

**PENERAPAN ALGORITMA K-NEAREST NEIGHBORS DAN
NAIVE BAYES UNTUK KLASIFIKASI CUACA
BERDASARKAN DATA HISTORIS METEOROLOGI**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MUHAMAD DEWANO ARLINGGA

21.11.3842

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PENERAPAN ALGORITMA K-NEAREST NEIGHBORS DAN
NAIVE BAYES UNTUK KLASIFIKASI CUACA
BERDASARKAN DATA HISTORIS METEOROLOGI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MUHAMAD DEWANO ARLINGGA

21.11.3842

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PENERAPAN ALGORITMA K-NEAREST NEIGHBORS DAN NAIVE
BAYES UNTUK KLASIFIKASI CUACA BERDASARKAN DATA
HISTORIS METEOROLOGI**

yang disusun dan diajukan oleh

Muhamad Dewano Arlingga

21.11.3842

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 25 Juli 2025

Dosen Pembimbing,



Subektiningsih, M.Kom.

NIK. 190302413

HALAMAN PENGESAHAN
SKRIPSI
PENERAPAN ALGORITMA K-NEAREST NEIGHBORS DAN NAIVE
BAYES UNTUK KLASIFIKASI CUACA BERDASARKAN DATA
HISTORIS METEOROLOGI

yang disusun dan diajukan oleh

Muhamad Dewano Arlingga

21.11.3842

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Nuri Cahyono, M.Kom.
NIK. 190302278

Subektiningsih, M.Kom.
NIK. 190302413

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhamad Dewano Arlingga
NIM : 21.11.3842

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Algoritma *K-Nearest Neighbors* dan *Naive Bayes* Untuk Klasifikasi Cuaca Berdasarkan Data Historis Meteorologi

Dosen Pembimbing : Subektiningsih, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 25 Juli 2025

Yang Menyatakan,

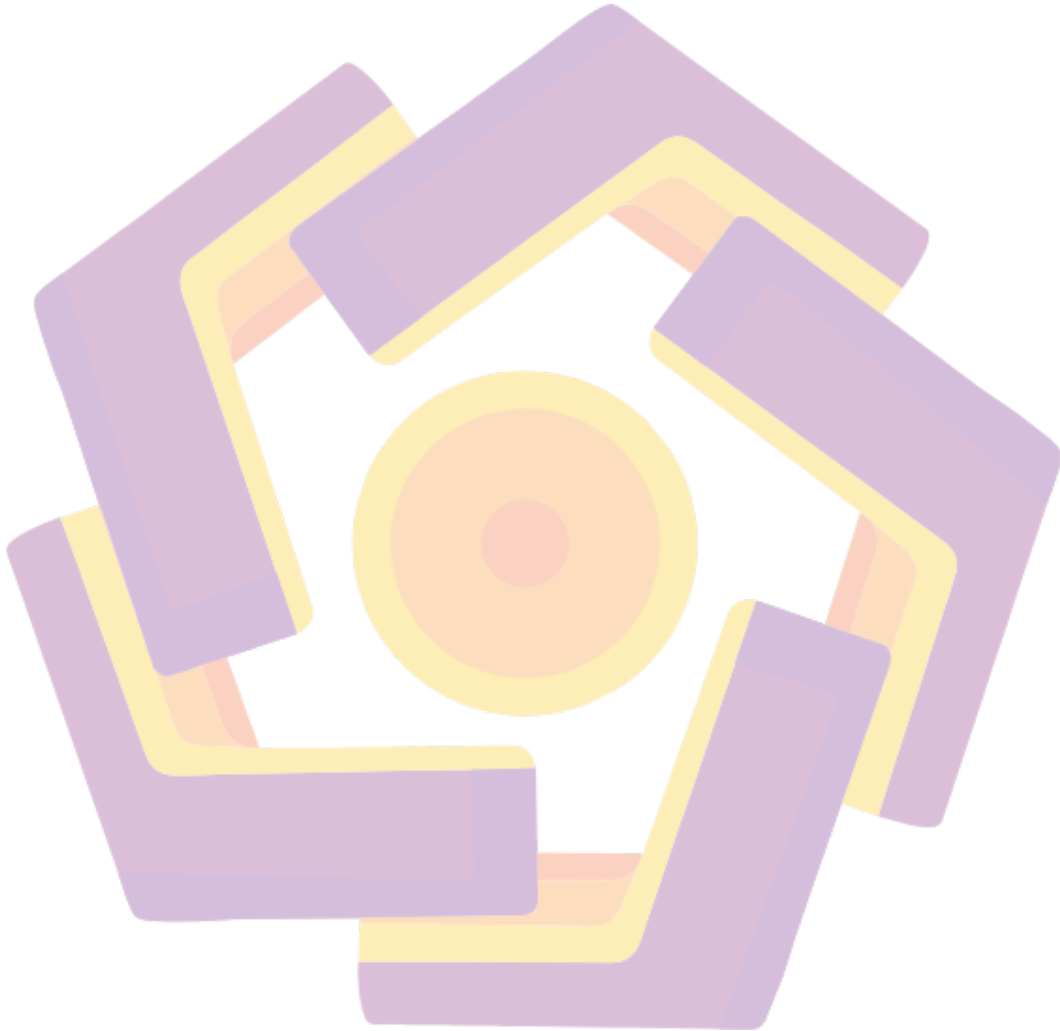


METERAI
TEMPEL
34732AMX425392027

Muhamad Dewano Arlingga

HALAMAN PERSEMBAHAN

(Bila ada) Halaman ini berisi kepada siapa skripsi dipersembahkan. Ditulis dengan singkat, resmi, sederhana, tidak terlalu banyak, serta tidak menjurus ke penulisan informal sehingga mengurangi sifat resmi laporan ilmiah.



KATA PENGANTAR

Bagian ini berisi pernyataan resmi yang ingin disampaikan oleh penulis kepada pihak lain, misalnya ucapan terima kasih kepada Dosen Pembimbing, Tim Dosen Penguji, dan semua pihak yang terkait dalam penyelesaian skripsi termasuk orang tua dan penyanggah dana.

Nama harus ditulis secara lengkap termasuk gelar akademik dan harus dihindari ucapan terima kasih kepada pihak yang tidak terkait. Bahasa yang digunakan harus mengikuti kaidah bahasa Indonesia yang baku.

Bagian ini tidak perlu dituliskan hal-hal yang bersifat ilmiah. Kata Pengantar diakhiri dengan mencantumkan kota dan tanggal penulisan diikuti di bawahnya dengan kata “Penulis” tanpa perlu menyebutkan nama dan tanda tangan.

Yogyakarta, 25 Juli 2025

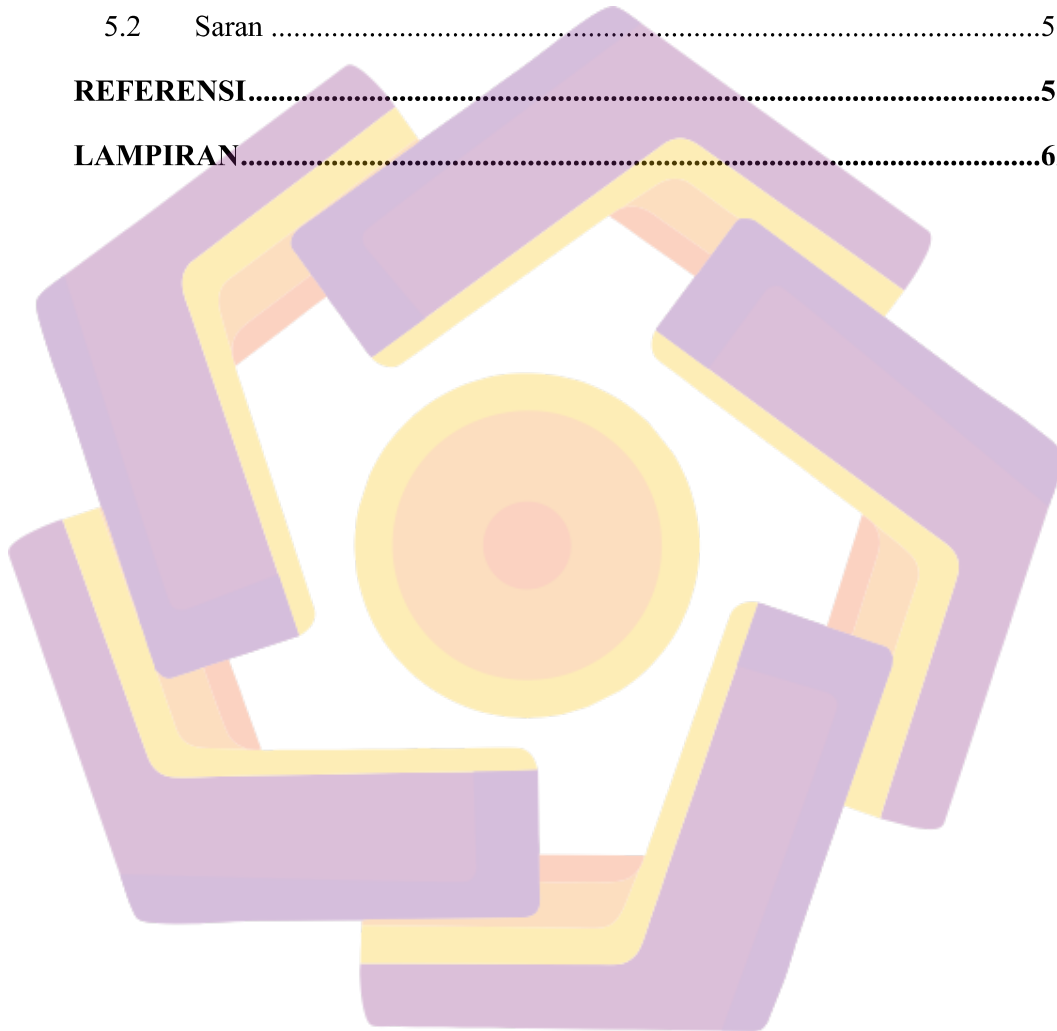
Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	12
2.2.1 <i>Artificial Intelligence</i>	12
2.2.2 <i>Machine Learning</i>	13

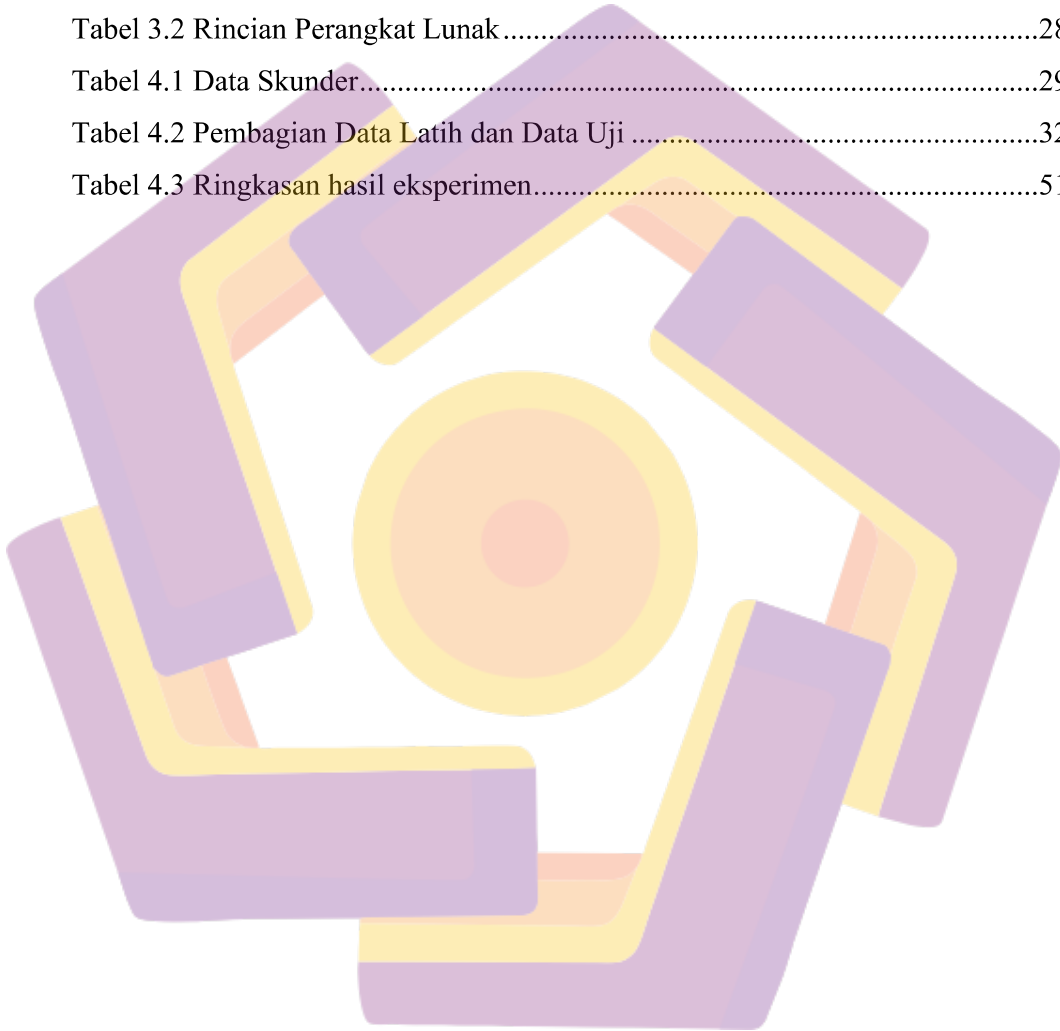
2.2.3	Klasifikasi	14
2.2.4	<i>K-Nearest Neighbours</i>	15
2.2.5	<i>Naive Bayes</i>	16
2.2.6	<i>Data Mining</i>	17
2.2.7	<i>Confusion Matrix</i>	19
BAB III METODE PENELITIAN		21
3.1	Objek Penelitian	21
3.2	Alur Penelitian	21
3.2.1	Identifikasi Masalah	22
3.2.2	Pemilihan Data	23
3.2.3	Praproses Data	24
3.2.4	Transformasi Data	25
3.2.5	<i>Data Mining</i>	26
3.2.6	Evaluasi Pola	27
3.2.7	Perbandingan	27
3.3	Alat dan Bahan	27
BAB IV HASIL DAN PEMBAHASAN		29
4.1	Persiapan <i>dataset</i>	29
4.2	<i>Praprocessing Data</i>	30
4.3	Transformasi Data	31
4.4	Pembagian Data Latih dan Data Uji	32
4.5	Algoritma KNN	33
4.4.1	Pembagian Rasio 70:30	33
4.4.2	Pembagian Rasio 80:20	38
4.4.3	Pembagian Rasio 90:10	44

4.6	Evaluasi Pola KNN	50
4.7	Perbandingan KNN dengan <i>Naive Bayes</i>	54
BAB V PENUTUP		56
5.1	Kesimpulan	56
5.2	Saran	56
REFERENSI.....		58
LAMPIRAN.....		62



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	9
Tabel 2.2 <i>Confusion Matrix</i>	19
Tabel 3.1 Rincian Perangkat Keras.....	28
Tabel 3.2 Rincian Perangkat Lunak.....	28
Tabel 4.1 Data Skunder.....	29
Tabel 4.2 Pembagian Data Latih dan Data Uji	32
Tabel 4.3 Ringkasan hasil eksperimen.....	51



DAFTAR GAMBAR

Gambar 2.1 Proses KDD.....	17
Gambar 3.1 Alur Penelitian	21
Gambar 4.1 Dataset Skunder	29
Gambar 4.2 Pengecekan data kosong	30
Gambar 4.3 Hasil Pengkategorian Perkolom.....	31
Gambar 4.4 Hasil <i>encoding</i>	31
Gambar 4.5 Hasil Normalisasi Data	32
Gambar 4.6 Laporan Klasifikasi Rasio 70:30 K = 3.....	33
Gambar 4.7 <i>Confusion Matrix</i> Rasio 70:30 K = 3	34
Gambar 4.8 Laporan Klasifikasi Rasio 70:30 K = 6.....	35
Gambar 4.9 <i>Confusion Matrix</i> Rasio 70:30 K = 6.....	36
Gambar 4.10 Laporan Klasifikasi Rasio 70:30 K = 9.....	37
Gambar 4. 11 <i>Confusion Matrix</i> Rasio 70:30 K = 9	38
Gambar 4.12 Laporan Klasifikasi Rasio 80:20 K = 3.....	39
Gambar 4.13 <i>Confusion Matrix</i> Rasio 80:20 K = 3	40
Gambar 4.14 Laporan Klasifikasi Rasio 80:20 K = 6.....	41
Gambar 4.15 <i>Confusion Matrix</i> Rasio 80:20 K = 6	42
Gambar 4.16 Laporan Klasifikasi Rasio 80:20 K = 9.....	43
Gambar 4.17 <i>Confusion Matrix</i> Rasio 80:20 K = 9	44
Gambar 4.18 Laporan Klasifikasi Rasio 90:10 K = 3.....	45
Gambar 4.19 <i>Confusion Matrix</i> Rasio 90:10 K = 3	46
Gambar 4.20 Laporan Klasifikasi Rasio 90:10 K = 6.....	47
Gambar 4.21 <i>Confusion Matrix</i> Rasio 90:10 K = 6	48
Gambar 4.22 Laporan Klasifikasi Rasio 90:10 K = 9.....	49
Gambar 4.23 <i>Confusion Matrix</i> Rasio 90:10 K = 9	50
Gambar 4.24 Evaluasi KNN Rasio 70:30	52
Gambar 4.25 Evaluasi KNN Rasio 80:20	53
Gambar 4.26 Evaluasi KNN Rasio 90:10	54

INTISARI

Penelitian ini bertujuan untuk memprediksi kondisi cuaca berdasarkan data historis meteorologi menggunakan algoritma *K-Nearest Neighbors* (KNN) serta mengevaluasi dan membandingkan performanya dengan algoritma *Naive Bayes* yang telah diterapkan pada penelitian sebelumnya. Data yang digunakan mencakup variabel *precipitation*, *temp_max*, *temp_min*, dan *wind*. Model KNN dibangun untuk mengklasifikasikan kondisi cuaca menjadi beberapa kategori, yaitu *drizzle*, *rain*, *sun*, *snow*, dan *fog*. Proses pengolahan data melibatkan tahapan pembersihan data, normalisasi menggunakan metode *Min-Max*, serta transformasi data kategorial menjadi numerik melalui label *encoding*. Selanjutnya, data dibagi ke dalam beberapa rasio pelatihan dan pengujian, yaitu 70:30, 80:20, dan 90:10 untuk menguji kestabilan performa model terhadap perubahan proporsi data latih. Hasil evaluasi menunjukkan bahwa akurasi model KNN meningkat seiring bertambahnya jumlah tetangga (nilai k) dan proporsi data pelatihan yang lebih besar. Akurasi tertinggi dicapai sebesar 85% pada rasio 90:10 dengan nilai $k=6$ dan $k=9$. Selain akurasi, metrik evaluasi lainnya seperti *precision*, *recall*, dan *F1-score* juga dianalisis, yang masing-masing bernilai 34%, 39%, dan 36%. Jika dibandingkan dengan algoritma *Naive Bayes* yang memperoleh akurasi 84,26%, KNN menunjukkan performa yang lebih baik dalam menangkap pola-pola *non-linear* pada data cuaca. Hasil ini mengindikasikan bahwa KNN memiliki potensi sebagai alternatif yang lebih unggul dalam klasifikasi kondisi cuaca berbasis data historis. Oleh karena itu, pemilihan algoritma prediksi perlu disesuaikan dengan kompleksitas data serta tujuan akhir dari prediksi yang diinginkan. Temuan ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem prediksi cuaca yang lebih akurat, adaptif, aplikatif, efisien, dan relevan untuk kebutuhan di berbagai sektor industri dan masyarakat.

Kata kunci : Prediksi cuaca, *K-Nearest Neighbors* (KNN), data historis meteorologi, klasifikasi cuaca.

ABSTRACT

This study aims to predict weather conditions based on meteorological historical data using the K-Nearest Neighbors (KNN) algorithm and evaluate and compare its performance with the Naive Bayes algorithm that has been applied in previous studies. The data used included the variables precipitation, temp_max, temp_min, and wind. The KNN model was built to classify weather conditions into several categories, namely drizzle, rain, sun, snow, and fog. The data processing process involves the stages of data cleaning, normalization using the Min-Max method, and the transformation of categorical data into numerical through label encoding. Next, the data was divided into several training and test ratios, namely 70:30, 80:20, and 90:10 to test the stability of the model's performance against changes in the proportion of the training data. The results of the evaluation showed that the accuracy of the KNN model increased as the number of neighbors increased (k-value) and the proportion of training data increased. The highest accuracy was achieved at 85% at a ratio of 90:10 with values of k=6 and k=9. In addition to accuracy, other evaluation metrics such as precision, recall, and F1-score were also analyzed, which were valued at 34%, 39%, and 36%, respectively. When compared to the Naive Bayes algorithm which obtained an accuracy of 84.26%, KNN showed better performance in capturing non-linear patterns in weather data. These results indicate that KNN has the potential to be a superior alternative in the classification of weather conditions based on historical data. Therefore, the selection of prediction algorithms needs to be adjusted to the complexity of the data as well as the final goal of the desired prediction. These findings are expected to contribute to the development of weather prediction systems that are more accurate, adaptive, applicative, efficient, and relevant to the needs of various industrial sectors and society.

Keywords: *Weather prediction, K-Nearest Neighbors (KNN), historical meteorological data, weather classification*