



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK V - Surabaya, 26 April 2025

Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : 10.31284/p.snestik.2025.7622

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Klasifikasi Cuaca Kota Bandung dengan Menggunakan Algoritma K-Nearest Neighbors

Sivana Amelia Irmayanti, Adhitya Arie Mahatma, Danu Septi Adi, Rega Wahyu Aji,
Rinci Kembang Hapsari*, M Ravi Ubaidillah

Prodi Teknik Informatika, Institut Teknologi Adhi Tama Surabaya, Jawa Timur
e-mail: rincikembang@itats.ac.id

ABSTRACT

Indonesia is one of the countries that often experiences abnormal weather cycles. Extreme weather occurs due to several factors, such as unusual atmosphere. Weather prediction is very important in helping people anticipate various possibilities caused by extreme weather changes. This study aims to classify the weather in Bandung using the K-Nearest Neighbor algorithm. The K-Nearest Neighbor algorithm was chosen because of its ty to handle data with not too many variables and its implementation in weather classification cases. The dataset is secondary data taken from kaggle.com. Testing was conducted with $k = 30$, $k = 50$ and $k = 70$. With the highest accuracy, precision, and recall values at $k = 50$.

Keywords: *classification, weather, k-Nearest-Neighbors*

ABSTRAK

Indonesia merupakan salah satu negara yang sering terjadi siklus cuaca tidak normal. Cuaca ekstrem terjadi dipengaruhi oleh beberapa faktor seperti atmosfer yang tidak biasa. Prediksi cuaca memiliki peran yang sangat penting dalam membantu masyarakat untuk mengantisipasi berbagai kemungkinan yang disebabkan oleh perubahan cuaca yang ekstrim. Dalam penelitian ini bertujuan melakukan klasifikasi cuaca kota Bandung dengan menggunakan algoritma K-Nearest Neighbor. Algoritma K-Nearest Neighbor dipilih karena kemampuannya yang baik dalam menangani data dengan variabel yang tidak terlalu banyak, serta kemudahan implementasinya dalam kasus klasifikasi cuaca. Dataset merupakan data sekunder yang diambil

dari kaggle.com. Pengujian dilakukan dengan nilai $k=30$, $k=50$ dan $k=70$. Dengan nilai akurasi, presisi dan recall tertinggi pada saat $k=50$.

Kata kunci: Klasifikasi; cuaca; k-Nearest-Neighbors.

PENDAHULUAN

Cuaca adalah gambaran kondisi udara di bumi. Informasi cuaca berperan penting di berbagai sektor, termasuk transportasi udara, pertanian, pelayaran, pertahanan, komunikasi, dan lainnya [1]. Prediksi cuaca penting dilakukan untuk kepentingan dan kebutuhan masyarakat, sebagai contoh bagi pelaut dan nelayan prediksi cuaca diperlukan untuk memprediksi bahaya semisal badai laut yang akan terjadi. Prakiraan cuaca penting untuk menentukan ekspektasi iklim di masa depan. Prakiraan cuaca sangat bergantung pada data cuaca dan prediksi yang akurat sangat penting untuk pengurangan risiko bencana [2]. Pengumpulan data dapat diperoleh melalui survei lapangan langsung, observasi, wawancara, dan data-data kearsipan yang diperoleh dari instansi terkait [3].

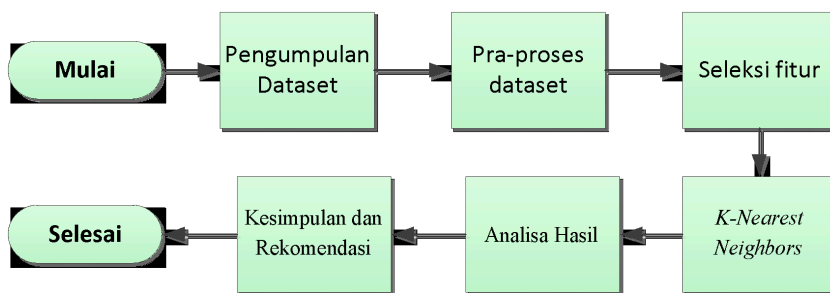
Dengan dilakukannya klasifikasi ini dapat membantu memberikan data yang diperlukan dalam pembuatan alat-alat untuk memprediksi cuaca tersebut. Dimana klasifikasi merupakan proses menemukan pola atau fungsi yang mendeskripsikan dan memisahkan kelas data, yang digunakan untuk memprediksi kelas dari objek data baru yang tidak diketahui kelasnya [4]. Untuk memprediksi intensitas curah hujan yang berbeda di berbagai wilayah di Indonesia, salah satunya adalah karena perbedaan kondisi geografis dan anomali suhu permukaan laut di kawasan laut Indonesia, Samudra Hindia, dan kawasan Pasifik [5]. Klasifikasi iklim secara global juga dipakai untuk memvisualisasikan perubahan iklim [6]. Klasifikasi cuaca membutuhkan metode klasifikasi yang tepat untuk mendapatkan hasil yang akurat, salah satunya dengan menggunakan data mining [7].

Data mining merupakan salah satu solusi untuk menganalisis data cuaca untuk mengklasifikasikan cuaca [8]. Dalam ranah data mining, klasifikasi menjadi sebuah teknik penting untuk mengelompokkan data berdasarkan karakteristik yang serupa [9]. Klasifikasi melibatkan tugas mengidentifikasi model atau fungsi yang memetakan setiap item data ke dalam salah satu dari serangkaian kelas atau kelompok yang telah ditentukan sebelumnya [10]. Sebagai alternatif, teknik data mining dan machine learning menyediakan pendekatan untuk meneliti elemen-elemen hidrologi secara mendalam, dengan tujuan mengungkap korelasi internal yang signifikan dan berpotensi bermanfaat antara berbagai komponen, yang kemudian dapat berfungsi sebagai prediktor yang efektif [11]. Dalam klasifikasi terdapat berbagai algoritma yang dapat digunakan, yaitu : Decision Trees, Neural Networks, Support Vector Machines, K-Nearest Neighbors, Naive Bayes, dan lainnya. Setiap algoritma memiliki kelebihan dan kekurangan masing-masing, sehingga pemilihan algoritma yang tepat sangat penting untuk mendapatkan hasil klasifikasi yang optimal.

Salah satu algoritma klasifikasi yang populer adalah K-Nearest Neighbors, algoritma ini dikenal karena kesederhanaan dan efektivitasnya dalam berbagai aplikasi. Metode K-Nearest Neighbor adalah algoritma klasifikasi yang sederhana namun efektif, yang mengklasifikasikan data baru berdasarkan mayoritas kelas dari k tetangga terdekatnya dalam ruang fitur [12] [13]. Dalam konteks klasifikasi cuaca di Bandung, algoritma K-Nearest Neighbors digunakan untuk memprediksi jenis cuaca berdasarkan data historis cuaca yang telah dikumpulkan. Algoritma ini bekerja dengan cara mencari sejumlah ' K ' data terdekat (tetangga) dalam dataset pelatihan, kemudian memprediksi kelas dari data baru berdasarkan kelas mayoritas dari tetangga-tetangga tersebut [14].

METODE

Dalam mengimplementasikan k-neares neighbor dalam klasifikasi cuaca kota Bandung dilakukan dengan langkah-langkah yang terstruktur, yang ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir Metodologi Penelitian

Tahap pertama dalam penelitian ini adalah pengumpulan data cuaca historis di kota Bandung. Dataset yang digunakan merupakan data sekunder yang diambil dari kaggle.com. Dataset terdiri dari 365 record. Data yang dikumpulkan mencakup berbagai parameter cuaca, seperti suhu udara, kelembaban, kecepatan angin, curah hujan, tekanan udara, dan parameter lainnya yang relevan. Dalam dataset dikelompokkan menjadi empat jenis cuaca, yaitu : berawan, berkabut, hujan dan cerah(bersinar).

Tahap selanjutnya adalah pra-pemrosesan data yang bertujuan untuk membersihkan dan mempersiapkan data agar sesuai untuk dianalisis. Data mentah seringkali mengandung nilai yang hilang (*missing values*), data yang tidak konsisten, atau data yang tidak relevan. Juga dilakukan normalisasi data untuk memastikan bahwa semua fitur memiliki rentang nilai yang sama, sehingga tidak ada fitur yang mendominasi fitur lainnya. Normalisasi yang digunakan adalah min-max normalisation, dengan rumus :

$$v' = \frac{v - \min_A}{\max_A - \min_A} (new_{\max_A} - new_{\min_A}) + new_{\min_A} \quad (1)$$

Setelah data selesai diproses dilakukan pemilihan fitur yang relevan dengan klasifikasi cuaca. Pemilihan fitur bertujuan untuk memilih fitur-fitur yang paling berpengaruh dalam memprediksi jenis cuaca. Fitur-fitur yaitu kelembapan, sinar, suhu dan angin.

Model k-Nearest Neighbors dibangun dengan menggunakan data pelatihan yang telah diproses. Algoritma k-Nearest Neighbors bekerja dengan mencari 'k' data terdekat dalam data pelatihan untuk setiap data baru yang akan diklasifikasikan. Jarak antara data baru dan data pelatihan dihitung menggunakan metrik jarak , yaitu Euclidean distance. Dengan rumus :

$$d_{euc}(x, y) = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (2)$$

Setelah menemukan k data terdekat, algoritma memprediksi kelas dari data baru berdasarkan kelas mayoritas dari tetangga-tetangga tersebut.

Setelah model k-Nearest Neighbors dibangun, model tersebut dievaluasi menggunakan data pengujian yang terpisah dari data pelatihan. Evaluasi model bertujuan untuk mengukur kinerja model dalam memprediksi jenis cuaca. Metrik evaluasi yang digunakan adalah akurasi, presisi, recall, dan F1-score. Rumus perhitungan metrik menggunakan persamaan 3 sampai persamaan 6.

$$Akurasi = \frac{TP + TN}{Total} \quad (3)$$

$$Presisi = \frac{TP}{FP + TP} \quad (4)$$

$$Recall = \frac{TP}{FN + TP} \quad (5)$$

$$F1 - Score = 2 * \frac{precision * recall}{precision + recall} \quad (6)$$

Akurasi memberikan gambaran tentang seberapa sering model memprediksi cuaca dengan benar, sementara presisi mengukur seberapa akurat model dalam memprediksi cuaca positif. Recall mengukur seberapa baik model menemukan semua kejadian cuaca positif, dan F1-score merupakan rata-rata harmonik dari presisi dan recall. Hasil klasifikasi ini akan dievaluasi untuk melihat seberapa akurat model dalam memprediksi cuaca [15].

HASIL DAN PEMBAHASAN

Data yang telah dikumpulkan kemudian dianalisis menggunakan algoritma K-Nearest Neighbors untuk mengklasifikasikan jenis cuaca di Bandung. Proses implementasi kode program pada python dengan pemanggilan paket library di python, paket library yang digunakan, adalah :

1. Memuat Dataset

Python

```
dataset = pd.read_csv('DATASET JURNAL.csv')
```

Pada bagian ini, dataset dari file CSV dengan nama 'DATASET JURNAL.csv' dimuat menggunakan library pandas (pd).

2. Pemisahan Fitur dan Label python

python

```
x = dataset.iloc[:, [0, 1, 2, 3, 4, 5, 6]].values
```

```
y = dataset.iloc[:, -1].values
```

Fitur atau atribut (x) dan label (y) diambil dari dataset. Fitur diambil dari kolom 0 hingga 6, sementara label diambil dari kolom terakhir.

3. Pemisahan Data Latih dan Uji

python

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=0)
```

Data dibagi menjadi data latih dan data uji menggunakan `train_test_split` dari `scikit-learn`. Data uji sebesar 25% dari total data. Pengujian dilakukan untuk nilai $k=30$, $k=50$ dan $k=70$. Hasil akurasi, recall dan presisi ditunjukkan pada tabel 1. Dari tabel dapat diketahui bahwa akurasi, recall dan presisi tertinggi ketika $k=50$. Sedangkan akurasi dan presisi terendah pada saat $k=30$ dan recall terendah pada saat $k=70$. Algoritma K-Nearest Neighbor menawarkan pendekatan yang sederhana namun efektif untuk klasifikasi cuaca, terutama dalam konteks di mana jumlah variabel yang relevan tidak terlalu besar, dan implementasinya relatif mudah dilakukan [15][16]. Kinerja sistem diukur menggunakan confusion matrix untuk mengevaluasi akurasi dan kemampuan klasifikasi model yang dibangun. K-Nearest Neighbors merupakan algoritma supervised learning yang memanfaatkan prinsip klasifikasi berdasarkan jarak terdekat antara data baru dengan data latih. Algoritma ini sangat mudah diimplementasikan dan memberikan hasil yang cukup baik untuk data yang memiliki pola yang jelas. K-Nearest Neighbors memiliki kemampuan untuk beradaptasi dengan data yang kompleks dan non-linear, serta relatif mudah diimplementasikan.

Tabel 1. Nilai evaluasi metrik

Pengukuran	K (dalam %)		
	30	50	70
Akurasi	60.7	66.7	62.6
Recall	67.7	68.8	67.4
Presisi	17.1	55.8	38.6

KESIMPULAN

Dalam penelitian ini bertujuan untuk mengimplementasikan algoritma K-Nearest Neighbors dalam klasifikasi cuaca di kota Bandung. Algoritma ini bekerja dengan cara mencari sejumlah 'K' data terdekat dalam dataset pelatihan, kemudian memprediksi kelas dari data baru berdasarkan kelas mayoritas dari tetangga-tetangga tersebut. Metode K-Nearest Neighbor adalah algoritma klasifikasi yang sederhana namun efektif, yang mengklasifikasikan data baru berdasarkan mayoritas kelas dari k tetangga terdekatnya dalam ruang fitur.

DAFTAR PUSTAKA

- [1] D. Setiawan, N. H. Prabowo, and N. B. Sukoco, "Rancang Bangun Alat Perekam Data Cuaca dengan Berbasis Internet Of Things," Jurnal HIDROPILAR, vol. 6, no. 2, p. 61, Dec. 2020, doi: 10.37875/hidropilar.v6i2.180.
- [2] I. Darmayanti, P. Subarkah, L. R. Anunggilarsa, and J. Suhaman, "PREDIKSI POTENSI SISWA PUTUS SEKOLAH AKIBAT PANDEMI COVID-19 MENGGUNAKAN ALGORITME K-NEAREST NEIGHBOR," JST (Jurnal Sains dan Teknologi), vol. 10, no. 2, p. 230, Nov. 2021, doi: 10.23887/jstundiksha.v10i2.39151.
- [3] A. Suzana, "Penyusunan Status Kerusakan Tanah untuk Produksi Biomassa di Kabupaten Bandung," Jurnal Civronlit Unbari, vol. 4, no. 1, p. 1, Jul. 2019, doi: 10.33087/civronlit.v4i1.41.
- [4] S. Mulyati, S. M. Husein, and R. Ramdhan, "RANCANG BANGUN APLIKASI DATA MINING PREDIKSI KELULUSAN UJIAN NASIONAL MENGGUNAKAN ALGORITMA (KNN) K-NEAREST NEIGHBOR DENGAN METODE EUCLIDEAN DISTANCE PADA SMPN 2 PAGEDANGAN," JIKA (Jurnal Informatika), vol. 4, no. 1, p. 65, Jan. 2020, doi: 10.31000/jika.v4i1.2288.
- [5] I. Peranginangin, A. Zakaria, and D. I. Kusumastuti, "Perbandingan spektrum curah hujan harian antara BMKG dan TRMM di DKI Jakarta," Rekayasa Jurnal Ilmiah Fakultas Teknik Universitas Lampung, vol. 24, no. 2, p. 53, Dec. 2021, doi: 10.23960/rekrjits.v24i2.41.
- [6] R. Shimabukuro, T. Tomita, and K. Fukui, "Update of global maps of Alisov's climate classification," Progress in Earth and Planetary Science, vol. 10, no. 1, Apr. 2023, doi: 10.1186/s40645-023-00547-1.
- [7] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," ILKOM Jurnal Ilmiah, vol. 12, no. 2, p. 104, Aug. 2020, doi: 10.33096/ilkom.v12i2.539.104-111.
- [8] N. Pan, "A Study on the Implications of Critical Discourse Analysis Theory for College English Teaching in the Context of Big Data," Applied Mathematics and Nonlinear Sciences, vol. 9, no. 1, Apr. 2023, doi: 10.2478/amns.2023.1.00104.
- [9] A. Winantu and C. Khatimah, "Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Dalam Memprediksi Prestasi Siswa," INTEK Jurnal Informatika dan Teknologi Informasi, vol. 6, no. 1, p. 58, May 2023, doi: 10.37729/intek.v6i1.3006.

-
- [10] N. Nigam and A. Rajavat, "A Systematic Literature Review of Data Classification Techniques," *International Journal of Computer Applications*, vol. 177, no. 44, p. 41, Mar. 2020, doi: 10.5120/ijca2020919971.
 - [11] A. P. PUTRA and D. P. RAMADHAN, "ANALISIS PENANGANAN BANJIR DI KAWASAN GENUK KOTA SEMARANG DENGAN MENGGUNAKAN PIPA RESAPAN HORIZONTAL," 2024.
 - [12] F. Akbar, S. Achmadi, and A. Mahmudi, "IMPLEMENTASI ANALISIS DATA KREDIT NASABAH MENGGUNAKAN METODE K-NEAREST NEIGHBORS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 4, no. 1, p. 82, May 2020, doi: 10.36040/jati.v4i1.2351.
 - [13] A. K. Wijaya, R. Syifa, I. N. Rahmadiano, and R. K. Hapsari, "Identifikasi Penyakit Ginjal Kronis Menggunakan Algoritma K-Nearest Neighbour (k-NN)," in *Seminar Nasional Informatika Bela Negara (SANTIKA)* ISSN, 2024, pp. 361–365.
 - [14] N. Yu and T. Haskins, "KNN, An Underestimated Model for Regional Rainfall Forecasting," *arXiv (Cornell University)*, Jan. 2021, doi: 10.48550/arxiv.2103.15235.
 - [15] I. Nopriansyah, A. Azwarman, and E. Raudhati, "Analisa Kapasitas Drainase Untuk Menanggulangi Banjir Pada Perumahan Kembar Lestari 1 Jambi," *Jurnal Talenta Sipil*, vol. 6, no. 1, p. 140, Feb. 2023, doi: 10.33087/talentasipil.v6i1.147.
 - [16] A. Rachmad, R. K. Hapsari, W. Setiawan, T. Indriyani, E. M. S. Rochman, and B. D. Satoto, *Classification of Tobacco Leaf Quality Using Feature Extraction of Gray Level Co-occurrence Matrix (GLCM) and K-Nearest Neighbor (K-NN)*, vol. 1. Atlantis Press