

**PERBANDINGAN ANTARA NAIVE BAYES DAN RANDOM
FOREST UNTUK PREDIKSI CUACA DI KOTA PALEMBANG**



TESIS

**NADYA KUSUMA WARDHANI
ENTERPRISE IT INFRASTRUCTURE
222420066**

**PROGRAM STUDI TEKNIK INFORMATIKA – S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2025**

PERBANDINGAN ANTARA *NAIVE BAYES* DAN *RANDOM FOREST* UNTUK PREDIKSI CUACA DI KOTA PALEMBANG



**Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar**

MAGISTER KOMPUTER

**NADYA KUSUMA WARDHANI
ENTERPRISE IT INFRASTRUCTURE
222420066**

**PROGRAM STUDI TEKNIK INFORMATIKA – S2
PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA
PALEMBANG
2025**

HALAMAN PENGESAHAN PEMBIMBING TESIS

Judul Tesis : PERBANDINGAN ANTARA *NAIVE BAYES* DAN *RANDOM FOREST* UNTUK PREDIKSI CUACA DI KOTA PALEMBANG

Oleh NADYA KUSUMA WARDHANI, NIM 222420066, Tesis ini telah disetujui dan disahkan oleh Pembimbing Program Studi Teknik Informatika – S2 Konsentrasi ENTERPRISE IT INFRASTRUCTURE Program Pascasarjana Universitas Bina Darma Palembang pada 12 September 2025 dan telah dinyatakan LULUS.

Palembang, 12 September 2025

Mengetahui,

Program Studi Teknik Informatika – S2

Universitas Bina Darma

Ketua,



Dr. Usman Ependi, S.Kom., M.Kom

Pembimbing,



Prof. Dr. Edi Surya Negara, M.Kom

HALAMAN PENGESAHAN PENGUJI TESIS

Judul Tesis : PERBANDINGAN ANTARA *NAIVE BAYES* DAN *RANDOM FOREST* UNTUK PREDIKSI CUACA DI KOTA PALEMBANG

Oleh NADYA KUSUMA WARDHANI, NIM 222420066, Tesis ini telah diujikan di hadapan Tim Penguji Program Studi Teknik Informatika – S2 Konsentrasi ENTERPRISE IT INFRASTRUCTURE Program Pascasarjana Universitas Bina Darma Palembang. Pada tanggal 12 September 2025 dan telah dinyatakan LULUS.

Palembang, 12 September 2025
Mengetahui,
Program Pascasarjana
Universitas Bina Darma
Direktur,



Logo of Universitas Bina Darma Program Pascasarjana, featuring a globe and the text 'Universitas Bina Darma PROGRAM PASCASARJANA'.

Prof. Dr. Ir. Achmad Syarifudin, M.Sc

Penguji I,



Handwritten signature of Prof. Dr. Edi Surya Negara.

Prof. Dr. Edi Surya Negara, M.Kom

Penguji II,



Handwritten signature of Dr. Yesi Novaria Kunang.

Dr. Yesi Novaria Kunang, S. T., M.Kom

Penguji III,



Handwritten signature of Zaid Amin.

Zaid Amin, M.Kom., Ph.D.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : NADYA KUSUMA WARDHANI

NIM : 222420066

Dengan ini menyatakan bahwa:

1. Karya tulis Saya (Tesis, Skripsi, Tugas Akhir) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baik (Magister, Sarjana, dan Ahli Madya) di Universitas Bina Darma;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka;
4. Karena yakin dengan keaslian karya tulis ini, Saya menyatakan bersedia Tesis/Skripsi/Tugas Akhir, yang Saya hasilkan di unggah ke internet;
5. Surat Pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 12 September 2025
Yang Membuat Pernyataan,



NADYA KUSUMA WARDHANI
NIM: 222420066

ABSTRAK

Perubahan iklim telah meningkatkan frekuensi kejadian hujan ekstrem, sehingga memperbesar risiko terjadinya bencana hidrometeorologi. Kota Palembang merupakan salah satu wilayah dengan curah hujan tinggi yang sering memicu banjir dan berdampak pada transportasi, ekonomi, serta kesehatan masyarakat. Oleh karena itu, prediksi curah hujan yang akurat sangat penting untuk mendukung upaya mitigasi bencana. Penelitian ini membandingkan kinerja algoritma Naïve Bayes dan Random Forest dengan menggunakan data cuaca harian dari BMKG Palembang periode 2019–2024. Tahapan penelitian meliputi pra-pemrosesan data, imputasi nilai hilang menggunakan KNN Imputer, normalisasi dengan MinMaxScaler, klasifikasi curah hujan ke dalam empat kategori (Ringan, Sedang, Lebat, dan Sangat Lebat), serta implementasi algoritma Naïve Bayes dan Random Forest. Evaluasi model dilakukan dengan Confusion Matrix dan K-Fold Cross-Validation menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC. Hasil penelitian menunjukkan bahwa Random Forest mencapai kinerja tertinggi dengan akurasi 99,60% dan AUC 100%, sementara Naïve Bayes memperoleh akurasi 97,17% dan AUC 99,79%. Namun, Naïve Bayes menghasilkan nilai presisi, recall, dan F1-score yang lebih tinggi, sehingga menunjukkan sensitivitas yang lebih baik dalam mendeteksi kelas minoritas. Kesimpulannya, Random Forest memberikan prediksi yang lebih stabil dan akurat untuk klasifikasi curah hujan secara keseluruhan, sedangkan Naïve Bayes lebih sesuai digunakan ketika fokus penelitian adalah pada deteksi kelas minoritas.

Kata kunci: Machine Learning, Naïve Bayes, Curah Hujan, Random Forest, Prediksi Cuaca

ABSTRACT

Climate change has increased the frequency of extreme rainfall events, leading to a higher risk of hydrometeorological disasters. Palembang City is one of the regions with high rainfall, often triggering floods that affect transportation, the economy, and public health. Therefore, accurate rainfall prediction is essential for effective disaster mitigation. This study compares the performance of the Naïve Bayes and Random Forest algorithms using daily weather data from BMKG Palembang for the period 2019-2024. The research stages include data preprocessing, missing value imputation using KNN Imputer, normalization with MinMaxScaler, classification of rainfall into four categories (Light, Moderate, Heavy, and Very Heavy), and implementation of Naïve Bayes and Random Forest algorithms. Model evaluation was carried out using a Confusion Matrix and K-Fold Cross-Validation with accuracy, precision, recall, F1-score, and AUC as the evaluation metrics. The results show that Random Forest achieves the highest performance with 99.60% accuracy and 100% AUC, while Naïve Bayes obtains 97.17% accuracy and 99.79% AUC. However, Naïve Bayes produces higher precision, recall, and F1-score, indicating better sensitivity in identifying minority classes. In conclusion, Random Forest provides more stable and accurate predictions for overall rainfall classification, whereas Naïve Bayes is more suitable when the focus is on maximizing minority class detection.

Keywords: *Machine Learning, Naive Bayes, Rainfall, Random Forest, Weather Prediction*

MOTTO DAN HALAMAN PERSEMBAHAN

"Kesabaran dan kerja keras akan selalu berbuah hasil"

Kupersembahkan tesis ini untuk :

- ✿ Allah SWT yang selalu memberikan kemudahan & kelancaran kepada saya
- ✿ Terkhusus Ibu dan Ayah tercinta yang selalu memberikan dukungan, semangat, serta do'a terbaiknya untuk masa depanku dan selamanya akan menjadi alasan utamaku untuk sukses, agar dapat memberikan rasa bangga kepada mereka.
- ✿ Dan rekan-rekan seperjuangan Magister Teknik Informatika Angkatan 28 khususnya Reguler A.

(Nadya Kusuma Wardhani, S. Tr. Kom., M. Kom.)

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan petunjuk-Nya sehingga penulis dapat menyelesaikan tesis dengan judul “Perbandingan Algoritma Naïve Bayes dan Random Forest dalam Prediksi Curah Hujan di Kota Palembang” tepat pada waktunya. Tesis ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan studi dan memperoleh gelar Pascasarjana Magister Teknik Informatika di Universitas Bina Darma. Penulis menyadari bahwa terselesaikannya tesis ini tidak lepas dari bantuan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT atas segala limpahan rahmat dan karunia-Nya.
2. Kedua orang tua tercinta atas doa, kasih sayang, dukungan, dan pengorbanan yang tiada henti.
3. Ibu Prof. Dr. Sunda Ariana, M.Pd., M.M. Selaku Rektor Universitas Bina Darma Palembang.
4. Prof. Dr. Ir. H. Achmad Syarifudin, M.Sc. Selaku Direktur Pascasarjana Universitas Bina Darma Palembang.
5. Dr. Usman Ependi, S.Kom., M.Kom Selaku Ketua Program Studi Magister Teknik Informatika yang telah memberi saran, kritik, arahan dan memberi dorongan dalam penyusunan tesis ini.
6. Prof. Dr. Edi Surya Negara., M.Kom. selaku pembimbing yang telah memberikan bimbingan dan arahan dalam penulisan tesis ini
7. Dr. Yesi Novaria Kunang, S. T., M.Kom. selaku penguji yang telah memberikan arahan dalam penulisan tesis ini
8. Zaid Amin, M.Kom., Ph.D. selaku penguji yang telah memberikan motivasi dan arahan dalam penulisan tesis ini.
9. Pihak Sekretariat Pascasarjana Universitas Bina Darma Palembang yang telah memberikan bimbingan pelayanan dengan baik.

Penulis
Nadya Kusuma Wardhani

DAFTAR ISI

HALAMAN DEPAN	i
HALAMAN PENGESAHAN PEMBIMBING TESIS	ii
HALAMAN PENGESAHAN PENGUJI TESIS	iii
SURAT PERNYATAAN	iv
ABSTRAK (BAHASA INDONESIA)	v
ABSTRACT (BAHASA INGGRIS)	vi
MOTTO DAN HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.1.1 Konteks Bencana Hidrometeorologi di Kota Palembang	2
1.2. Identifikasi Masalah.....	3
1.3. Perumusan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Kebaruan	4
1.6. Manfaat Penelitian	5
1.7. Pembatasan Masalah	6
1.8. Sistematika Penulisan	6

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	8
2.1. Analisis Sintesis Penelitian Terdahulu	8
2.2. Penelitian Terdahulu.....	9
2.3. Landasan Teori	10
2.2.1. Cuaca.....	10
2.2.2. <i>Data Mining</i> dan <i>Machine Learning</i>	10
2.2.3. <i>Naïve Bayes</i>	11
2.2.4. <i>Random Forest</i>	12
2.2.5. <i>K-Nearest Neighbor Imputation</i>	13
2.2.6. <i>K-Fold Cross-Validation</i>	14
BAB III METODOLOGI PENELITIAN	15
3.1. Waktu Dan Tempat Penelitian.....	15
3.2. Prosedur Penelitian	15
3.2.1 <i>Data Selection</i>	16
3.2.2 <i>Preprocessing</i> atau <i>Cleaning</i>	19
3.2.3 <i>Transformation</i>	21
3.2.4 Analisis dan Seleksi Fitur.....	23
3.2.5 <i>Data Mining</i>	24
3.2.6 <i>Interpretation/Evaluation</i>	27
3.3. Jadwal Penelitian	27
3.4. Kerangka Pemikiran	28
BAB IV HASIL PENELITIAN.....	29
4.1 Proses Analisis Data.....	29

4.2	<i>Data Mining</i>	30
4.2.1	Pengujian dengan <i>Naïve Bayes</i>	30
4.2.2	Pengujian dengan <i>Random Forest</i>	33
4.3	<i>Evaluation / Interpretation</i>	38
BAB V KESIMPULAN DAN SARAN		41
5.1.	Kesimpulan	41
5.2.	Saran	42
DAFTAR PUSTAKA		43



DAFTAR GAMBAR

Gambar 3.1. Tahapan <i>Knowledge Discovery In Database</i> (KDD)	16
Gambar 3.2. Coding Klasifikasi data pada Google Collab	18
Gambar 3.3. Hasil klasifikasi curah hujan	19
Gambar 3.4. Proses inputasi data menggunakan KNNI.....	20
Gambar 3.5. Hasil inputasi data <i>missing</i> menggunakan KNNI	20
Gambar 3.6. Penentuan fitur dan target serta pembagian dataset	21
Gambar 3.7. Proses transformasi data.....	22
Gambar 3.8. Tahap <i>Data Mining</i> Menggunakan Algoritma <i>Naïve Bayes</i>	25
Gambar 3.9. Tahap <i>Data Mining</i> Menggunakan Algoritma <i>Random Forest</i>	25
Gambar 3.10. Proses <i>K – Fold Cross Validation</i>	26
Gambar 3.11. Kode pemrograman tabel perbandingan hasil algoritma <i>Naïve Bayes</i> dan Algoritma <i>Random Forest</i>	27
Gambar 3.12. Kerangka Pemikiran.....	28
Gambar 4.1. Hasil Analisis Data Curah Hujan Kota Palembang.....	29
Gambar 4.2. <i>Confusion Matrix Naive Bayes</i>	30
Gambar 4.3. <i>Confusion Matrix Random Forest</i>	34
Gambar 4.4. hasil rata-rata metrik evaluasi model <i>Naïve Bayes</i> dan <i>Random Forest</i> setelah penerapan <i>K-Fold Cross-Validation</i>	38
Gambar 4.5. Perbandingan rata-rata akurasi <i>Random Forest</i> dan <i>Naïve Bayes</i> pada <i>K-Fold Cross-Validation</i>	40

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	9
Tabel 3.1. Data cuaca Kota Palembang.....	17
Tabel 3.2. Kategori Hujan.....	18
Tabel 4.1. Hasil Evaluasi Perhitungan <i>Confusion Matrix Naive Bayes</i>	33
Tabel 4.2. Hasil Evaluasi Perhitungan <i>Confusion Matrix Random Forest</i>	37
Tabel 4.3. Hasil Evaluasi <i>Confusion Matrix Naive Bayes</i> dan <i>Random forest</i>	38

DAFTAR LAMPIRAN

DAFTAR RIWAYAT HIDUP.....	47
Hasil Pengolahan Data.....	48
Lembar Konsultasi Pembimbing.....	56
SK Pembimbing.....	58
Lembar Perbaikan Tesis.....	59

