

Implementasi Metode K-Nearest Neighbor Untuk Memprediksi Kondisi Cuaca

Diterima:

10 Juni 2024

Revisi:

10 Juli 2024

Terbit:

1 Agustus 2024

¹Evania Priyanto, ² Erna Daniati, ³ Aidina Ristyawan

¹⁻³Universitas Nusantara PGRI Kediri

¹evaniapriyanto@gmail.com, ²ernadaniati@unpkediri.ac.id,

³adinaristi@unpkediri.ac.id

Abstrak— Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor (KNN) menggunakan aplikasi RapidMiner untuk memprediksi kondisi cuaca. Mengingat cuaca memiliki pola yang tidak menentu, prediksi manual menjadi sangat sulit. Meskipun cuaca tidak dapat diprediksi dengan kepastian absolut, perkiraan masih dapat dibuat. Dalam konteks ini, data mining memungkinkan mesin untuk mengenali dan mempelajari pola data yang kompleks. Oleh karena itu, machine learning dapat digunakan untuk mempelajari pola data cuaca guna melakukan prediksi. Penelitian ini menggunakan enam variabel sebagai kriteria, yaitu tanggal, curah hujan, suhu maksimum, suhu minimum, kecepatan angin, dan kondisi cuaca. Pengujian yang dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) melalui aplikasi RapidMiner dengan nilai K=2 menunjukkan akurasi sebesar 100%, berdasarkan 1462 data yang diambil dari tahun 2012 hingga 2015. Hasil analisis membuktikan bahwa metode KNN yang diterapkan dengan aplikasi RapidMiner efektif dalam memprediksi cuaca.

Kata Kunci— prediksi, cuaca, KNN

Abstract— This study aims to apply the K-Nearest Neighbor (KNN) algorithm using the RapidMiner application to predict weather conditions. Given that weather patterns are unpredictable, manual prediction becomes very difficult. Although weather cannot be predicted with absolute certainty, estimations can still be made. In this context, data mining enables machines to recognize and learn complex data patterns. Therefore, machine learning can be used to study weather data patterns for prediction purposes. This study employs six variables as criteria: date, rainfall, maximum temperature, minimum temperature, wind speed, and weather conditions. Testing conducted using the K-Nearest Neighbor (KNN) algorithm through the RapidMiner application with a K value of 2 showed an accuracy of 100%, based on 1462 data points taken from 2012 to 2015. The analysis results demonstrate that the KNN method implemented with the RapidMiner application is effective in predicting weather conditions..

Keywords— prediction, weather, KNN

This is an open access article under the CC BY-SA License.

Penulis Korespondensi:

Nama Penulis, Erna Daniati

Departemen Penulis, Sistem Informasi

Institusi Penulis, Universitas Nusantara PGRI Kediri

Email: ernadaniati@unpkediri.ac.id

ID Orcid: [<https://orcid.org/0009-0008-9471-4421>]

Handphone: 081335242202

I. PENDAHULUAN

Keadaan udara di atmosfer bersifat fluktuatif dan terus mengalami perubahan seiring berjalannya waktu serta bervariasi pada setiap lokasi yang berbeda. Faktor-faktor seperti curah

hujan, temperatur udara, tingkat kelembapan, dan kecepatan angin menjadi parameter umum yang digunakan untuk mengklasifikasikan kondisi cuaca yang terjadi di suatu wilayah pada hari tertentu [1]. Menentukan kondisi cuaca merupakan aplikasi dari ilmu pengetahuan dan teknologi untuk mengamati keadaan atmosfer bumi. Proses penentuan cuaca menggunakan berbagai metode untuk menghasilkan prediksi cuaca. Penggunaan metode yang akurat dan mutakhir menjadi prioritas utama bagi para peneliti di bidang cuaca dan atmosfer. Untuk mendapatkan hasil prediksi yang lebih baik dan tepat, para peneliti terus mengkaji secara mendalam aspek-aspek atmosfer yang memengaruhi cuaca serta mengembangkan metode-metode baru guna memenuhi kebutuhan akan informasi cuaca dan kondisi atmosfer yang akurat [2]. Di bidang pertanian, fenomena alam menunjukkan peran yang semakin signifikan belakangan ini, terutama melalui intensitas curah hujan yang berfluktuasi setiap harinya [3]. Seiring dengan kemajuan teknologi di bidang data sains, prediksi cuaca masa depan menjadi semakin penting. Badan Meteorologi Klimatologi dan Geofisika (BMKG) bertanggung jawab atas pengelolaan informasi prediksi cuaca di Indonesia. Namun, untuk menentukan tingkat akurasi data prediksi tersebut, diperlukan metode khusus. Salah satu metode yang sesuai adalah data mining. Metode ini menggabungkan matematika, teknik statistik, kecerdasan buatan, dan machine learning untuk mengolah, mengekstraksi, dan mengidentifikasi pengetahuan dari berbagai database, serta memanfaatkan data tersebut untuk menghasilkan informasi berharga. Data mining dapat dikategorikan berdasarkan tugas dan fungsinya, seperti prediksi, klasifikasi, pengelompokan, deskripsi, estimasi, dan asosiasi[4]. Penelitian ini bertujuan memprediksi cuaca berdasarkan dataset yang di dapatkan sebelumnya. Penelitian ini menggunakan metode K-Nearest Neighbor (KNN). KNN termasuk kedalam bagian Data Mining yang mampu mengenali suatu pola data sequensia [5]. Ada beberapa penelitian yang menggunakan metode K-Nearest Neighbor (KNN), Seperti penelitian yang dilakukan oleh Muhammad Yusuf Rizqon Rangkuti dkk., dengan judul "Penerapan Algoritma KNN Dalam Memprediksi dan Menghitung Tingkat Akutasi Data Cuaca di Indonesia" Dengan tingkat akurasi data 90% ke akuratan data testing dengan data training hampir menyentuh sempurna [6]. Dan juga penelitian Iftar Ramadhan dkk., untuk memprediksi bencana gunung berapi dengan tingkat akurasi yang di peroleh 100%[7]. Dari berbagai penelitian tersebut, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (KNN) menghasilkan tingkat akurasi yang tinggi. Terkait hal ini, diusulkan penelitian berjudul "Implementasi Metode K-Nearest Neighbor untuk Memprediksi Kondisi Cuaca" dengan tujuan meningkatkan ketepatan prediksi cuaca. Akurasi informasi cuaca berdampak signifikan pada pengambilan keputusan strategis di sektor-sektor krusial seperti pertanian, transportasi, dan pariwisata. Hasil prediksi diharapkan dapat dimanfaatkan secara efektif oleh pengguna non-teknis dan menjadi upaya untuk mengatasi permasalahan terkait

prediksi cuaca yang akurat. Penggunaan algoritma KNN ini bertujuan untuk meningkatkan efisiensi dan efektifitas pengolahan data [8]. Metode ini memungkinkan kita untuk mempelajari beberapa faktor historis guna meramalkan kejadian di masa mendatang. Dengan demikian, kita dapat lebih siap menghadapi peristiwa yang akan datang [9].

II. METODE

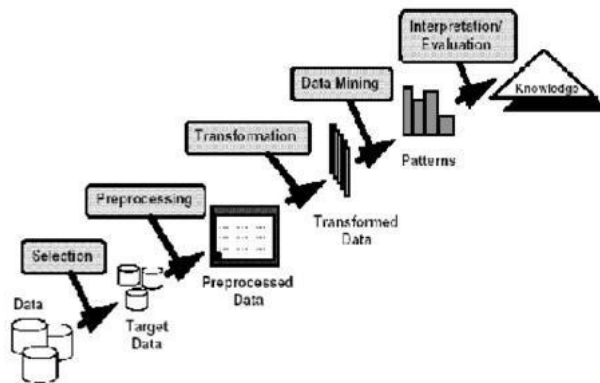
2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data yang diperoleh dari situs web *Kaggle* sebagai berikut:

<https://www.kaggle.com/datasets/ananthr1/weather-prediction>

2.2 KDD

KDD (Knowledge Discovery in Databases) atau penemuan pengetahuan dalam basis data adalah serangkaian tahapan yang dilakukan untuk menemukan informasi yang tersembunyi atau tidak tampak secara eksplisit dari data yang terakumulasi dalam pangkalan data. Prosesnya bertujuan untuk mengekstrak pengetahuan bernilai dari sekumpulan data yang terakumulasi dalam basis data [10].



Gambar 1. Metode knowledge discovery in databases (KDD)

Berikut adalah Langkah- Langkah yang perlu dilakukan dalam penelitian berdasarkan tahapan knowledge discovery in database:

1. Data selection: Dari kumpulan data operasional sebelum memulai proses Knowledge Discovery in Database (KDD). Data yang telah dipilih untuk digunakan dalam proses data mining kemudian disimpan dalam file yang terpisah dari basis data operasional[11].

2. Preprocessing atau Cleaning: Langkah ini bertujuan untuk menghilangkan data yang tidak memiliki nilai (null). Setelah dilakukan pengecekan, tidak ditemukan nilai yang hilang pada data yang akan diproses [12].
3. Transformasi data: Bertujuan untuk memudahkan pengolahan data yang akan diproses. Dalam fase ini, tindakan yang diambil adalah mengubah data yang tidak memiliki entitas yang jelas menjadi bentuk data yang valid atau dapat diproses untuk keperluan Data Mining [13].
4. Data mining: Suatu proses yang diterapkan pada basis data besar dengan memanfaatkan kecerdasan buatan, teknik matematika, statistik, dan machine learning untuk menyimpulkan serta mengidentifikasi informasi berharga dan pengetahuan yang relevan[14].
5. Evaluasi: Tahap evaluasi dilakukan dengan mengamati hasil prediksi menggunakan algoritma K-Nearest Neighbor (KNN). Melalui proses ini, dapat diketahui algoritma mana yang memiliki akurasi lebih baik dalam memprediksi cuaca [15].

III. HASIL DAN PEMBAHASAN

3.1 Pengumpulan data

Langkah awal dalam penelitian ini adalah mengumpulkan data dalam bentuk dataset. Dataset yang digunakan berisi data cuaca yang diperoleh dari situs Kaggle. Dataset tersebut memuat 1.462 catatan data dengan total 6 atribut yang mencakup periode dari 1 Januari 2012 hingga 31 Desember 2015.

Tabel 1. Dataset cuaca

No	Date	Precipitation	Temp_max	Temp_min	Wind	Weather
1	2012-01-01	0.0	12.8	5.0	4.7	drizzle
2	2012-01-02	10.9	10.6	10.6	4.5	rain
3	2012-01-03	0.8	11.7	7.2	2.3	rain
..
1460	2015-12-29	0.0	7.2	0.6	2.6	fog
1461	2015-12-30	0.0	5.6	-1	3.4	sun
1462	2015-12-31	0.0	5.6	-2.1	3.5	sun

Data tersebut mencakup informasi cuaca dengan atribut seperti tanggal, curah hujan, suhu minimum, suhu maksimum, kecepatan angin, dan kondisi cuaca.

3.2 Preprocessing Data

Pra-pemrosesan data merupakan tahapan yang harus dilalui sebelum memasukkan data untuk diklasifikasikan. Pada tahap ini, dilakukan proses pembersihan data di mana field atau atribut yang dianggap tidak mendukung dapat dihilangkan. Selain itu, data dikonfigurasi ulang

untuk menjamin format tetap konsisten. Tahapan ini meliputi penanganan terhadap nilai-nilai yang hilang atau kosong pada data, sebagaimana diperlihatkan pada gambar.

Name	Type	Missing	Statistics	Filter (12 / 12 attributes)	Search for Attributes
Label: weather	Polynomial	0	Least snow (26) Most rain (641)	Values rain (641), sun (640), ... [3 more]	
Prediction: prediction(weather)	Polynomial	0	Least snow (23) Most rain (701)	Values rain (701), sun (608), ... [3 more]	
Confidence: confidence(drizzle)	Real	0	Min 0 Max 1	Average 0.036	
Confidence: confidence(rain)	Real	0	Min 0 Max 1	Average 0.429	
Confidence: confidence(sun)	Real	0	Min 0 Max 1	Average 0.438	
Confidence: confidence(snow)	Real	0	Min 0 Max 1	Average 0.018	
Confidence: confidence(fog)	Real	0	Min 0 Max 1	Average 0.009	
date	Date	0	Earliest date Jan 1, 2012 Latest date Dec 31, 2015	Duration 1460 days	
precipitation	Real	0	Min 0 Max 55.900	Average 3.029	

Gambar 2. Cek Missing Value.

3.3 Transformation

Transformasi data menjadi format yang sesuai untuk diproses oleh algoritma data mining dapat dilihat dalam gambar.

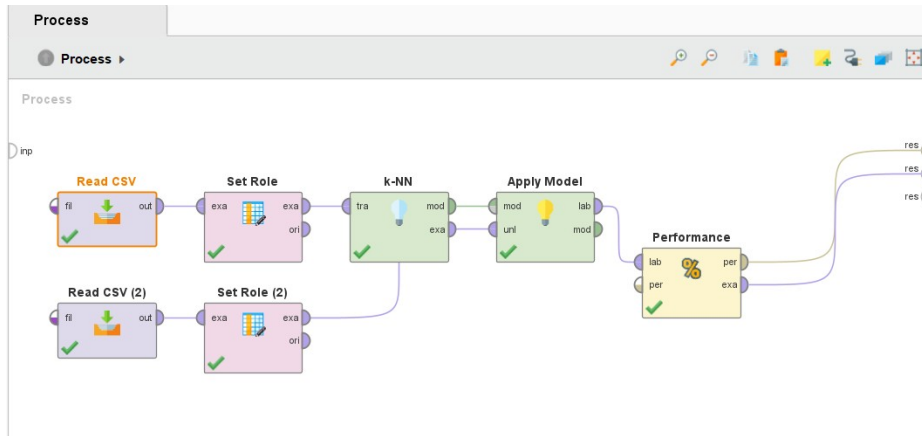
Format your columns.						
date	precipitation	temp_max	temp_min	wind	weather	
date	real	real	real	real	polynomial label	
1	Jan 1, 2012	0.000	12.800	5.000	4.700	drizzle
2	Jan 2, 2012	10.900	10.600	2.800	4.500	rain
3	Jan 3, 2012	0.800	11.700	7.200	2.300	rain
4	Jan 4, 2012	20.300	12.200	5.600	4.700	rain
5	Jan 5, 2012	1.300	8.900	2.800	6.100	rain
6	Jan 6, 2012	2.500	4.400	2.200	2.200	rain
7	Jan 7, 2012	0.000	7.200	2.800	2.300	rain
8	Jan 8, 2012	0.000	10.000	2.800	2.000	sun
9	Jan 9, 2012	4.300	9.400	5.000	3.400	rain
10	Jan 10, 2012	1.000	6.100	0.600	3.400	rain
11	Jan 11, 2012	0.000	6.100	-1.100	5.100	sun
12	Jan 12, 2012	0.000	6.100	-1.700	1.900	sun
13	Jan 13, 2012	0.000	5.000	-2.800	1.300	sun
14	Jan 14, 2012	4.100	4.400	0.600	5.300	snow
15	Jan 15, 2012	5.300	1.100	-3.300	3.200	snow
16	Jan 16, 2012	2.500	1.700	-2.800	5.000	snow
17	Jan 17, 2012	8.100	3.300	0.000	5.600	snow
18	Jan 18, 2012	19.800	0.000	-2.800	5.000	snow

Gambar 3. Transformasi Data.

Pada Gambar 3, transformasi data dilakukan dengan menginisialisasi data yang memiliki nilai-nilai tertentu agar sesuai dengan jenis data yang dibutuhkan oleh algoritma K-Nearest Neighbor. Tipe data yang digunakan mencakup nominal dan real, yang kemudian diubah menjadi label.

3.4 Data Mining

Gambar berikut menunjukkan penggunaan model Algoritma K-Nearest Neighbor untuk memprediksi cuaca.



Gambar 4. RapidMiner

Penelitian ini memanfaatkan perangkat lunak RapidMiner untuk memperoleh pengetahuan, dengan menggunakan operator KNN dari RapidMiner guna menghasilkan model pengujian yang sesuai pada setiap siklus yang telah direncanakan.

3.5 Evaluasi

Pada tahap ini hasil penerapan metode K-Nearest Neighbor dengan $K=2$ untuk klasifikasi Rapid Miner menunjukkan hasil akurasi sebagai berikut:

	true drizzle	true rain	true sun	true snow	true fog	class precision
pred. drizzle	53	0	0	0	0	100.00%
pred. rain	0	641	0	0	0	100.00%
pred. sun	0	0	640	0	0	100.00%
pred. snow	0	0	0	26	0	100.00%
pred. fog	0	0	0	0	101	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	

Gambar 5. Akurasi

Dari hasil percobaan yang dilakukan, dengan 1462 sampel data yang masing-masing mewakili sampel kondisi cuaca *drizzle*, *rain*, *sun*, *snow*, *fog* menghasilkan akurasi sebesar 100%.

IV. KESIMPULAN

Berdasarkan penjelasan tersebut, dapat disimpulkan bahwa penerapan metode klasifikasi menggunakan algoritma K-Nearest Neighbor dengan dukungan Rapid Miner terbukti efektif dalam memprediksi kondisi cuaca berdasarkan data prediksi cuaca dari tahun 2012 hingga 2015. Melalui tahapan-tahapan pengolahan data awal, prapemrosesan data, transformasi data, serta penggunaan operator seperti Read CSV, Set Role, K-Nearest Neighbor, Apply Model, dan Performance (*Classification*), dengan menggunakan K=2 model yang dihasilkan mencapai akurasi 100%. Hasil ini menunjukkan bahwa model K-Nearest Neighbor mampu memberikan prediksi cuaca dengan tingkat akurasi yang sempurna.

UCAPAN TERIMAKASIH

Puji syukur kami haturkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga jurnal berjudul "Implementasi Metode K-Nearest Neighbor untuk Memprediksi Kondisi Cuaca" ini dapat diselesaikan dengan baik. Penulis juga mengucapkan terima kasih yang tulus kepada Universitas Nusantara PGRI Kediri atas dukungan finansial yang diberikan untuk penelitian ini.

DAFTAR PUSTAKA

- [1] A. N. Kirana, B. Nurhakim, S. E. Permana, W. Prihartono, and G. Dwilestari, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer," (JATI) Journal Mahasiswa, vol. 8, no.2, 2024. DOI: <https://doi.org/10.36040/jati.v8i2.8967>, Available: <https://www.kaggle.com/datasets/ananthr1/weather->
- [2] A. Ahmada Akhbar and D. Hartanti, "Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner Untuk Memprediksi Hujan Kota Surakarta.," *SENATIB*, 2023. DOI: <https://doi.org/10.47701/senatib.v3i1.2996>
- [3] R. Harun, K. Chandra Pelangi, and Y. Lasena, "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K Nearest Neighbor (KNN)," vol. 3, no. 1, 2020. DOI: <https://doi.org/10.36595/misi.v3i1.125> Available: <http://e-journal.stmiklombok.ac.id/index.php/misi>
- [4] 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.
- [5] D. Martia Nanda, T. Hendro Pudjiantoro, P. Nurul Sabrina, and A. Yani, "Metode K-Nearest Neighbor (KNN) dalam Memprediksi Curah Hujan di Kota Bandung," *SNESTIK Seminar*

Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika, p. 387. DOI: 10.31284/p.snestik.2022.2750.

- [6] M. Yusuf *et al.*, “Penerapan Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia,” *Hexagon*, vol. 2, no. 2, 2021. DOI: <https://doi.org/10.36761/hexagon.v2i2.1082>
- [7] I. Ramadhan, S. Elmi, and L. Efrizoni, “IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR (KNN) UNTUK PREDIKSI BENCANA GUNUNG BERAPI,” *Jurnal Ilmiah Betrik*, vol. 15, no. 1, 2024. DOI: <https://doi.org/10.36050/betrik.v15i01%20APRIL.250>
- [8] M. Zuhdi Sasongko and E. Daniati, “Perencanaan Sistem Manajemen Evaluasi Belajar Mahasiswa Berbasis Elearning Dengan Pendekatan Berorientasi Objek”. *Jurnal Teknologi Informasi Respati*, vol. 8, no. 2, 2018. DOI: <https://dx.doi.org/10.35842/jtir.v13i2.228>
- [9] A. C. Pradikdo and A. Ristyawan, “Model Klasifikasi Abstrak Skripsi Menggunakan Text Mining Untuk Pengkategorian Skripsi Sesuai Bidang Kajian,” *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 9, no. 2, 2018, DOI: 10.24176/simet.v9i2.2513.
- [10] R. Prasetya, “Penerapan Teknik Data Mining Dengan Algoritma Classification Tree Untuk Prediksi Hujan,” *Jurnal Widya Climago*, vol. 2, no.2, 2020.
- [11] I. Rizky Mahartika and A. Wibowo, “Data Mining Klasterisasi dengan Algoritma K-Means untuk Pengelompokkan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional.” *Prosiding SISFOTEK*, vol. 3, no. 2, pp. 87-91, 2019. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/108>
- [12] F. Harits Muzaki, W. Joko Pranoto, and M. Kalimantan Timur, “Analisis Regresi Linear Dalam Data Mining Untuk Prediksi Sijil Off Di Ksop Kelas I Samarinda,” *Jurnal Ilmu Teknik*, vol. 1, no. 2, pp. 261–266, 2024. DOI: 10.62017/tektonik.
- [13] I. Ali, D. Ade Kurnia, M. A. Pratama, and F. Al Ma’ruf, “Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor”, *Jurnal Informatika*, vol. 5, no. 1, pp. 93-104, 2020. DOI: <https://doi.org/10.51211/itbi.v5i1.1431>. Available: <http://jurnal.kopertipindonesia.or.id/index.php/kopertip>
- [14] R. Bahtiar, “Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor.” *BINER: Jurnal Ilmu Komputer, Teknik dan Multimedia*, vol. 1, no.3, pp. 650-658, 2023. Available: <https://jurnal.publikasitecno.id/index.php/jim203>
- [15] B. Lareno, T. Informatika, and S. PPKIA Tarakanita Rahmawati Tarakan, “Konferensi Nasional Sistem & Informatika,” 2015.