

## K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI KONDISI CUACA MENGGUNAKAN RAPIDMINER

<sup>1</sup>Dela Alaina Lailatul Maqfiroh <sup>2</sup>Zaehol Fatah

<sup>1</sup>Sistem Informasi, Universitas Ibrahimy

<sup>2</sup>Sistem Informasi, Universitas Ibrahimy

<sup>1</sup>delaalainalailatulmaqfiroh@gmail.com, <sup>2</sup>zaeholfatah2@gmail.com

### ABSTRACT

*The use of machine learning technology for weather prediction has become an increasingly important approach in modern meteorology. This study implements the K-Nearest Neighbor (KNN) algorithm to predict air temperature using a dataset from the High-Performance Wireless Research and Education Network (HPWREN). The dataset covers a period of three months with various weather parameters such as air pressure, wind direction, wind speed, and relative humidity, while the rain accumulation and rain duration variables are excluded because they do not have significant values. The research methodology consists of four main stages: data collection, preprocessing using RapidMiner, KNN implementation, and model evaluation. The preprocessing stage includes data normalization and selecting relevant attributes to ensure the quality of model input. The KNN model is configured with optimal parameters, namely  $k = 5$ , weighted vote, and MixedEuclideanDistance to improve prediction accuracy. The dataset is divided into training data (70%) and testing data (30%) to validate model performance. Model evaluation using three standard metrics yielded a Root Mean Squared Error (RMSE) of 3,680, Mean Absolute Error (MAE) of 2,216, and Root Relative Squared Error (RRSE) of 31.5%. These results indicate that the KNN algorithm effectively predicts air temperature with a level of accuracy sufficient for practical applications. This study provides a foundation for the development of more accurate weather prediction systems in the future, with recommendations to expand the data collection period and further optimize model parameters.*

**Keywords:** K-Nearest Neighbor, weather forecast, RapidMiner, Machine Learning, HPWREN

### ABSTRAK

Penggunaan teknologi machine learning untuk prediksi cuaca telah menjadi pendekatan yang semakin penting dalam meteorologi modern. Penelitian ini mengimplementasikan algoritma K-Nearest Neighbor (KNN) untuk memprediksi suhu udara menggunakan dataset dari High Performance Wireless Research and Education Network (HPWREN). Dataset mencakup periode tiga bulan dengan berbagai parameter cuaca seperti tekanan udara, arah angin, kecepatan angin, dan kelembaban relatif, sementara variabel *rain\_accumulation* dan *rain\_duration* dikeluarkan karena tidak memiliki nilai signifikan. Metodologi penelitian terdiri dari empat tahap utama: pengumpulan data, preprocessing menggunakan RapidMiner, implementasi KNN, dan evaluasi model. Tahap preprocessing meliputi normalisasi data dan pemilihan atribut yang relevan untuk memastikan kualitas input model. Model KNN dikonfigurasi dengan parameter optimal yaitu nilai  $k=5$ , *weighted vote*, dan *MixedEuclideanDistance* untuk meningkatkan akurasi prediksi. Dataset dibagi menjadi data training (70%) dan testing (30%) untuk validasi performa model. Evaluasi model menggunakan tiga metrik standar menghasilkan Root Mean Squared Error (RMSE) sebesar 3.680, Mean Absolute Error (MAE) 2.216, dan Root Relative Squared Error (RRSE) 31.5%. Hasil ini menunjukkan bahwa algoritma KNN efektif dalam memprediksi suhu udara dengan tingkat akurasi yang memadai untuk aplikasi praktis. Penelitian ini memberikan landasan untuk pengembangan sistem prediksi cuaca yang lebih akurat di masa depan, dengan rekomendasi untuk memperluas periode pengambilan data dan mengoptimalkan parameter model lebih lanjut.

**Kata Kunci:** K-Nearest Neighbor, Prediksi Cuaca, RapidMiner, Machine Learning, HPWREN

## I. PENDAHULUAN

Prediksi cuaca menjadi kebutuhan yang semakin kritis di era modern ini, terutama dengan meningkatnya frekuensi kejadian cuaca ekstrem akibat perubahan iklim global. Keakuratan prediksi cuaca tidak hanya penting untuk kegiatan sehari-hari masyarakat, tetapi juga vital bagi berbagai sektor seperti pertanian, penerbangan, pelayaran, dan manajemen bencana. Data cuaca yang dikumpulkan oleh HPWREN (High Performance Wireless Research and Education Network) menyediakan parameter-parameter penting seperti tekanan udara, suhu, arah angin, kecepatan angin, dan kelembaban relatif yang dapat dianalisis untuk menghasilkan prediksi cuaca yang lebih akurat [1].

Perkembangan teknologi machine learning membuka peluang baru dalam meningkatkan akurasi prediksi cuaca. K-Nearest Neighbor (KNN) merupakan salah satu algoritma yang menjanjikan untuk kasus ini karena kemampuannya dalam mengenali pola dari data historis cuaca. KNN bekerja dengan prinsip mencari kemiripan antara kondisi cuaca saat ini dengan data historis, memungkinkan prediksi yang lebih presisi dibandingkan metode konvensional. Keunggulan KNN terletak pada kesederhanaannya dalam implementasi namun tetap mampu menghasilkan prediksi yang akurat, terutama untuk data numerik seperti parameter cuaca [2].

RapidMiner hadir sebagai platform yang ideal untuk mengimplementasikan algoritma KNN dalam analisis data cuaca. Software ini menyediakan berbagai tools yang komprehensif untuk preprocessing data, implementasi model, dan evaluasi performa. Dengan menggunakan RapidMiner, proses analisis data cuaca menjadi lebih efisien dan dapat menghasilkan model prediksi yang lebih handal. Platform ini juga memungkinkan visualisasi data yang membantu dalam memahami pola-pola cuaca secara lebih mendalam [3].

Penelitian ini fokus pada pengembangan model prediksi cuaca menggunakan algoritma KNN dengan memanfaatkan dataset HPWREN. Dataset ini mencakup berbagai parameter cuaca yang direkam secara real-time, memberikan dasar yang kuat untuk analisis prediktif. Tujuan utama penelitian adalah menghasilkan model prediksi cuaca yang dapat memberikan informasi akurat tentang kondisi cuaca di masa depan [4]. Hal ini penting mengingat dampak signifikan cuaca terhadap berbagai aspek kehidupan, mulai dari keselamatan publik hingga produktivitas ekonomi.

Implementasi model prediksi cuaca berbasis KNN ini diharapkan dapat memberikan kontribusi nyata dalam meningkatkan akurasi prediksi cuaca. Dengan prediksi yang lebih akurat, berbagai sektor dapat

mengambil keputusan yang lebih tepat dalam menghadapi kondisi cuaca yang akan datang. Selain itu, penelitian ini juga dapat menjadi landasan untuk pengembangan sistem prediksi cuaca yang lebih canggih di masa depan, mengintegrasikan berbagai teknologi machine learning untuk hasil yang lebih optimal [5].

## II. METODE PENELITIAN

### 2.1 Cuaca

Cuaca merupakan kondisi atmosfer yang terbentuk dari interaksi berbagai parameter meteorologi seperti suhu, tekanan udara, kelembaban, dan pergerakan massa udara. Pengamatan dan pemahaman terhadap parameter-parameter ini sangat penting karena cuaca mempengaruhi berbagai aspek kehidupan, mulai dari pertanian hingga transportasi [6].

Dalam kajian meteorologi modern, kondisi cuaca dapat dipahami melalui analisis data yang dikumpulkan dari berbagai sensor dan peralatan pengamatan. Data-data ini memungkinkan para ahli untuk memahami pola cuaca dan melakukan prediksi yang lebih akurat. Pemanfaatan teknologi dan metode analisis data terkini, seperti machine learning, telah membuka peluang baru dalam meningkatkan akurasi prediksi cuaca untuk berbagai keperluan praktis.

### 2.2 KNN

K-Nearest Neighbor (KNN) adalah algoritma machine learning yang bekerja dengan prinsip menentukan klasifikasi atau prediksi berdasarkan kedekatan/kemiripan data baru dengan data yang sudah ada. Metode ini termasuk dalam kategori supervised learning, di mana model belajar dari data training yang sudah memiliki label atau nilai target. Cara kerja KNN relatif sederhana namun efektif. Ketika ada data baru yang ingin diprediksi, algoritma akan mencari  $k$  data terdekat (tetangga) dari data training berdasarkan ukuran jarak tertentu, seperti Euclidean distance. Untuk kasus prediksi numerik seperti prediksi cuaca, nilai prediksi dihitung dari rata-rata nilai  $k$  tetangga terdekat tersebut. Parameter  $k$  menjadi faktor penting yang mempengaruhi performa model, di mana nilai  $k$  yang terlalu kecil dapat membuat model sensitif terhadap noise, sementara nilai  $k$  yang terlalu besar dapat mengaburkan pola-pola lokal dalam data [7].

### 2.3 Rapid Miner

RapidMiner adalah platform analisis data dan machine learning yang menyediakan antarmuka visual untuk memproses dan menganalisis data tanpa perlu menulis kode program. Platform ini dilengkapi dengan berbagai operator untuk preprocessing data, pemodelan, evaluasi, dan visualisasi hasil. Dalam

konteks prediksi cuaca, RapidMiner menyediakan tools yang diperlukan seperti operator untuk membaca data (Read Excel), normalisasi data numerik, pembagian dataset (Split Data), implementasi algoritma KNN, dan evaluasi performa model melalui berbagai metrik seperti RMSE dan MAE. Antarmuka drag-and-drop memudahkan pengguna dalam membangun alur kerja analisis data, mulai dari tahap preprocessing hingga evaluasi model. RapidMiner juga menyediakan fitur visualisasi yang membantu dalam memahami karakteristik data dan hasil analisis [8].

## 2.4 Metodologi Penelitian

Metodologi penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan dataset cuaca dari HPWREN untuk mengembangkan model prediksi cuaca menggunakan algoritma K-Nearest Neighbor. Dataset ini mencakup parameter penting seperti tekanan udara, suhu, arah angin, kecepatan angin, dan kelembaban relatif yang digunakan sebagai variabel input untuk prediksi kondisi cuaca.



GAMBAR 1.  
METODOLOGI PENELITIAN

Tahap pertama penelitian dimulai dengan preprocessing data menggunakan RapidMiner. Proses ini meliputi pembersihan data dari nilai yang hilang atau tidak valid, normalisasi data numerik, dan pemilihan atribut yang relevan untuk prediksi cuaca. Normalisasi data penting dilakukan mengingat parameter cuaca memiliki rentang nilai yang berbeda-beda, sehingga perlu distandarisasi untuk memastikan akurasi model KNN [9].

Selanjutnya, dataset dibagi menjadi data training (80%) dan data testing (20%) untuk memvalidasi performa model. Pembagian ini penting untuk menguji kemampuan generalisasi model terhadap data baru yang belum pernah dilihat sebelumnya. Algoritma KNN diterapkan dengan menggunakan operator k-NN di RapidMiner, dengan nilai k awal ditetapkan sebagai 5. Parameter ini dapat dioptimalkan lebih lanjut berdasarkan hasil evaluasi model.

Evaluasi performa model menggunakan beberapa metrik standar untuk kasus prediksi numerik, termasuk Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan R-squared ( $R^2$ ). Metrik-metrik ini memberikan gambaran komprehensif tentang akurasi prediksi model dari berbagai aspek. RMSE menunjukkan besarnya kesalahan prediksi, MAE mengukur rata-rata kesalahan absolut, sedangkan  $R^2$

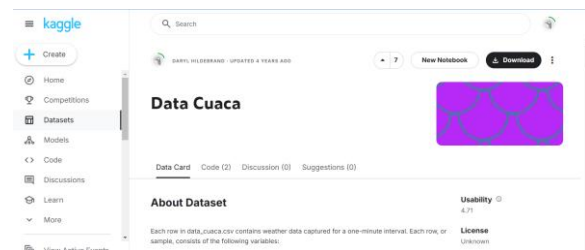
menunjukkan seberapa baik model menjelaskan variabilitas data [10].

## III. HASIL DAN PEMBAHASAN

### 3.1 Deskripsi Dataset

Pada bagian deskripsi dataset, penelitian ini menggunakan dataset cuaca dari HPWREN (High Performance Wireless Research and Education Network) yang terdiri dari 12 parameter penting untuk prediksi cuaca. Dataset mencakup:

- a. rowID: Identifier unik untuk setiap rekaman data
- b. hpwren\_timestamp: Waktu pengambilan data
- c. air\_pressure: Tekanan udara dalam hPa (hectopascal)
- d. air\_temp: Suhu udara dalam derajat Celsius
- e. avg\_wind\_direction: Arah angin rata-rata dalam derajat (0-360°)
- f. avg\_wind\_speed: Kecepatan angin rata-rata dalam m/s
- g. max\_wind\_direction: Arah angin maksimum dalam derajat
- h. max\_wind\_speed: Kecepatan angin maksimum dalam m/s
- i. min\_wind\_direction: Arah angin minimum dalam derajat
- j. min\_wind\_speed: Kecepatan angin minimum dalam m/s
- k. rain\_accumulation: Akumulasi curah hujan dalam mm
- l. rain\_duration: Durasi hujan dalam detik
- m. relative\_humidity: Kelembaban relatif dalam persen



GAMBAR 2.  
DATASET CUACA KAGGLE

Dataset ini menyediakan basis yang kuat untuk pengembangan model prediksi cuaca karena mencakup parameter-parameter kunci yang saling

kombinasi parameter numerik dengan skala berbeda. `MixedEuclideanDistance` memungkinkan perhitungan jarak yang lebih akurat antara data point, terutama penting mengingat dataset mencakup berbagai parameter cuaca seperti tekanan udara, arah angin, dan kelembaban yang memiliki karakteristik berbeda.

Parameter-parameter ini dipilih berdasarkan karakteristik data cuaca yang kompleks dan saling berkaitan. Kombinasi nilai  $k$  yang moderat ( $k=5$ ), weighted vote, dan penggunaan MixedEuclideanDistance memungkinkan model untuk menangkap pola-pola lokal dalam data sambil tetap mempertahankan kemampuan generalisasi yang baik.

GAMBAR 4.  
TAHAPAN IMPLEMENTASI KNN

Evaluasi model KNN untuk prediksi cuaca menghasilkan beberapa metrik performa yang menunjukkan kemampuan model dalam memprediksi suhu udara. Root Mean Squared Error (RMSE) sebesar 3.680 mengindikasikan rata-rata kesalahan prediksi sekitar 3.7 derajat Celsius. Nilai ini menunjukkan tingkat akurasi yang cukup baik mengingat rentang suhu udara yang cukup luas dalam dataset tiga bulan yang dianalisis.

[illegible]

```

class PerformanceVector:
    cost_mean_squared_error: 3.480 w/- 0.000
    absolute_error: 2.318 w/- 2.385
    cost_relative_squared_error: 0.315

```

Mean Absolute Error (MAE) sebesar 2.216 memberikan gambaran lebih detail tentang rata-rata kesalahan absolut prediksi model. Nilai ini lebih kecil dari RMSE karena MAE tidak memberikan penalti



lebih besar pada kesalahan yang besar, menunjukkan bahwa sebagian besar prediksi memiliki deviasi yang moderat dari nilai aktual.

Root Relative Squared Error (RRSE) sebesar 0.315 atau 31.5% mengindikasikan bahwa model memiliki performa yang lebih baik dibandingkan dengan prediksi menggunakan nilai rata-rata sederhana. Nilai ini menunjukkan bahwa model KNN berhasil menangkap pola-pola dalam data cuaca dan memberikan prediksi yang lebih akurat dibandingkan baseline sederhana. Kombinasi ketiga metrik ini menunjukkan bahwa model KNN dengan parameter yang dipilih ( $k=5$ , weighted vote, dan MixedEuclideanDistance) mampu memprediksi suhu udara dengan tingkat akurasi yang memadai untuk aplikasi praktis dalam prediksi cuaca.

#### IV. KESIMPULAN

Penelitian implementasi algoritma K-Nearest Neighbor untuk prediksi suhu udara menggunakan dataset HPWREN selama periode tiga bulan telah berhasil menunjukkan hasil yang menjanjikan. Model KNN dengan konfigurasi  $k=5$ , weighted vote, dan MixedEuclideanDistance mampu menghasilkan prediksi dengan tingkat akurasi yang baik, dibuktikan dengan nilai RMSE 3.680, MAE 2.216, dan RRSE 31.5%. Hasil ini mengonfirmasi bahwa pemilihan parameter dan preprocessing data yang tepat berperan penting dalam mengoptimalkan performa model untuk prediksi cuaca.

#### V. SARAN

Untuk pengembangan penelitian ke depan, beberapa aspek dapat ditingkatkan untuk memperoleh hasil yang lebih optimal. Periode pengambilan data dapat diperluas untuk mencakup variasi musiman yang lebih lengkap, memberikan pemahaman yang lebih komprehensif tentang pola cuaca. Eksplorasi nilai  $k$  yang berbeda dan penambahan variabel cuaca lain juga dapat meningkatkan akurasi prediksi. Perbandingan dengan algoritma machine learning lainnya dapat memberikan wawasan tentang pendekatan terbaik untuk prediksi cuaca.

Implementasi praktis dari penelitian ini dapat dikembangkan lebih lanjut melalui pengembangan sistem prediksi real-time yang terintegrasi dengan sumber data cuaca. Analisis sensitivitas terhadap parameter-parameter cuaca juga dapat dilakukan untuk memahami kontribusi masing-masing variabel terhadap akurasi prediksi. Pengembangan ini penting untuk meningkatkan kegunaan praktis dari model prediksi cuaca dalam berbagai aplikasi, dari pertanian hingga manajemen bencana.

#### DAFTAR PUSTAKA

- [1] I Wayan Sudiarta, "Pengantar Meteorologi". Universitas Mataram, Juni 2013.
- [2] A. Yogiarto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- [3] A. Sigit and A. S. Yuita, "Implementasi Data Mining Menggunakan Weka" Penerbit UB Pres Cetakan Pertama. Malang. Januari. 2018.
- [4] Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al- Musa, and Edi Tohidi, "Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 2, pp. 58–64, 2022, doi: 10.32485/kopertip.v4i2.120.
- [5] A. Jalil, A. Homaidi, and Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2070–2079, 2024, doi: 10.33379/gtech.v8i3.4811.
- [6] A. Nur Kirana, B. Nurhakim, S. Eka Permana, W. Prihartono, and G. Dwilestari, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1637–1642, 2024, doi: 10.36040/jati.v8i2.8967.
- [7] R. Aprilian, R. Habibi, and M. Y. H. Setyawan, "Algoritma KNN dalam Memprediksi Cuaca untuk Menentukan Tanaman yang Cocok Sesuai Musim". Penerbit; Kreatif Industri Nusantara, Bandung. 2020.
- [8] A. Mutoi Siregar and A. Puspabhuana, "Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner". Penerbit; CV. Kekata Group, 2017.
- [9] I. A. Gellysa Urva, Desyanti, "Penerapan Data Mining Di Berbagai Bidang: Konsep, Metode, Dan Studi Kasus". Penerbit; PT Sonpedia Publishing Indonesia, Mei, 2023.

- [10] H. Alrasyid, A. Homaidi, M. Kom, Z. Fatah, and M. Kom, "Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes," vol. 1, no. 1, pp. 447–453, 2024.