiliki letak yang berdekatan dengan garis khatulistiwa, hal ini dapat mempengaruhi perubahan iklim di daerah tersebut salah satunya curah hujan. Perubahan curah hujan yang tidak menentu mengakibatkan timbulnya bencana seperti banjir yang dapat mempengaruhi gerak aktivitas masyarakat sehari-hari. Salah satu hal yang perlu dilakukan untuk mengantisipasi dengan prediksi cuaca. Pemanfaatan metode data mining dapat membantu dalam melakukan prediksi serta akurasi data dengan baik. Penelitian ini menggunakan dataset BMKG di Provinsi Sulawesi Tengah yang dikumpulkan dari 1 Januari 2019 sampai 31 Oktober 2023 serta klasifikasi dibagi menjadi 5 kelas menggunakan algoritma K-Nearest Neighbor (KNN). Tujuan penelitian ini memperoleh informasi dengan mengelompokkan data guna memprediksi curah hujan di BMKG Sulawesi Tengah. Hasil evaluasi menunjukkan bahwa nilai  $K = 23$  dengan akurasi sebesar 83,0%, dan algoritma K-Nearest Neighbor (KNN) memiliki kinerja yang cukup baik dalam melakukan klasifikasi cuaca.

## ABSTRACT

The Central Sulawesi Province is located near the equator, which can affect climate changes in the region, one of which is rainfall. Uncertain changes in rainfall can lead to disasters such as floods that can affect the daily activities of the community. One of the things that needs to be done to anticipate this is weather prediction. The utilization of data mining methods can assist in making predictions and ensuring data accuracy. This research utilizes BMKG datasets in the Central Sulawesi Province collected from January 1, 2019, to October 31, 2023, with classification divided into 5 classes using the K-Nearest Neighbor (KNN) algorithm. The aim of this research is to obtain information by grouping data to predict rainfall in BMKG Central Sulawesi. The evaluation results show that the value of  $K = 23$  with an accuracy of 83,0%, and the K-Nearest Neighbor (KNN) algorithm performs fairly well in weather classification.

## I. PENDAHULUAN

SULAWESI Tengah merupakan salah satu provinsi di Indonesia yang terletak di tengah pulau Sulawesi. Dengan letak geografi yang berdekatan dengan garis khatulistiwa dapat mempengaruhi kondisi iklim[1] Salah satu perubahan iklim yang terjadi di Provinsi Sulawesi Tengah adalah curah hujan. Berdasarkan data bencana dari Badan Penanggulangan Bencana (BPBD) Sulawesi Tengah dengan persentase bencana dari tahun 2020, 2021, 2022, 2023 yang di sebabkan hujan adalah bencana banjir, tanah longsor, dan puting beliung. Hal ini memberikan dampak kerugian baik secara materialistis serta dapat mempengaruhi gerak aktivitas masyarakat sehari-hari. Beberapa cara dapat dilakukan untuk memperkecil dampak ditimbulkan salah satunya dengan melakukan prediksi cuaca. Dengan prediksi cuaca yang akurat dapat memberikan dampak baik dalam sektor kegiatan manusia[2]

Cuaca adalah keadaan atmosfer pada wilayah yang terbatas dan dalam periode waktu yang singkat[3] Akan tetapi setiap daerah memiliki jenis iklim yang dipengaruhi oleh garis lintang[4] Faktor yang dapat mempengaruhi keadaan cuaca berubah-ubah, antara lain suhu, angin, tekanan, kelembapan, lamanya penyinaran matahari, temperatur maximum, temperatur minimum, dan curah hujan. Dari faktor tersebut dapat menjadi sebuah atribut penentu dalam

melakukan prediksi cuaca. Dengan kata lain cuaca dapat diartikan sebagai keadaan atmosfer dalam waktu tertentu disuatu daerah atau tempat dan menjadi salah satu dari banyak variabel yang menjadi penentu kondisi iklim[5]

Kemajuan dan perkembangan ilmu di bidang data science dan teknologi yang dapat menghasilkan informasi dalam memprediksi perubahan cuaca. Informasi tentang prediksi cuaca saat ini sangat mudah diakses melalui Badan Meteorologi Klimatologi dan Geofisika (BMKG) sebuah Lembaga yang berfokus dalam mengelola aspek terkait dengan kondisi iklim dan cuaca yang tengah berlangsung, namun untuk menilai keakuratan data tersebut, diperlukan suatu metode khusus.

*Data Mining* merupakan sebuah proses metode pembelajaran mesin dengan model statistik dan algoritma matematika dalam mengumpulkan data yang besar serta menggunakan teknik analisis data tingkat lanjut[6]. Estimasi, prediksi, klasifikasi, clustering, dan asosiasi merupakan lima aspek penting dalam *data mining*[7] Pada Penelitian ini akan mencoba melakukan klasifikasi data curah hujan berdasarkan ketentuan BMKG pada probabilitik curah hujan harian yang terdiri dari lima kelas, berawan/tidak hujan, hujan ringan, hujan sedang, hujan lebat dan hujan sangat lebat. CRISP-DM akan menjadi panduan utama dalam melakukan penelitian data mining ini. Proses penambangan data bisa dilakukan dengan berbagai cara, termasuk dengan menggunakan metode CRISP-DM[8]

CRISP-DM merupakan standar yang diakui secara luas dan digunakan secara independen di berbagai industri/penelitian untuk menjalankan proyek penambangan data dari awal hingga akhir[9]. Metodologi ini mengadopsi pendekatan terstruktur yang sangat efektif dan telah terbukti handal untuk perancangan proyek data mining serta metode ini sangat populer karena praktis, fleksibel, dan aplikatif, mampu menyelesaikan masalah bisnis yang kompleks[10]. Metode CRIPS-DM menyajikan setiap langkah dan tugas secara rinci melalui dokumentasi resmi dan contoh penerapannya di dunia nyata [11]. Selain melakukan pemilihan metode, algoritma yang digunakan juga memiliki pengaruh besar terhadap metode *data mining*.

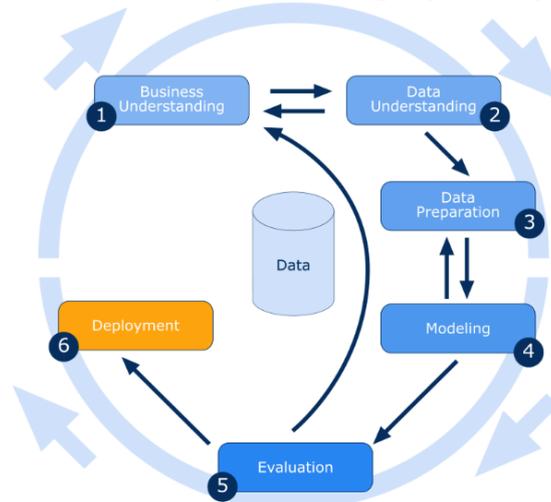
Ada banyak penelitian yang telah membahas tentang algoritma *data mining*, Pada tahun 2023 penelitian dilakukan oleh [12] melakukan prediksi curah hujan menggunakan metode *categorical boosting (CATBOOSTS)*. Model ini mendapatkan hasil prediksi curah hujan serta evaluasi berdasarkan *Confusion Matrix* sebesar 94,22%. Penelitian lainnya pada tahun 2021 yang dilakukan oleh [13] melakukan perbandingan algoritma dari neural network yaitu Backpropagation (BPNN) dengan Recurrent neural network (RNN) dengan menggunakan data cuaca dari tahun 2010-2019 untuk memprediksi curah hujan dan kecepatan angin dan di dapatkan hasil bahwa BPNN memiliki akurasi yang baik dalam prediksi curah hujan dan RNN lebih baik dalam memprediksi kecepatan angin.

Penelitian yang lain melakukan implementasi data mining untuk memprediksi data curah hujan pada tahun 2023 yang dilakukan oleh [14] dengan lokasi penelitian di Sulawesi Selatan menggunakan data sebanyak 1036 dengan menggunakan algoritma *K-Nearest Neighbor (KNN)* di dapatkan hasil akurasi dari model menggunakan *Confusion Matrix* sebesar 82,1% dengan nilai  $K=3$ . Selain itu penelitian pada tahun 2021 yang dilakukan [15] membuat prediksi curah hujan berpotensi banjir dengan mengimplementasikan CRIPS-DM dengan menggunakan algoritma *CART*. Penelitian ini menggunakan beberapa parameter seperti kelembapan, suhu, lama penyinaran matahari, dan juga curah hujan. Dari hasil pengujian algoritma ini memiliki kinerja yang cukup baik dan memberikan akurasi sebesar 89,4% dengan jumlah prediksi yang benar terdiri dari 110 data dari total keseluruhan data uji yaitu 123 data.

Berdasarkan dari penelitian yang telah dilakukan sebelumnya menunjukkan ada banyak algoritma yang dapat digunakan. Pada penelitian ini akan menggunakan salah satu algoritma yang menentukan kelas dari suatu data baru berdasarkan kelas data yang ada dalam tetangga terdekatnya yaitu algoritma *K-Nearest Neighbor (KNN)*. Penggunaan *K-Nearest Neighbor (KNN)* dalam penelitian ini dikarenakan algoritma tersebut dapat unggul dalam menangani data pelatihan dengan banyak noise dan ukuran yang besar secara stabil dan efektif [16]. Tujuan penelitian ini melakukan klasifikasi curah hujan di Sulawesi Tengah dengan mengimplementasikan metode CRIPS-DM serta mengetahui hasil kinerja dari algoritma dengan memanfaatkan data cuaca dari tahun 2019 sampai tahun 2023 dengan total 1764 data, dengan parameter yang digunakan terdiri dari suhu minimum (TN), suhu maximum ( $^{\circ}\text{C}$ ), curah hujan (RR), suhu rata-rata ( $^{\circ}\text{C}$ ), kelembapan rata-rata (RH\_avg), lamanya penyinaran matahari (ss), kecepatan angin maximum, arah angin saat kecepatan maximum (dd\_x), kecepatan angin rata-rata (ff\_avg). Selain memberikan pengetahuan dalam melakukan klasifikasi dengan algoritma *K-Nearest Neighbor (KNN)*, Hasil dari penelitian ini dapat memeberikan prediksi guna melakukan tindakan pencegahan seperti peringatan dini yang lebih efisien terhadap bencana alam seperti banjir, tanah longsor, dan angin kencang, serya memungkinkan pemerintah dan masyarakat untuk melakukan tindakan pencegahan dengan lebih awal, dan mengurangi dampak buruk yang mungkin timbul.

## II. METODE PENELITIAN

Dalam penelitian ini menggunakan salah satu metode yang dikenal dalam *data mining* yaitu CRIPS-DM. CRIPS-DM (*Cross-Industry Standard Process for Data Mining*) dianggap sebagai kerangka kerja yang sangat efektif untuk proses *knowledge discovery process in databases* (KDD) [17]. Metodologi ini memiliki enam tahapan terstruktur yang terdiri dari *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Dalam melakukan enam tahapan di atas, penelitian ini menggunakan pemrograman dengan bahasa Python yang banyak digunakan untuk menjalankan *machine learning* seperti klasifikasi, clustering dan regresi, serta Google Colaboratory sebagai text editor. Tahapan metodologi dijelaskan pada Gambar 1.



Gambar. 1. Alur Metode CRIPS-DM

### A. *Business Understanding*

Tahap ini berfokus dalam memahami dari segi tujuan serta keperluan dalam sebuah bisnis atau penelitian serta perlu mendefinisikan masalah yang akan dipecahkan, kemudian memastikan rancangan penelitian selaras dengan tujuan data mining.

### B. *Data Understanding*

Tahap ini melakukan pengumpulan data, eksplorasi data serta menganalisis data awal. Dengan tujuan untuk memahami karakteristik dan mengidentifikasi data secara keseluruhan, agar dapat membuat keputusan yang lebih tepat dalam tahap selanjutnya.

### C. *Data Preparation*

Pada proses ini melakukan persiapan data sebelum data digunakan pada proses modeling. Tahapan yang dilakukan antara lain pembersihan data (*Data Cleaning*), penyaringan data (*Data Selection*), dan perubahan data (*Data Transformation*). Tahapan ini dilakukan agar dapat memudahkan dalam proses pemodelan.

### D. *Modeling*

Tahap ini menggunakan *Machine Learning* untuk memilih teknik data mining, alat bantu data mining, dan algoritma data mining yang tepat. Pada penelitian ini menggunakan model klasifikasi *K-Nearest Neighbor* (KNN). Pada dasarnya, Algoritma *K-Nearest Neighbor* digunakan untuk mengelompokkan objek dengan membandingkannya dengan tetangga terdekat atau yang memiliki nilai ketidaksamaan paling sedikit dalam data pelatihan [18]

Algoritma KNN terdiri dari tiga tahap berurutan, yaitu: Pertama Menentukan nilai awal *K* pada proses pelatihan, kemudian menghitung jarak antara data baru dan data pelatihan menggunakan metrik jarak selama pengujian, terakhir memilih *K* tetangga terdekat dan menentukan kelas data baru berdasarkan mayoritas kelas dari tetangga tersebut Berikut penghitungan jarak tersebut pada Persamaan (1).

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Di mana :

$d(x_i, x_j)$  = Jarak *Euclidean*

$(x_i, x_j)$  = *record* ke-*m*, *record* ke-*n*

$(a_r)$  = data ke-r  
 $m, n = 1, 2, 3, \dots, n$   
 $n$  = dimensi objek

Ada banyak perhitungan jarak yang dimiliki Algoritma KNN namun pada penelitian ini menggunakan Penghirtungan jarak metric euclidiencia yang meruapakan perhitungan yang paling umum digunakan. Perhitungan jarak euclidiencia ini memiliki akurasi yang cukup baik dalam memberikan hasil evaluasi dari model [19].

### E. Evaluation

Tahap ini melakukan pengujian dari model yang telah dibuat untuk melihat performa yang dihasilkan oleh algoritma KNN. Namun sebelum itu perlu dilakukan pengukuran perlu melakukan pembagian dengan metode pembagian *train-test split* antara data latih sebanyak 80% dan data uji sebanyak 20%. Kegunaan dalam pembagian data ini sangat perlu guna menilai kinerja model machine learning. Data latih digunakan untuk melatih model, sementara data uji digunakan untuk menguji model yang telah dilatih, memastikan model dapat melakukan prediksi dengan baik pada data baru yang belum pernah dilihat sebelumnya. Langkah ini membantu mengukur akurasi dan mencegah overfitting[20].

Tahap pengukuran menggunakan metode analitik prediktif yaitu Confusion Matrix, atau sering disebut matriks kesalahan, sebuah tabel yang memberikan gambaran umum tentang seberapa baik model klasifikasi bekerja. Perbandingan tersebut digunakan untuk perhitungan sejumlah metrik evaluasi, termasuk pada hasil nilai akurasi (accuracy), presisi (precision), recall, dan skor F1, atau sering disebut sebagai F-Measure, berikut perhitungan persamaannya.

#### 1) Akurasi (Accuracy)

Mengukur sejauh mana model klasifikasi dapat membuat prediksi yang benar secara keseluruhan. Dihitung dengan membagi jumlah prediksi yang benar (True Positive dan True Negative) dengan total jumlah prediksi. Berikut rumus akurasi ditunjukkan pada Persamaan (2).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

Di mana :

TP = True Positive yaitu jumlah data benar dan hasil prediksi benar.

TN = True Negative yaitu jumlah data benar namun hasil prediksi salah.

FP = False Positive yaitu jumlah data salah dan hasil prediksi salah.

FN = False Negative yaitu jumlah data salah namun hasil prediksi benar.

#### 2) Presisi (Precision)

Mengukur seberapa banyak dari prediksi positif model yang benar. Dihitung dengan membagi True Positive dengan jumlah semua prediksi positif (True Positive dan False Positive). Berikut rumus presisi ditunjukkan pada Persamaan (3).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

#### 3) Recall

Mengukur sejauh mana model dapat menemukan semua instance dari kelas positif yang sebenarnya. Dihitung dengan membagi true positive dengan jumlah semua instance kelas positif (true positive dan false negative). Berikut rumus recall ditunjukkan pada persamaan (4).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

#### 4) F1-Score

Nilai rata-rata harmonis antara presisi dan recall. Memberikan keseimbangan antara kedua metrik tersebut. Dihitung dengan menggunakan rumus  $2 * (Presisi * Recall) / (Presisi + Recall)$ . Berikut rumus F1-Score ditunjukkan pada Persamaan (5).

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \quad (5)$$

Untuk menentukan nilai K terbaik dengan akurasi tertinggi, dilakukan iterasi dan dicari akurasi untuk setiap nilai K yang dimulai.

#### F. Deployment

Tahap ini melakukan publikasi artikel jurnal dan presentasi hasil penelitian berdasarkan hasil model yang didapatkan.

### III. HASIL PEMBAHASAN

#### A. Business Understanding

Penelitian ini menggunakan data mining untuk mengeksplorasi pengetahuan tentang data curah hujan dan intensitas hujan dengan berfokus langsung dalam melakukan klasifikasi data curah hujan dengan menggunakan algoritma K-Nearest Neighbor yang dapat unggul dalam menangani data pelatihan dengan banyak noise dan ukuran yang besar secara stabil dan efektif guna mendapatkan hasil klasifikasi dan mengetahui parameter apa saja yang mempengaruhi intensitas curah hujan di Provinsi Sulawesi Tengah.

#### B. Data Understanding

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari website pusat data BMKG dengan alamat : <https://dataonline.bmkg.go.id/home> yang merupakan hasil pengamatan BMKG Stasiun Meteorologi Kelas II Mutiara Sis Al-Jufrie Palu. Data berisikan data cuaca harian, lima tahun kebelakang terhitung dari 1 Januari 2019 sampai 31 Oktober 2023 dengan total data yang dikumpulkan sebanyak 1764 data dan memiliki 11 atribut di dalam format (.xls).

Adapun 11 atribut didalam dataset tersebut terdiri atas tanggal, suhu minimum (Tn), suhu maximum (Tx), temperatur rata-rata (Tavg), kelembaban rata-rata (RH\_avg), curah hujan (RR), lamanya penyinaran matahari (ss), kecepatan angin maximum (ff\_x), arah angin saat kecepatan maximum (ddd\_x), kecepatan angin rata-rata (ff\_avg) dan arah angin terbanyak (ddd\_car). Sedangkan untuk penentuan kelas klasifikasi hujan berdasarkan ketentuan probabilistik BMKG. Berikut menampilkan Tabel 1 kategori kelas curah hujan dan Tabel 2 dataset BMKG.

Tabel I  
Kategori Kelas Curah Hujan

Kelas	Kategori	Intensitas Curah Hujan (mm/hari)
1	Berawan/Tidak Hujan	0
2	Hujan Ringan	0.5 – 20
3	Hujan Sedang	20 – 50
4	Hujan Lebat	50 – 100
5	Hujan Sangat Lebat	100 – 150

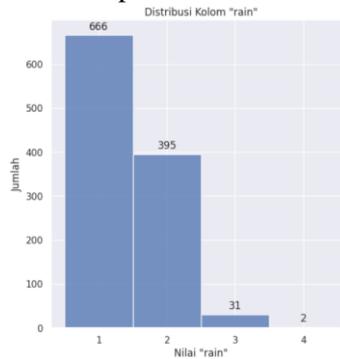
Tabel II  
Dataset BMKG

No	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
1	01-01-2019	24	35.2	28.2	73	0	4	5	340	2	S
2	02-01-2019	23.4	28	25.3	85	0	0	4	350	1	S
3	03-01-2019	22.4	34	27.4	80	3.7	0.2	7	350	2	N
4	04-01-2019	24	34.2	28.1	81	1.8	4.9	7	350	2	C
5	05-01-2019	23	34	27.8	76	0	4.4	9	360	2	C
6	06-01-2019	23.4	35.2	28.2	73	0	4	8	330	3	N
7	07-01-2019	24.2	35.2	28.9	73	0	8.5	8	350	3	S
8	08-01-2019	24.6	34.2	28.6	78	8888	8.3	6	360	2	N
...	....	...	...	...	...	...	...	...	...	...	...
1764	31-10-2023	25.2	36.6	30.5	69	0	9.3	7	360	3	N

#### C. Data Preparation

Dalam pembersihan data (Data Cleaning), dengan dataset yang digunakan sebanyak 1764 data, namun dalam tahap ini data memiliki beberapa data bernilai Null, data yang tidak terukur, serta data yang tidak dilakukan pengukuran, hal ini biasa di sebut dengan Missing Value. Untuk menangani hal tersebut, maka dilakukan penghapusan data dikarenakan data yang didapatkan memiliki banyak nilai null (kosong). Tujuan dilakukan penghapusan untuk mencegah terjadinya distorsi atau bias dalam hasil analisis. Beberapa atribut juga dilakukan penghapusan dikarenakan, type data yang dimiliki adalah string dan object yang dimana dimiliki oleh atribut Tanggal dan Arah Angin (ddd\_car). Setelah melalui proses pembersihan data didapatkan sebanyak 1094 data dengan 10 atribut yang tersimpan dalam format (.xls).

Selanjutnya penyaringan data (Data Selection) tahap ini menyaring atau memilih atribut serta variabel data yang akan digunakan seperti suhu minimum (TN), suhu maximum ( $^{\circ}\text{C}$ ), curah hujan (RR), suhu rata-rata ( $^{\circ}\text{C}$ ), kelembaban rata-rata (RH\_avg), lamanya penyinaran matahari (ss), kecepatan angin maximum, arah angin saat kecepatan maximum (dd\_x), kecepatan angin rata-rata (ff\_avg) yang akan menjadi variabel independen berisi atribut yang mempengaruhi faktor dari curha hujan dan atribut rain yang berisikan data nilai hasil klasifikasi. Berikut tampilan grafik klasifikasi pada atribut rain pada Gambar 2.



Gambar 2. Grafik Klasifikasi Atribut Rain

Dari tampilan Gambar 4 dapat diketahui bahwa hasil klasifikasi yang akan digunakan untuk proses algoritma KNN, memberikan 666 data pada kelas 1 yakni (Berawan atau tidak hujan), 395 data pada kelas 2 (hujan ringan), 31 data pada kelas 3 (hujan sedang) dan 2 data untuk kelas 4 (hujan lebat).

Kemudian tahap perubahan data (Data Transformation) data yang telah melewati tahap sebelumnya akan dilakukan perubahan bentuk format (.xls) menjadi format (.csv) kemudian data tersebut disimpan dan dimasukkan kedalam text editor yaitu aplikasi Google Collaboratory menggunakan bahasa python. Tampilan data setelah dilakukannya pada proses preparation di tunjukan pada Tabel 3.

Tabel III  
 Dataset Setelah Preparation

No	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	Rain
1	24	35.2	28.2	73	0	4	5	340	2	1
2	23.4	28	25.3	85	0	0	4	350	1	1
3	22.4	34	27.4	80	3.7	0.2	7	350	2	2
4	24	34.2	28.1	81	1.8	4.9	7	350	2	2
5	23	34	27.8	76	0	4.4	9	360	2	1
6	23.4	35.2	28.2	73	0	4	8	330	3	1
7	24.2	35.2	28.9	73	0	8.5	8	350	3	1
8	24.6	34.2	28.6	78	8888	8.3	6	360	2	2
...	...	...	...	...	...	...	...	...	...	...
1094	25.2	36.6	30.5	69	0	9.3	7	360	3	1

#### D. Modeling

Tahapan pembuatan model machine learning penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) dalam melakukan klasifikasi data curah hujan di Provinsi Sulawesi Tengah. Dengan kemampuan menemukan jarak terdekat antara data yang sedang dievaluasi dan k tetangga terdekatnya dalam dataset pelatihan. Dalam membantu proses pembuatan model ini, menggunakan tools berupa teks editor yaitu Google Collaboratory dengan menggunakan bahasa pemrograman Python. Tahapan ini akan menampilkan hasil kinerja serta informasi klasifikasi dari model KNN yang telah dibuat.

Dengan melakukan pembagian data dari total keseluruhan data yang di dapatkan dalam proses preparation sebelumnya sebanyak 1094 data, dan dilakukan pembagian data menggunakan *train-test split* dengan data latih 80% atau 875 data dan data uji 20% sebanyak 219 data dengan menggunakan *sklearn model\_selection*.

#### E. Evaluasi

Setelah melakukan modeling tahapan ini akan menampilkan hasil kinerja serta informasi klasifikasi dari model KNN yang telah dibuat.

Proses klasifikasi dilakukan berdasarkan tahapan awal dengan menentukan nilai K awal. Pada penentuan nilai K awal dari proses pelatihan dilakukan proses eksperimen dengan melihat perubahan dari hasil model pada tahap pelatihan.

Pada penelitian ini di dapatkan dari nilai K awal pada proses pelatihan adalah 3 dan juga dilakukan pengukuran

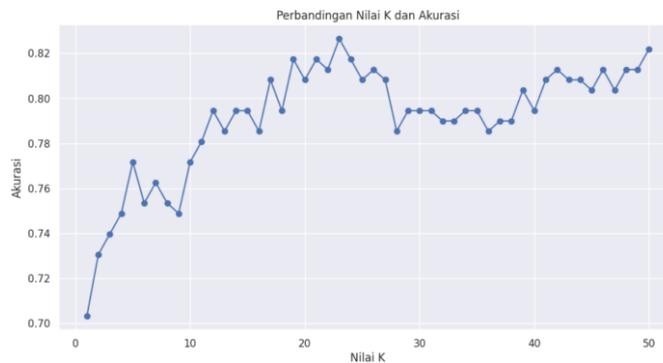
yang terhadap data latih menggunakan  $K = 3$  sebesar 78,0%, kemudian pengujian dilakukan untuk melihat akurasi terhadap data uji menggunakan  $K = 3$  dan di dapatkan akurasi sebesar 74,0% dengan ditampilkan hasil akurasi didalam confusion matrix classification report pada Tabel IV.

Tabel IV  
 Hasil Matrix Classification

Kelas	Precision	Recall	F1-Score	Support
1	0.82	0.83	0.83	150
2	0.57	0.60	0.58	62
3	0.00	0.00	0.00	6
4	0.00	0.00	0.00	1
accuracy			0.74	219
marco avg	0.35	0.36	0.35	219
Weighted avg	0.72	0.74	0.73	219

Dari Gambar 3, dapat diketahui bahwa pada pembagian 5 kelas terdapat Kelas 1 (Berawan) dengan nilai F1-Score yang dimiliki 83%, kelas 2 (hujan ringan) dengan 58%, Kelas 3 (hujan sedang) 00% begitu pula dengan kelas 4 (hujan lebat) dan kelas 5 (hujan sangat lebat) tidak menampilkan hasil yang signifikan dikarenakan data yang dimiliki tidak banyak mengakibatkan model tidak dapat untuk melakukan perhitungan dari kelas tersebut.

Dengan melihat hasil yang diberikan, peneliti ingin mengetahui akurasi terbaik yang dapat dihasilkan oleh model. Maka dilakukan proses untuk mencari nilai  $K$  dengan akurasi terbaik dengan cara membandingkan nilai  $K$  secara keseluruhan menggunakan range. Nilai  $K$  yang digunakan menggunakan range 1 sampai 50. Hasil perbandingan nilai  $K$  ditampilkan dalam Gambar 3.



Gambar 3. Grafik K Range 1 sampai 50

Pada Gambar 3, menunjukan grafik tertinggi dimiliki  $K = 23$  kemudian dilakukan pengujian akurasi untuk melihat secara detail pada confusion matrix classification report dan didapatkan adalah  $K = 23$  dengan akurasi terbaik 83,0%. Hasil pengujian ditampilkan pada Tabel V.

Tabel V  
 Hasil Matrix Classification  $K = 23$

Kelas	Precision	Recall	F1-Score	Support
1	0.85	0.95	0.89	150
2	0.76	0.63	0.69	62
3	0.00	0.00	0.00	6
4	0.00	0.00	0.00	1
accuracy			0.83	219
marco avg	0.40	0.39	0.40	219
Weighted avg	0.80	0.83	0.81	219

Model KNN dengan  $K=23$  menunjukkan performa yang lebih baik secara keseluruhan dibandingkan dengan  $K=3$ . Peningkatan akurasi dan weighted averages mengindikasikan bahwa model dengan  $K=23$  lebih handal dan lebih akurat dalam prediksi. Namun, perlu dicatat bahwa model tidak dapat memprediksi kelas dengan sangat sedikit dukungan (kelas 3 dan 4). Hal ini terjadi dikarenakan data yang dimiliki kelas tersebut sangat sedikit.

#### F. Deployment

Setelah melalui berbagai tahapan, pada tahap ini penelitian tidak hanya menyelesaikan penelitiannya, namun dapat memberikan informasi serta manfaat dari hasil penelitian. Dengan melakukan klasifikasi curah hujan menggunakan algoritma KNN dapat membantu memberikan sebuah prediksi cuaca dalam mendukung

pengambilan keputusan guna melakukan tindakan pencegahan seperti peringatan dini yang lebih efisien terhadap bencana alam seperti banjir, tanah longsor, dan angin kencang, memungkinkan pemerintah dan masyarakat untuk melakukan tindakan pencegahan dengan lebih awal, dan mengurangi dampak buruk yang mungkin timbul di Provinsi Sulawesi Tengah.

Penelitian ini memiliki referensi dari penelitian sebelumnya yang dilakukan oleh [14] yang membedakan pada penelitian ini menggunakan dataset yang cukup banyak dengan lokasi yang berbeda, namun penelitian ini dapat memberikan kontribusi bahwa KNN dapat memberikan hasil yang konsisten dengan akurasi yang cukup baik dalam konteks prediksi curah hujan di wilayah yang berbeda. Dengan menggunakan dataset yang lebih besar dan periode waktu yang lebih panjang, penelitian ini menegaskan keandalan KNN dalam kondisi data yang bervariasi.

#### IV. KESIMPULAN

Pada penelitian ini, telah melakukan klasifikasi data curah hujan dengan mengimplementasikan metode CRISP-DM serta menggunakan algoritma KNN. Dari dataset awal yang dimiliki 1764 data dengan 11 atribut didalamnya, dilakukan proses preparation data di dapatkan data sebanyak 1094 dengan 10 atribut yang akan digunakan dalam proses modeling. Dalam menentukan kelas target berdasarkan ketentuan BMKG terdapat 5 kelas yaitu, berawan, hujan ringan, hujan sedang, hujan lebat dan hujan sangat lebat. Dari hasil evaluasi model di dapatkan nilai K awal dalam proses pengujian yaitu  $K = 3$  dengan akurasi 74,0% serta dilakukan pencarian nilai K dengan akurasi terbaik yang di hasilkan model  $K = 23$  dengan akurasi 83%. Penelitian ini dapat disimpulkan bahwa algoritma KNN memiliki akurasi yang cukup baik dalam melakukan klasifikasi curah hujan dengan melihat hasil evaluasi dari akurasi yang diberikan.

Saran untuk penelitian selanjutnya dapat mencoba dalam menguji metode perhitungan jarak lainnya, atau melakukan eksplorasi berupa perbandingan algoritma dengan menggunakan data yang lebih besar dan dapat mengimplementasikan di lokasi yang sama atau berbeda.

#### UCAPAN TERIMA KASIH

Ucapan terima kasih peneliti ucapkan kepada Universitas Tadulako karena sudah memberikan ruang serta kesempatan peneliti dalam menyelesaikan penelitian melalui program Kampus Merdeka (MBKM) dengan mengikuti program Riset MBKM pada Stasiun Meteorologi Kelas II Mutiara Sis Al-Jufrie Palu.

#### V. DAFTAR PUSTAKA

- [1] Solih Alfandy *et al.*, "Analisis Iklim Provinsi Sulawesi Tengah berdasarkan Data Pemantau Cuaca Otomatis BMKG," 2020.
- [2] T. Lattifia, P. Wira Buana, N. Kadek, and D. Rusjyanthi, "Model Prediksi Cuaca Menggunakan Metode LSTM," 2022.
- [3] M. Yulianto, D. Afriyantari, and P. Putri, "Pengembangan Game Edukasi Pengenalan Iklim Dan Cuaca Untuk Siswa Kelas III Sekolah Dasar," *Jurnal Teknik Elektro*, vol. 20, no. 2, pp. 128–133, 2020.
- [4] D. Martia Nanda, T. Hendro Pudjiantoro, P. Nurul Sabrina, and A. Yani, "SNESTIK Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika Metode K-Nearest Neighbor (KNN) dalam Memprediksi Curah Hujan di Kota Bandung," p. 387, 2022, doi: 10.31284/p.snes-tik.2022.2750.
- [5] S. Khairunnisa and M. I. Jambak, "Pengelompokan Cuaca Kota Palembang Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Pola Karakteristik Cuaca," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2352, Oct. 2022, doi: 10.30865/mib.v6i4.4810.
- [6] A. De Wibowo Muhammad Sidik, I. Himawan Kusumah, A. Suryana, M. Artiyasa, and A. Pradiftha Junfithrana, "Gambaran Umum Metode Klasifikasi Data Mining," vol. 2, no. 2, pp. 34–38, 2020.
- [7] B. Poemomo, R. Dewi, and I. Sari, "PENERAPAN DATA MINING UNTUK PRAKIRAAN CUACA DI KOTA MALANG MENGGUNAKAN ALGORITMA ITERATIVE DICHOTOMISER TREE (ID3)," 2017.
- [8] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," in *Procedia CIRP*, Elsevier B.V., 2019, pp. 403–408. doi: 10.1016/j.procir.2019.02.106.
- [9] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.
- [10] A. Pambudi, "PENERAPAN CRISP-DM MENGGUNAKAN MLR K-FOLD PADA DATA SAHAM PT. TELKOM INDONESIA (PERSERO) TBK (TLKM) (STUDI KASUS: BURSA EFEK INDONESIA TAHUN 2015-2022)," *Jurnal Data Mining dan Sistem Informasi*, vol. 4, no. 1, p. 1, Mar. 2023, doi: 10.33365/jdmsi.v4i1.2462.
- [11] H. J. G. Palacios, R. A. J. Toledo, G. A. H. Pantoja, and Á. A. M. Navarro, "A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change," *Advances in Science, Technology and Engineering Systems*, vol. 2, no. 3, pp. 598–604, 2017, doi: 10.25046/aj020376.
- [12] Andrian Febriansyah Istianto, Asep Id Hadiana, and Fajri Rakhmat Umbara, "PREDIKSI CURAH HUJAN MENGGUNAKAN METODE CATEGORICAL BOOSTING (CATBOOST)," 2023.
- [13] D. A. H. Panggabean, F. M. Sihombing, and N. M. Aruan, "PREDIKSI TINGGI CURAH HUJAN DAN KECEPATAN ANGIN BERDASARKAN DATA CUACA DENGAN PENERAPAN ALGORITMA ARTIFICIAL NEURAL NETWORK (ANN)," *SEMINASTIKA*, vol. 3, no. 1, pp. 1–7, Nov. 2021, doi: 10.47002/seminastika.v3i1.237.
- [14] A. S. Agung, A. A. Fauzi, A. A. Nur Risal, and F. Adiba, "Implementasi Teknik Data Mining terhadap Klasifikasi Data Prediksi Curah Hujan BMKG Di Sulawesi Selatan," *Jurnal Tekno Insentif*, vol. 17, no. 1, pp. 22–23, Apr. 2023, doi: 10.36787/jti.v17i1.955.
- [15] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>

- [16] M. F. Nasrullah, R. Rohmat Saedudin, and F. Hamami, "JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Journal homepage: <https://jurnal.stkipgritlungagung.ac.id/index.php/jipi> PERBANDINGAN AKURASI ALGORITMA C4.5 DAN K-NEAREST NEIGHBORS UNTUK KLASIFIKASI CURAH HUJAN BERDASARKAN IKLIM INDONESIA," vol. 9, no. 2, pp. 628–638, 2024, doi: 10.29100/jipi.v9i2.4655.
- [17] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans Knowl Data Eng*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.
- [18] M. Kubat, *An Introduction to Machine Learning*. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-81935-4.
- [19] G. Hendro Martono and N. Sulistianingsih, "Perbandingan Matriks jarak pada Algoritma K-NN untuk Prediksi Penyakit Diabetes Comparison of Distance Matrices in the K-NN Algorithm for Predicting diabetes," *JoMI: Journal of Millennial Informatics*, vol. 2, no. 1, pp. 1–6, 2024.
- [20] A. Novareza, R. Herdiana, and I. Iin, "ANALISIS DATA MINING PADA PRODUKSI IKAN AIR TAWAR DI KECAMATAN PANJALU MENGGUNAKAN METODE C4.5," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 252–258, Feb. 2024, doi: 10.36040/jati.v8i1.8209.