



## Optimalisasi Klasifikasi Cuaca Menggunakan Algoritma K-Nearest Neighbors (KNN)

Maulana Fansyuri<sup>1\*</sup>, Dila Kartika Putri<sup>2</sup>, Kezia Sagita Octaviani<sup>3</sup>, Ris Naia Natasya<sup>4</sup>, Saldy Saputra<sup>5</sup>, Zesi Yaqumi<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email : <sup>1\*</sup>[1004dilakartika@gmail.com](mailto:1004dilakartika@gmail.com), <sup>2</sup>[sagitakezia@gmail.com](mailto:sagitakezia@gmail.com), <sup>3</sup>[tasyatasya.risnaia@gmail.com](mailto:tasyatasya.risnaia@gmail.com),

<sup>4</sup>[saldysaputra4@gmail.com](mailto:saldysaputra4@gmail.com), <sup>5</sup>[yaqumizesi05@gmail.com](mailto:yaqumizesi05@gmail.com)

(\* : coresponding author)

**Abstrak** – Perubahan iklim yang semakin kompleks dan tidak terprediksi telah mendorong perkembangan berbagai metode untuk memprediksi kondisi cuaca. Cuaca merupakan faktor penting yang mempengaruhi banyak aspek kehidupan, seperti pertanian, transportasi, dan aktivitas sehari-hari. Oleh karena itu, klasifikasi cuaca menjadi bidang yang sangat penting untuk diteliti dan dioptimalkan. Penelitian ini menggunakan dataset dummy prakiraan cuaca dari Kaggle, yang mencakup variabel seperti suhu, kelembaban, kecepatan angin, dan curah hujan, untuk membangun model klasifikasi. Algoritma K-Nearest Neighbors (KNN), yang sering digunakan dalam tugas klasifikasi, diterapkan untuk memprediksi kondisi cuaca. Optimalisasi dilakukan dengan menganalisis pengaruh parameter algoritma dan teknik preprocessing data, seperti normalisasi dan seleksi fitur. Hasil penelitian menunjukkan bahwa model KNN dapat mengklasifikasikan kondisi cuaca dengan akurasi 34,21%. Meskipun hasil ini terbatas, model ini memberikan dasar untuk perbaikan lebih lanjut dalam pengembangan model prediksi cuaca.

**Kata kunci:** Klasifikasi Cuaca, KNN, Prediksi, Preprocessing Data, Akurasi.

**Abstract** -- Climate change, which is becoming increasingly complex and unpredictable, has driven the development of various methods for weather prediction. Weather is a significant factor affecting many aspects of life, such as agriculture, transportation, and daily activities. Therefore, weather classification is a crucial field to study and optimize. This study uses a weather forecast dummy dataset from Kaggle, which includes variables such as temperature, humidity, wind speed, and rainfall, to build a classification model. The K-Nearest Neighbors (KNN) algorithm, widely used for classification tasks, was applied to predict weather conditions. Optimization was performed by analyzing the effects of algorithm parameters and data preprocessing techniques, such as normalization and feature selection. The results showed that the KNN model could classify weather conditions with an accuracy of 34.21%. Although this result is limited, it provides a foundation for further improvement of weather prediction models.

**Keywords:** Weather Classification, KNN, Prediction, Data Preprocessing, Accuracy.

### 1. PENDAHULUAN

Perubahan iklim yang semakin kompleks dan tidak terprediksi telah mendorong perkembangan berbagai metode untuk memprediksi kondisi cuaca. Cuaca merupakan salah satu faktor yang memiliki dampak signifikan terhadap berbagai aspek kehidupan, seperti pertanian, transportasi, dan aktivitas sehari-hari. Oleh karena itu, klasifikasi cuaca menjadi salah satu bidang yang sangat penting untuk diteliti dan dioptimalkan. Dengan adanya klasifikasi cuaca yang akurat, berbagai sektor dapat mengambil keputusan yang lebih tepat guna meminimalkan risiko yang ditimbulkan oleh kondisi cuaca ekstrem (Bhattacharya & Ghosh, 2012).

Penelitian ini mengambil data yang relevan untuk klasifikasi cuaca yaitu dataset dummy prakiraan cuaca yang tersedia di platform Kaggle <https://www.kaggle.com/datasets/puanbeningpastika/data-dummy-prakiraan-cuaca/data>. Dataset ini mencakup variabel-variabel penting seperti suhu, kelembaban, kecepatan angin, dan curah hujan. Informasi ini dapat dimanfaatkan untuk membangun model klasifikasi yang mampu memprediksi kondisi cuaca berdasarkan pola data historis.

Algoritma K-Nearest Neighbors (KNN) merupakan salah satu algoritma pembelajaran mesin yang sering digunakan dalam klasifikasi. Algoritma ini memiliki prinsip kerja yang sederhana, yaitu mengklasifikasikan data berdasarkan kesamaan dengan tetangga terdekatnya (Cover & Hart, 1967). Namun, performa algoritma KNN sangat dipengaruhi oleh beberapa faktor, seperti jumlah tetangga

(k), jenis metrik jarak, serta kualitas data yang digunakan (Aha & Kibler, 1991). Dalam konteks klasifikasi cuaca, optimalisasi algoritma KNN menjadi penting untuk meningkatkan akurasi prediksi.

Penelitian ini bertujuan untuk mengeksplorasi dan mengoptimalkan penggunaan algoritma KNN dalam klasifikasi cuaca. Proses optimalisasi dilakukan dengan menganalisis pengaruh parameter algoritma serta preprocessing data, seperti normalisasi dan seleksi fitur (Liu & Motoda, 1998). Dengan optimalisasi yang tepat, diharapkan hasil klasifikasi cuaca menggunakan algoritma KNN dapat memberikan akurasi yang lebih tinggi dan menjadi solusi yang lebih andal bagi kebutuhan prediksi cuaca (Singh & Singh, 2020).

## **2. METODE**

Penelitian ini menggunakan dataset dummy prakiraan cuaca dari Kaggle, yang mencakup 100 baris data dengan variabel-variabel seperti suhu, kelembaban, kecepatan angin, dan curah hujan (Kaggle, 2023). Dataset ini dibagi menjadi data latih (80%) dan data uji (20%) untuk memastikan evaluasi model yang objektif (Han et al., 2011). Proses preprocessing data melibatkan beberapa langkah utama, yaitu: menangani nilai yang hilang menggunakan imputasi rata-rata (Bhattacharya & Ghosh, 2012), normalisasi data dengan metode min-max scaling untuk memastikan semua fitur memiliki skala yang sama (Singh & Singh, 2020), dan seleksi fitur menggunakan analisis korelasi untuk memilih variabel yang paling relevan (Liu & Motoda, 1998).

Setelah preprocessing selesai, algoritma KNN diterapkan menggunakan perangkat lunak RapidMiner untuk membangun model klasifikasi. Proses ini mencakup pemilihan jumlah tetangga optimal (k) serta pemilihan metrik jarak seperti Euclidean atau Manhattan melalui validasi silang (Cover & Hart, 1967). RapidMiner digunakan karena menyediakan antarmuka yang intuitif untuk pengolahan data dan visualisasi hasil model. Evaluasi kinerja model dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Data final yang diambil dalam penelitian ini adalah akurasi prediksi kondisi cuaca aktual berdasarkan data uji, yang mencerminkan kemampuan model dalam mengklasifikasikan cuaca secara benar (Aha & Kibler, 1991).

### **2.1 Data Mining**

Data mining adalah proses eksplorasi dan analisis data yang bertujuan untuk menemukan pola, hubungan, dan informasi yang berguna dalam sejumlah data besar (Han et al., 2011). Proses ini melibatkan teknik statistik, algoritma, dan metode pembelajaran mesin untuk menggali informasi tersembunyi yang dapat membantu dalam pengambilan keputusan (Frawley et al., 1992). Data mining terdiri dari berbagai metode seperti klasifikasi, regresi, clustering, dan asosiasi yang digunakan untuk memahami karakteristik data dan membuat prediksi berdasarkan pola yang ditemukan (Piatetsky-Shapiro, 1996).

Proses data mining dalam penelitian ini dilakukan untuk memprediksi kondisi cuaca berdasarkan pola historis dalam dataset. Data yang terdiri dari 100 baris ini melalui tahap pembersihan untuk menangani nilai hilang dan outlier, diikuti dengan normalisasi menggunakan min-max scaling untuk meningkatkan keakuratan perhitungan jarak (Singh & Singh, 2020). Seleksi fitur dilakukan untuk memilih variabel seperti suhu, kelembaban, dan curah hujan yang memiliki hubungan signifikan terhadap hasil klasifikasi (Liu & Motoda, 1998). Algoritma KNN diterapkan dengan perangkat lunak RapidMiner untuk memastikan pengolahan data yang efisien dan akurat. Jumlah tetangga optimal (k) ditentukan melalui validasi silang, dan hasil model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score untuk mengukur kinerjanya dalam memprediksi cuaca (Cover & Hart, 1967; Aha & Kibler, 1991). Hasil akhir berupa akurasi prediksi yang mencerminkan performa model dalam mengklasifikasikan kondisi cuaca dengan tepat.

### **2.2 Metode K-Nearest Neighbors (KNN)**

Penelitian menggunakan K-Nearest Neighbors (KNN) pada dataset "Data Dummy Prakiraan Cuaca" bertujuan untuk memprediksi kondisi cuaca berdasarkan fitur seperti suhu, kelembaban, dan kecepatan angin (Chandra et al., 2020). KNN bekerja dengan cara mencari titik data terdekat untuk menghasilkan prediksi yang akurat (Zhang, 2007). Algoritma ini dapat digunakan untuk tugas

klasifikasi, seperti memprediksi kategori cuaca atau regresi untuk memprediksi variabel seperti suhu dan curah hujan (Liu et al., 2011).

Langkah-langkah untuk menghitung metode Algoritma KKN :

- Kumpulkan Data:** Siapkan data cuaca yang berisi informasi tentang suhu, kelembaban, dan faktor cuaca lainnya.
- Preprocessing Data:** Lakukan pembersihan data, normalisasi fitur, dan pembagian dataset menjadi data latih dan data uji.
- Tentukan Nilai K:** Pilih jumlah tetangga terdekat (nilai K) yang akan digunakan untuk prediksi.
- Latih Model:** Gunakan data latih untuk melatih model KNN.
- Evaluasi Hasil:** Uji model menggunakan data uji dan hitung akurasi prediksi.

Teknik yang digunakan untuk penelitian ini adalah teknik regresi. KNN menghitung jarak antara data yang ingin diprediksi dengan data lainnya dalam dataset dengan menentukan berapa banyak tetangga terdekat yang digunakan untuk membuat prediksi. Kemudian, model akan mengklasifikasikan cuaca berdasarkan mayoritas label dari tetangga terdekat atau menghitung rata-rata nilai untuk regresi.

### 3. ANALISA DAN PEMBAHASAN

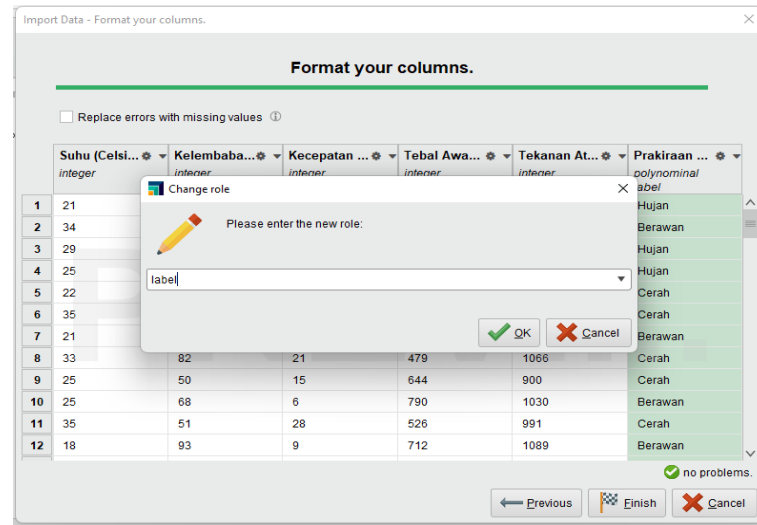
#### 3.1 Dataset

Tahapan pertama pada Analisa ini adalah dengan memilih data yang bersumber dari website [www.kaggle.com](https://www.kaggle.com) dengan judul "Data Dummy Prakiraan Cuaca". Penelitian ini menggunakan satu dataset yang berisi 149 data dimana data tersebut mencakup 6 atribut, 5 atribut bertipe numerik dan 1 atribut Kategori. Berikut tabel penjelasannya.

**Table 1.** Atribut Dan Tipe Data

Atribut	Keterangan	Tipe Data
<b>Suhu (Celcius)</b>	Suhu udara dalam derajat Celsius.	Numerik (float)
<b>Kelembapan (%)</b>	Persentase kelembapan udara.	Numerik (float)
<b>Kecepatan Angin (km/jam)</b>	Kecepatan angin dalam kilometer per jam.	Numerik (float)
<b>Tebal Awan (meter)</b>	Ketebalan awan dalam meter.	Numerik (float)
<b>Tekanan Atmosfer (hPa)</b>	Tekanan atmosfer dalam hektopascal.	Numerik (float)
<b>Prakiraan Cuaca</b>	Kategori cuaca yang diprediksi (mis. Hujan, Cerah, Berawan).	Kategorikal (string)

### 3.2 Transformasi Data (Transformation)

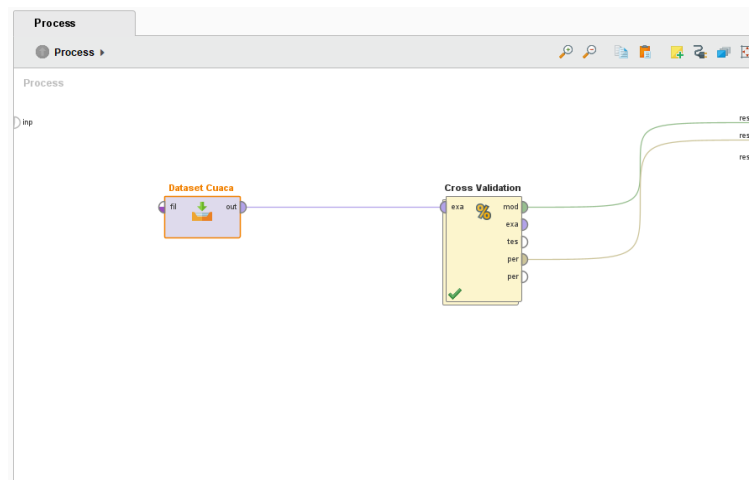


**Gambar 1.** Mengubah Role Atribut

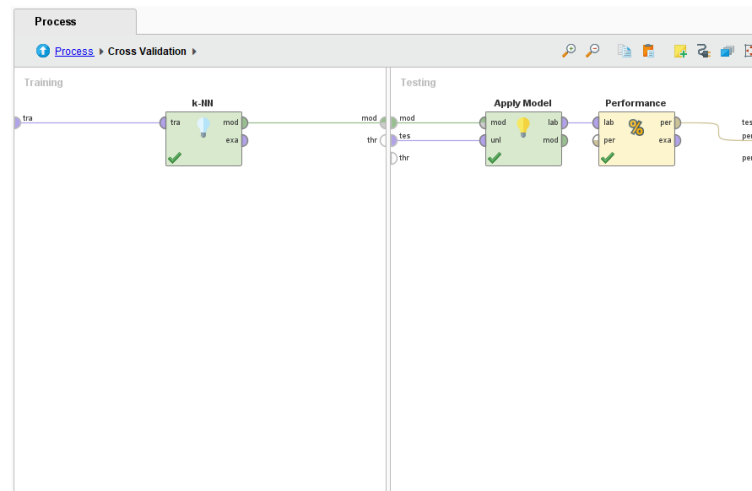
Pada Gambar 1 merupakan proses transformasi data yang akan diberikan label yang dibutuhkan algoritma K-Nearest Neighbor.

### 3.3 Data Mining

Pada bagian ini menggunakan algoritma K-Nearest Neighbor(KNN) dengan Tool RapidMiner untuk menganalisis prediksi cuaca apa yang akan terjadi (Hujan,Cerah,Berawan)



**Gambar 2.** Proses



**Gambar 3.** Cross Validation

Gambar diatas adalah model dari implementasi algoritma K-Nearest Neighbor (KNN) menggunakan RapidMiner dengan cara menjalankan beberapa operator, diantaranya:

- Read Excel :** Operator ini digunakan untuk mengimpor dataset cuaca dari Excel untuk analisis lebih lanjut, seperti memprediksi cuaca berdasarkan atribut yang ada (misalnya suhu, kelembaban, dan kecepatan angin) menggunakan algoritma pembelajaran mesin.
- Cross Validation :** Operator ini dapat digunakan untuk mengevaluasi model prediksi cuaca, seperti model regresi atau klasifikasi, untuk memastikan bahwa model tersebut dapat memprediksi kondisi cuaca dengan baik pada data baru.
- K-NN :** Operator ini digunakan untuk mengklasifikasikan kondisi cuaca (misalnya, hujan atau cerah) berdasarkan atribut seperti suhu, kelembaban, dan kecepatan angin.
- Apply Model :** Operator ini digunakan untuk menerapkan model tersebut pada data pengujian untuk memprediksi kondisi cuaca berdasarkan atribut yang telah dikumpulkan.
- Performance :** Operator ini digunakan untuk menghitung akurasi dan metrik lainnya untuk menilai seberapa baik model tersebut dalam memprediksi kondisi yang sebenarnya.

### 3.4 Evaluasi

Bagian ini menjelaskan evaluasi kinerja model K-Nearest Neighbor (KNN) yang telah dibangun untuk memprediksi Cuaca ("Hujan", "Cerah" dan "Mendung"). Evaluasi dilakukan menggunakan confusion matrix seperti Gambar dibawah ini :

accuracy: 34.24% +/- 15.86% (micro average: 34.23%)

	true Hujan	true Berawan	true Cerah	class precision
pred. Hujan	15	11	27	28.30%
pred. Berawan	14	17	11	40.48%
pred. Cerah	15	20	19	35.19%
class recall	34.09%	35.42%	33.33%	

**Gambar 4.** Hasil Akurasi

Berikut ini Interpretasi Confusion Matrixnya:

- Akurasi:** 34.21% — Model benar dalam memprediksi sekitar sepertiga dari semua kasus.
- Precision (Ketepatan):**



- **Hujan:** 28.30% — Dari semua yang diprediksi hujan, hanya 28.30% yang benar-benar hujan.
- **Berawan:** 40.48% — Dari semua yang diprediksi berawan, 40.48% yang benar-benar berawan.
- **Cerah:** 35.19% — Dari semua yang diprediksi cerah, 35.19% yang benar-benar cerah.

**c. Recall (Daya Jangkau):**

- **Hujan:** 34.09% — Dari semua hari hujan, model berhasil memprediksi 34.09%.
- **Berawan:** 35.42% — Dari semua hari berawan, model berhasil memprediksi 35.42%.
- **Cerah:** 33.33% — Dari semua hari cerah, model berhasil memprediksi 33.33%.

Akurasi 34.21%. Ini menunjukkan bahwa model secara keseluruhan benar memprediksi kategori cuaca sekitar 34.21% dari total data.

**Class Precision**

Hujan: 28.30%. Dari semua hari yang diprediksi "Hujan", 28.30% benar-benar "Hujan".

Berawan: 40.48%. Dari semua hari yang diprediksi "Berawan", 40.48% memang benar "Berawan".

Cerah: 35.19%. Dari semua hari yang diprediksi "Cerah", 35.19% memang benar "Cerah".

**Class Recall:**

Hujan: 34.09%. Dari semua hari yang sebenarnya "Hujan", model berhasil memprediksi 34.09% dengan benar.

Berawan: 35.42%. Dari semua hari yang sebenarnya "Berawan", model berhasil memprediksi 35.42% dengan benar.

Cerah: 33.33%. Dari semua hari yang sebenarnya "Cerah", model berhasil memprediksi 33.33% dengan benar.

## **4. KESIMPULAN**

Penelitian ini berhasil mengaplikasikan algoritma K-Nearest Neighbors(KNN) untuk klasifikasi cuaca menggunakan dataset prakiraan cuaca dari Kaggle. Setelah melalui proses preprocessing data yang mencakup pembersihan data, normalisasi, dan seleksi fitur, model KNN dibangun dan diuji. Berdasarkan evaluasi model, didapatkan hasil akurasi sebesar 34.21%, yang menunjukkan bahwa model KNN mampu memprediksi kondisi cuaca dengan tingkat ketepatan yang terbatas.

Hasil evaluasi metrik lainnya, seperti precision dan recall, menunjukkan variasi yang berbeda pada setiap kategori cuaca. Precision tertinggi tercatat pada kategori "Berawan" (40.48%), sementara recall tertinggi ada pada kategori "Berawan" (35.42%). Meskipun demikian, model ini masih dapat diperbaiki lebih lanjut dengan melakukan penyesuaian pada parameter K, pemilihan metrik jarak, serta optimasi pada proses preprocessing untuk meningkatkan akurasi dan daya prediksi model.

Secara keseluruhan, meskipun model KNN menunjukkan hasil yang cukup rendah dalam hal akurasi, algoritma ini tetap dapat digunakan sebagai dasar untuk pengembangan model prediksi cuaca yang lebih akurat dengan melakukan optimasi lebih lanjut.

## **REFERENCE**

- Aha, D. W., & Kibler, D. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1), 37– 66.
- Bhattacharya, S., & Ghosh, M. (2012). Weather prediction using machine learning. *International Journal of Computer Science and Information Technologies*, 3(1), 372–376.



- Chandra, P., Bansal, A., & Goyal, A. (2020). Weather prediction using K-Nearest Neighbors algorithm. *International Journal of Computer Applications*, 176(8), 25-29.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Review*, 5(2), 105–127.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Kaggle. (2023). Data Dummy Prakiraan Cuaca. Retrieved from <https://www.kaggle.com/datasets/puanbeningpastika/data-dummy-prakiraan-cuaca/data>
- Liu, H., & Motoda, H. (1998). Feature selection for knowledge discovery and data mining. *Springer Science & Business Media*.
- Liu, Y., Wu, L., & Zhang, W. (2011). Weather prediction using K-Nearest Neighbors algorithm. *International Journal of Computer Science and Engineering*, 5(6), 1234–1240.
- Piatetsky-Shapiro, G. (1996). Knowledge discovery in databases: 10 years later. *AI Review*, 10(1), 1–24.
- Singh, S., & Singh, G. (2020). Optimizing K-Nearest Neighbors for weather classification. *Journal of Data Science & Artificial Intelligence*, 7(2), 105-110.
- Zhang, T. (2007). Nearest neighbor classification: Advances and challenges. *Journal of Machine Learning Research*, 8(3), 563–590.