

**KOMPARASI METODE NAÏVE BAYES DAN REGRESI LOGISTIK
UNTUK KLASIFIKASI PRODUKTIVITAS PADI DI JAWA TIMUR**

THESIS

**Oleh :
HAMIM TOHARI
NIM. 210605210006**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**KOMPARASI METODE NAÏVE BAYES DAN REGRESI LOGISTIK
UNTUK KLASIFIKASI PRODUKTIVITAS PADI DI JAWA TIMUR**

THESIS

**Diajukan kepada :
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :
HAMIM TOHARI
NIM. 210605210006**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**KOMPARASI METODE NAÏVE BAYES DAN REGRESI LOGISTIK
UNTUK KLASIFIKASI PRODUKTIVITAS PADI DI JAWA TIMUR**

THESIS

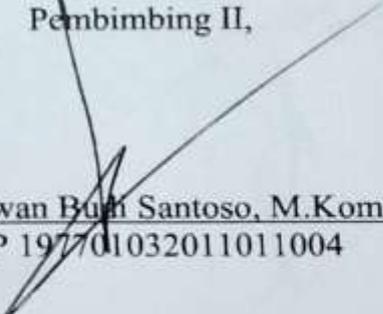
Oleh :
HAMIM TOHARI
NIM. 210605210006

Telah diperiksa dan disetujui untuk diuji:
Tanggal: 23 Mei 2025

Pembimbing I,


Prof. Dr. Sri Harini, M.Si
NIP 197310142001122002

Pembimbing II,


Dr. Irwan Budi Santoso, M.Kom
NIP 197701032011011004

Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Grysdiyan, M.Cs
NIP 19740424 2009011008

**KOMPARASI METODE NAÏVE BAYES DAN REGRESI LOGISTIK
UNTUK KLASIFIKASI PRODUKTIVITAS PADI DI JAWA TIMUR**

THESIS

**Oleh :
HAMIM TOHARI
NIM. 210605210006**

Telah Dipertahankan di Depan Dewan Penguji Thesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 03 Juni 2025

Susunan Dewan Penguji

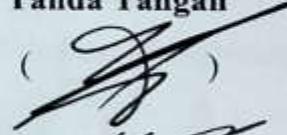
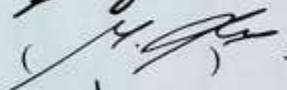
Tanda Tangan

Penguji Utama : Dr. Cahyo Crysdian, M.Cs
NIP 197404242009011008

Ketua Penguji : Dr. M. Ainul Yaqin, M.Kom
NIP 197610132006041004

Sekretaris Penguji : Prof. Dr. Sri Harini, M.Si
NIP 197310142001122002

Anggota Penguji : Dr. Irwan Budi Santoso, M.Kom
NIP 197701032011011004

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Cahyo Crysdian, M.Cs
NIP 197404242009011008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Hamim Tohari

NIM : 210605210006

Program Studi : Magister Informatika

Fakultas : Sains dan Teknologi

Judul Thesis : “Komparasi Metode Naïve Bayes Dan Regresi Logistik Untuk
Klasifikasi Produktivitas Padi Di Jawa Timur”

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 25 Juni 2025

Yang Membuat Pernyataan



Hamim Tohari

NIM: 210605210006

HALAMAN PERSEMBAHAN

“Thesis ini dipersembahkan untuk Istriku tercinta Erma Nur Khasanah dan anakku tersayang Samir Yusuf Almustofa yang menjadi sumber energi dan semangatku”

KATA PENGANTAR

Puji syukur penulis haturkan ke hadirat Allah SWT atas segala limpahan rahmat, taufik, serta hidayah-Nya sehingga penulis dapat menyelesaikan thesis yang berjudul:

“Komparasi Metode Naïve Bayes dan Regresi Logistik untuk Klasifikasi Produktivitas Padi di Jawa Timur”

Thesis ini disusun dalam rangka memenuhi salah satu syarat untuk memperoleh gelar Magister Komputer pada Program Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Adapun penelitian ini bertujuan untuk mengkaji dan membandingkan kinerja dua metode klasifikasi, yaitu Naïve Bayes dan Regresi Logistik dalam memprediksi produktivitas padi di Jawa Timur dengan memanfaatkan data numerik hasil kompilasi dari Badan Pusat Statistik (BPS) Jawa Timur. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem pendukung keputusan di bidang pertanian serta menjadi acuan bagi penelitian lanjutan di bidang data mining dan klasifikasi produktivitas pertanian.

Dalam kesempatan ini, penulis menyampaikan terima kasih dan penghargaan yang setulus-tulusnya kepada berbagai pihak yang telah memberikan dukungan, bimbingan, dan bantuan selama proses penyusunan tesis ini, terutama kepada:

1. Ibu Prof. Dr. Sri Harini, M.Si, selaku dosen pembimbing utama, atas segala arahan, bimbingan, dan motivasi yang diberikan sejak awal hingga terselesaikannya thesis ini.
2. Bapak Dr. Irwan Budi Santoso, M. Kom, selaku dosen pembimbing kedua, atas masukan berharga, kritik yang membangun, serta saran yang sangat membantu dalam penyempurnaan penelitian ini.
3. Bapak Dr. Cahyo Crys dian, M.Cs, selaku Kepala Program Studi Magister Informatika UIN Maulana Malik Ibrahim Malang, seluruh dosen dan staff administrasi di Program Studi Magister Informatika UIN Maulana Malik Ibrahim Malang, atas ilmu, pengalaman dan bantuan yang telah diberikan selama masa studi.
4. Orang tua yang hebat, Ibu Anikah dan Ibu Ninik Astutik serta Istri tercinta Erma Nur Khasanah, Anak tersayang Samir Yusuf Almustofa, atas doa dukungan moril, dan

motivasi yang tiada henti sehingga penulis dapat menyelesaikan thesis ini dengan baik.

5. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu baik secara langsung maupun tidak langsung dalam proses penyusunan thesis ini.

Penulis menyadari bahwa thesis ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang konstruktif sangat diharapkan demi perbaikan dan penyempurnaan di masa mendatang. Semoga thesis ini dapat memberikan manfaat dan kontribusi bagi perkembangan ilmu pengetahuan, khususnya di bidang data mining dan klasifikasi produktivitas pertanian.

Malang, 25 Juni 2025

Hamim Tohari

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
ABSTRAK	xii
ABSTRACT	xiv
المخلص	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	6
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
2.1 Klasifikasi Dalam Pertanian	8
2.2 Naïve Bayes	9
2.3 Regresi Logistik	13
2.4 Kerangka Teori	17
2.5 Kerangka Konsep	20
BAB III METODOLOGI PENELITIAN	23
3.1 Tahapan Penelitian	23
3.2 Desain Sistem	25
3.3 Pengumpulan Data	27

3.4 Pra-Pemrosesan Data	27
3.5 Metode yang Diusulkan	30
3.6 Hyperparameter Tuning	30
3.7 Validasi Data	33
3.8 Evaluasi Metode	37
3.9 Novelty Penelitian	40
BAB IV HASIL DAN PEMBAHASAN	42
4.1 Statistik Deskriptif	42
4.2 Hasil Hyperparameter Tuning	42
4.3 Hasil Eksperimen I	43
4.4 Hasil Eksperimen II	46
4.5 Hasil Eksperimen III	48
4.6 Hasil Eksperimen IV	50
4.7 Pembahasan Komparasi Eksperimen	52
4.8 Integrasi Penelitian dengan Islam	55
BAB V PENUTUP	57
5.1 Kesimpulan	57
5.2 Saran	58
DAFTAR PUSTAKA	60
LAMPIRAN	68

DAFTAR GAMBAR

Gambar 2.1 Kerangka Teori	17
Gambar 2.2 Kerangka Konsep	20
Gambar 3.1 Tahapan Penelitian	23
Gambar 3.2 Desain Sistem	25
Gambar 4.1 Confusion Matrix Eksperimen 1	45
Gambar 4.2 Kurva ROC Eksperimen 1	45
Gambar 4.3 Confusion Matrix Eksperimen 2	47
Gambar 4.4 Kurva ROC Eksperimen 2	47
Gambar 4.5 Confusion Matrix Eksperimen 3	49
Gambar 4.6 Kurva ROC Eksperimen 3	49
Gambar 4.7 Confusion Matrix Eksperimen 4	51
Gambar 4.8 Kurva ROC Eksperimen 4	51

DAFTAR TABEL

Tabel 3.1 Sampel Data Penelitian	27
Tabel 3.2 Ilustrasi Missing Values	29
Tabel 3.3 Hasil Transformasi dan Seleksi Fitur	29
Tabel 3.4 Nilai Grid Hyperparameter Gaussian Naïve Bayes	31
Tabel 3.5 Nilai Grid Hyperparameter Regresi Logistik	32
Tabel 3.6 Kombinasi Eksperimen	36
Tabel 3.7 Model Confusion Matrix	37
Tabel 4.1 Statistik Deskriptif Data	42
Tabel 4.2 Hasil Evaluasi Eksperimen 1	44
Tabel 4.3 Hasil Evaluasi Eksperimen 2	46
Tabel 4.4 Hasil Evaluasi Eksperimen 3	48
Tabel 4.5 Hasil Evaluasi Eksperimen 4	50

ABSTRAK

Tohari, Hamim. 2025. **Komparasi Metode Naïve Bayes Dan Regresi Logistik Untuk Klasifikasi Produktivitas Padi Di Jawa Timur**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Sri Harini, M.Si (II) Dr. Irwan Budi Santoso, M.Kom

Kata Kunci: Gaussian Naïve Bayes, Regresi Logistik, klasifikasi, produktivitas padi, Jawa Timur

Produktivitas padi merupakan faktor strategis yang memengaruhi ketahanan pangan dan perekonomian daerah, khususnya di Provinsi Jawa Timur yang dikenal sebagai salah satu sentra produksi padi nasional. Penelitian ini bertujuan untuk mengembangkan model klasifikasi produktivitas padi dengan membandingkan kinerja Gaussian Naïve Bayes (GNB) dan Regresi Logistik (LR), dua algoritma klasifikasi berbasis statistik yang umum digunakan dalam bidang data mining. Dataset yang digunakan dalam penelitian ini diperoleh dari hasil kompilasi data sekunder Badan Pusat Statistik (BPS) Jawa Timur yang mencakup atribut numerik seperti suhu udara, kelembaban, kecepatan angin, tekanan udara, curah hujan, dan intensitas penyinaran. Label produktivitas padi dikategorikan ke dalam dua kelas: Tinggi dan Rendah berdasarkan rata-rata produktivitas di wilayah studi. Penelitian ini menerapkan dua skenario pembagian data, yaitu rasio 80:20 dan 70:30 untuk data latih dan uji, dengan masing-masing divalidasi menggunakan 5-fold dan 10-fold cross-validation. Hasil eksperimen menunjukkan bahwa model GNB dengan konfigurasi rasio 70:30 dan validasi 10-fold menghasilkan performa terbaik, dengan akurasi 66%, F1-score 65%, dan AUC 72%, yang mencerminkan kemampuan model yang baik dalam membedakan kelas produktivitas. Sebaliknya, model Regresi Logistik menunjukkan performa yang lebih rendah pada seluruh skenario, dengan capaian terbaik akurasi 61% dan AUC 65% pada konfigurasi yang sama. Temuan penelitian ini mengindikasikan bahwa Gaussian Naïve Bayes lebih unggul dibandingkan Regresi Logistik dalam klasifikasi produktivitas padi berbasis data cuaca dan lingkungan di Jawa Timur.

ABSTRACT

Tohari, Hamim. 2025. **A Comparison of Naïve Bayes and Logistic Regression Methods for the Classification of Rice Productivity in East Java**. Thesis. Master of Informatics Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisors: (I) Prof. Dr. Sri Harini, M.Si (II) Dr. Irwan Budi Santoso, M.Kom

Keywords: Gaussian Naïve Bayes, Logistic Regression, classification, rice productivity, East Java

Rice productivity is a crucial factor influencing food security and the regional economy, particularly in East Java Province, which is recognized as a major center for national rice production. This study aims to develop a rice productivity classification model by comparing the performance of Gaussian Naïve Bayes (GNB) and Logistic Regression (LR), two statistically-based classification algorithms commonly used in the field of data mining. The dataset used in this study was sourced from a compilation of secondary data from the Central Bureau of Statistics (BPS) of East Java, which includes numerical attributes such as air temperature, humidity, wind speed, air pressure, rainfall, and solar radiation intensity. Rice productivity labels were categorized into two classes: High and Low, based on the average productivity in the study area. This research applied two data splitting scenarios—80:20 and 70:30 ratios for training and testing data—each validated using 5-fold and 10-fold cross-validation. The experimental results showed that the GNB model, under the 70:30 ratio and 10-fold validation configuration, achieved the best performance with an accuracy of 66%, an F1-score of 65%, and an AUC of 72%, reflecting the model's good capability to distinguish between productivity classes. Conversely, the Logistic Regression model exhibited lower performance across all scenarios, with its best achievement being an accuracy of 61% and an AUC of 65% under the same configuration. The findings of this study indicate that Gaussian Naïve Bayes outperforms Logistic Regression in classifying rice productivity based on weather and environmental data in East Java.

المخلص

طوهاري، حاميم. 2025. مقارنة بين طريقتي نايف بايز والانحدار اللوجستي لتصنيف إنتاجية الأرز في جاوة الشرقية. رسالة ماجستير. برنامج ماجستير في المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانج. المشرفون: (١) الأستاذة الدكتورة سري هريني، ماجستير العلوم (٢) الدكتور إيروان بودي سانتوسو، ماجستير الحاسوب

الكلمات المفتاحية: نايف بايز الغاوسي، الانحدار اللوجستي، التصنيف، إنتاجية الأرز، جاوة الشرقية.

إنتاجية الأرز عامل استراتيجي يؤثر على الأمن الغذائي والاقتصاد الإقليمي، وخاصة في مقاطعة جاوة الشرقية المعروفة بأنها أحد مراكز إنتاج الأرز الوطنية. تهدف هذه الدراسة إلى تطوير نموذج تصنيف إنتاجية الأرز من خلال ، وهما خوارزميتان للتصنيف (LR) والانحدار اللوجستي (GNB) Gaussian Naïve Bayes مقارنة أداء الإحصائي تُستخدمان بشكل شائع في استخراج البيانات. تم الحصول على مجموعة البيانات المستخدمة في هذه والتي تتضمن سمات رقمية (BPS) الدراسة من تجميع البيانات الثانوية من وكالة الإحصاء المركزية لشرق جاوة مثل درجة حرارة الهواء والرطوبة وسرعة الرياح وضغط الهواء وهطول الأمطار وكثافة الإشعاع. يتم تصنيف تسميات إنتاجية الأرز إلى فئتين: عالية ومنخفضة بناءً على متوسط الإنتاجية في منطقة الدراسة. تطبق هذه الدراسة سيناريوهين لمشاركة البيانات، وهما نسبة 80:20 و70:30 لبيانات التدريب والاختبار، مع التحقق من صحة كل ، بتكوين نسبة GNB منهما باستخدام التحقق المتبادل 5 أضعاف و10 أضعاف. أظهرت النتائج التجريبية أن نموذج ، ومساحة تحت المنحنى 72%، مما يعكس 65% F1 70:30 وتحقق عشري، حقق أفضل أداء، بدقة 66%، ودرجة قدرة النموذج الجيدة على تمييز فئات الإنتاجية. في المقابل، أظهر نموذج الانحدار اللوجستي أداءً أقل في جميع السيناريوهات، مع أفضل تحقيق لدقة 61% ومساحة تحت المنحنى 65% في التكوين نفسه. تشير نتائج هذه الدراسة إلى تفوق خوارزمية بايز الغوسية الساذجة على الانحدار اللوجستي في تصنيف إنتاجية الأرز بناءً على بيانات الطقس والبيئة في شرق جاوة

BAB I

PENDAHULUAN

1.1 Latar Belakang

Demi terwujudnya ketahanan pangan bagi masyarakat Indonesia, sektor pertanian diandalkan sebagai penyedia utama. Hal ini didasari oleh fakta bahwa pangan adalah kebutuhan pokok yang wajib dipenuhi, yang berarti harus selalu tersedia setiap saat dengan harga yang dapat dijangkau oleh siapa pun. (Cahyono et al., 2023). Di antara berbagai komoditas pertanian, padi memegang peran sentral sebagai bahan makanan pokok untuk mayoritas masyarakat di Indonesia. Produksi padi tersebar di berbagai wilayah Indonesia, dengan Jawa Timur menjadi provinsi yang menghasilkan padi paling besar. Tingkat produktivitas padi di Jawa Timur sangat penting untuk diperhatikan demi menjaga ketahanan pangan nasional dan menjamin ketersediaan beras bagi masyarakat. Namun, tingkat produktivitas padi di Jawa Timur sangat dipengaruhi oleh berbagai faktor, di antaranya kondisi iklim, jenis tanah, penggunaan pupuk, serangan hama dan teknologi budidaya. Fluktuasi produktivitas padi dapat terjadi antar musim tanam, bahkan antar wilayah di Jawa Timur. Untuk mengoptimalkan produksi padi, diperlukan analisis dan prediksi yang akurat mengenai produktivitas padi di setiap wilayah (Hidayati et al., 2022).

Pemanfaatan teknologi informasi, khususnya metode *data mining*, dapat menjadi solusi untuk menganalisis dan memprediksi produktivitas padi. *Data mining* merupakan proses pengumpulan dan pengolahan data untuk menemukan pola dan informasi penting yang tersembunyi di dalam data. Dalam *data mining*

terdapat metode yang sering digunakan untuk klasifikasi adalah Naïve Bayes dan Regresi Logistik.

Naïve Bayes merupakan metode klasifikasi probabilistik sederhana yang didasarkan pada penerapan teorema Bayes dengan memiliki asumsi independensi yang kuat antar fitur. Metode ini mudah diimplementasikan dan efektif untuk dataset berdimensi tinggi. Meskipun memiliki asumsi yang terkadang tidak realistis, Naïve Bayes terbukti bisa memberikan hasil yang baik dalam berbagai aplikasi klasifikasi (Sani et al., 2022).

Regresi Logistik merupakan metode statistik yang digunakan dalam memodelkan probabilitas dari kejadian yang memiliki hasil biner. Inti dari regresi logistik adalah penggunaan **fungsi logit**, yaitu kurva berbentuk 'S' yang mampu memetakan input apapun (dari negatif tak hingga hingga positif tak hingga) ke dalam rentang probabilitas antara 0 dan 1. Model ini menentukan hubungan linier antara variabel prediktor (independen) dengan log-odds (logaritma peluang) dari hasil biner (Fauzan et al., 2024).

Meskipun keduanya adalah metode klasifikasi yang banyak digunakan, Naïve Bayes dan Regresi Logistik memiliki pendekatan fundamental yang berbeda: Naïve Bayes adalah berfungsi sebagai pengklasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan memiliki asumsi independensi fitur yang kuat, membuatnya sederhana dan efisien bahkan pada dataset berdimensi tinggi, sedangkan Regresi Logistik memodelkan hubungan linier antara variabel prediktor dan log-odds dari hasil biner menggunakan fungsi logit, menghasilkan probabilitas kelas yang interpretatif dan sering digunakan sebagai tolak ukur yang kuat karena

kemampuannya untuk mengukur pengaruh setiap fitur (Purba et al., 2024). Oleh karena itu, perlu dilakukan studi komparatif untuk mengetahui metode mana yang lebih efektif dalam mengklasifikasikan produktivitas padi di Jawa Timur.

Yuliana et al. (2021) menggunakan decision tree C4.5 untuk memprediksi hasil panen padi di Jawa Barat dan memperoleh akurasi sebesar 78%. Pratama dan Lestari (2020) menerapkan K-Nearest Neighbor untuk klasifikasi produktivitas padi di Lampung dengan akurasi mencapai 81%. Handayani et al. (2022) membandingkan metode Naïve Bayes dan decision tree untuk klasifikasi produktivitas padi di Kalimantan Selatan dan menunjukkan bahwa Naïve Bayes unggul pada data dengan fitur-fitur yang saling independen, dengan akurasi 75%. Suryani et al. (2019) memanfaatkan regresi logistik untuk memprediksi produktivitas padi di Sulawesi Selatan berdasarkan data agroklimat, dengan hasil akurasi sebesar 68% dan menunjukkan interpretasi yang kuat terhadap pengaruh variabel iklim. Saputra dan Ahmad (2021) melaporkan bahwa regresi logistik lebih unggul pada data produksi padi yang memiliki keterkaitan linier antar variabel, dibandingkan dengan metode KNN.

Studi komparatif ini diharapkan dapat memberikan informasi yang berguna bagi para pengambil keputusan di sektor pertanian, seperti pemerintah daerah, penyuluh pertanian, dan petani, dalam mengoptimalkan produksi padi di Jawa Timur. Hasil klasifikasi produktivitas padi dapat digunakan untuk mengidentifikasi wilayah dengan produktivitas padi tinggi dan rendah. Informasi ini dapat digunakan untuk menentukan prioritas program pengembangan pertanian. Selain itu, hasil

klasifikasi produktivitas padi dapat juga digunakan untuk menganalisis faktor-faktor yang mempengaruhi produktivitas padi.

Penelitian ini menggunakan pendekatan komparatif dengan membandingkan kinerja metode Naïve Bayes dan Regresi Logistik dalam mengklasifikasikan produktivitas padi di Jawa Timur berdasarkan data historis. Data yang digunakan meliputi berbagai faktor yang dapat mempengaruhi produktivitas padi, seperti suhu udara, kelembaban udara, kecepatan angin, tekanan udara, curah hujan, hari hujan, luas panen dan produksi padi. Hasil penelitian ini diharapkan dapat memberikan rekomendasi metode klasifikasi yang lebih akurat dan efisien untuk memprediksi produktivitas padi di Jawa Timur.

Perintah Allah bagi umat manusia untuk berfikir dan merenungi tanda-tanda kekuasaan-Nya yang tersebar di alam semesta menjadi sumber inspirasi bagi penelitian ini, sebagaimana tertuang dalam Q.S. Al-Baqarah ayat 164 :

"Sesungguhnya dalam penciptaan langit dan bumi, silih bergantinya malam dan siang, bahtera yang berlayar pada laut membawa apa yang berguna bagi manusia dan apa yang Allah turunkan dari langit berupa air lalu dengan air itu Dia hiduskan bumi sesudah mati (kering)-nya dan Dia sebarkan di bumi itu segala jenis hewan, dan berkisarnya angin dan awan yang dikendalikan antara langit dan bumi. Sungguh (terdapat) tanda-tanda (keesaan dan kebesaran Allah) bagi kamu yang memikirkan."

1.2 Pernyataan Masalah

Uraian pada latar belakang mengarahkan pada perumusan masalah penelitian sebagai berikut :

1. Bagaimana penerapan metode Naïve Bayes dan Regresi Logistik untuk klasifikasi produktivitas padi di Jawa Timur?
2. Bagaimana perbandingan kinerja (*accuracy, precision, recall*) metode Naïve Bayes dan Regresi Logistik dalam klasifikasi produktivitas padi di Jawa Timur?
3. Metode manakah yang lebih efektif untuk klasifikasi produktivitas padi di Jawa Timur berdasarkan hasil perbandingan kinerja?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan metode Naïve Bayes dan Regresi Logistik untuk klasifikasi produktivitas padi di Jawa Timur.
2. Membandingkan kinerja (*accuracy, precision, recall*) metode Naïve Bayes dan Regresi Logistik dalam klasifikasi produktivitas padi di Jawa Timur.
3. Menentukan metode yang lebih efektif untuk klasifikasi produktivitas padi di Jawa Timur berdasarkan hasil perbandingan kinerja.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang sains data dan data mining, terlebih dalam penerapan metode Naïve Bayes dan Regresi Logistik untuk klasifikasi produktivitas padi.
2. Memberikan rekomendasi metode klasifikasi yang lebih akurat dan efisien untuk memprediksi produktivitas padi di Jawa Timur.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Dalam penelitian ini, analisis dilakukan terhadap data sekunder yang dirilis secara resmi oleh Badan Pusat Statistik (BPS) Provinsi Jawa Timur.
2. Faktor-faktor yang digunakan dalam klasifikasi produktivitas padi dibatasi pada suhu, kelembaban udara, kecepatan angin, tekanan udara, curah hujan, hari hujan, penyinaran, luas panen padi dan jumlah panen padi.

1.6 Sistematika Penulisan

Bab I Pendahuluan

Pada bab ini berisi latar belakang, pernyataan masalah, tujuan, manfaat, batasan masalah dan sistematika penulisan dalam penelitian ini.

Bab II Kajian Pustaka

Pada bab ini berisi kajian pustaka dari teori-teori yang mendasari penyelesaian masalah serta penelitian-penelitian yang telah dilakukan sebelumnya yang terkait dengan penelitian ini.

Bab III Metodologi Penelitian

Pada bab ini berisi mengenai langkah-langkah penelitian yang dilakukan, yaitu desain penelitian, sumber data, metode pengumpulan data dan tahap-tahap penelitian.

Bab IV Pembahasan

Pada bab ini berisi mengenai hasil penelitian, pembahasan, langkah-langkah perhitungan dan analisa dari metode Naïve Bayes dan Regresi Logistik, serta hasil perbandingan akurasi kedua metode tersebut.

Bab V Penutup

Pada bab ini berisi kesimpulan dari penulisan pada bab-bab sebelumnya dan saran untuk penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1 Klasifikasi Dalam Pertanian

Berdasarkan studi pustaka, peneliti menemukan beberapa referensi penelitian terdahulu yang berkaitan dengan penelitian yang sedang dilakukan peneliti. Studi penelitian terdahulu sangat penting sebagai bahan acuan yang membantu peneliti dalam merumuskan asumsi dasar dalam penelitian.

Penelitian pertama berasal dari Siti Maesaroh dan Kusrin (2021) yang melakukan analisis terhadap peningkatan produktivitas padi. Dalam penelitiannya disebutkan bahwa ada beberapa faktor yang memiliki peran yang penting terhadap produktivitas padi, yaitu cuaca, pupuk, kondisi tanah dan kesediaan pupuk. Selain itu, kemampuan dalam memprediksi produktivitas sangat diperlukan dalam tahap berikutnya. Data mining digunakan untuk membangun sistem prediksi. Prediksi dilakukan dengan mengolah data pendukung yang tersedia menggunakan 3 metode, yaitu K Nearest Neighbor (KNN), Decision Tree (Algoritma C45) dan *Naïve Bayes*. Selanjutnya hasil prediksi diuji menggunakan K Fold Cross Validation. Hasilnya, tiga metode tersebut memiliki kesalahan prediksi di bawah 30% menggunakan 100 data training dan *Naïve Bayes* memiliki kesalahan terkecil, yaitu 13%.

Penelitian kedua berasal dari Rahmat Deni Putra Setiawan (2022) yang melakukan penerapan metode K-Means *Clustering* untuk menganalisis potensi produksi komoditi buah di Kabupaten Lumajang. Berdasarkan data Dinas Pertanian

Lumajang di tahun 2014, potensi buah-buahan di Kabupaten Lumajang mengalami peningkatan sebesar 80%. Untuk menunjang potensi tersebut dibutuhkan sebuah sistem pengelompokan komoditi buah-buahan berdasarkan data yang diperoleh. Pengelompokan data komoditi buah-buahan menggunakan metode K-Means *Clustering*. Dengan menggunakan metode K-Means *Clustering* data buah-buahan yang sudah ada akan dikelompokkan untuk menentukan potensi produksi supaya dapat membantu pemerintah dalam pengambilan keputusan untuk dilakukan pengembangan lebih lanjut.

Penelitian ketiga berasal dari Dewi Puspa Lamondjong dan Mardi Hardjianto (2021) yang melakukan prediksi menggunakan metode regresi linear pada jumlah hasil pertanian di Kabupaten Banggai, Sulawesi Tengah. Adapun hasil pertaniannya meliputi padi, jagung, ubi kayu dan kedelai. Sumber data diambil dari Dinas Pertanian Kabupaten Banggai dan BMKG Kabupaten Banggai. Penelitian yang dilakukan yaitu memberikan prediksi jumlah penjualan tanaman pangan tersebut dengan memprediksi jumlah dari setiap hasil panen tanaman pangan tersebut pada waktu tertentu. Dengan menggunakan *Confusion Matrix*, dari penelitian tersebut didapatkan hasil prediksi dengan akurasi 96%, presisi 90% dan recall 100%.

2.2 Naïve Bayes

2.2.1 Konsep Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi yang memanfaatkan teori probabilitas untuk memprediksi kelas dari suatu data. Inti dari algoritma ini adalah

menghitung peluang terbesar dari kemungkinan klasifikasi yang ada, dengan menganalisis frekuensi kemunculan tiap kelas pada data latih (training data). Teorema Bayes adalah fondasi utama dari algoritma Naïve Bayes. Teorema ini ditemukan oleh Thomas Bayes, seorang ahli matematika asal Inggris (Sari et al., 2022). Julukan 'naïve' disematkan karena model ini mengasumsikan bahwa setiap fitur input bersifat independen terhadap fitur lainnya jika dikondisikan pada kelas target. Meskipun asumsi ini jarang sepenuhnya terpenuhi di dunia nyata, Naïve Bayes tetap populer karena kesederhanaannya, kecepatan komputasi, dan kemampuannya memberikan hasil yang kompetitif pada banyak aplikasi. Naïve Bayes telah diterapkan di berbagai bidang, termasuk pertanian (Harahap et al., 2023; Putri et al., 2023), deteksi spam (Yuliana & Rahmawati, 2023), klasifikasi berita (Amalia et al., 2022), dan diagnosis penyakit (Rizki & Handayani, 2023).

Algoritma Naïve Bayes memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi (Riani et.al, 2023). Bentuk umum dari teorema bayes seperti di bawah ini:

$$P(C|X) = \frac{[P(X|C) * P(C)]}{P(X)} \quad (2.1)$$

Keterangan :

$P(C|X)$ adalah probabilitas kelas C diberikan fitur X.

$P(X|C)$ adalah probabilitas fitur X diberikan kelas C.

$P(C)$ adalah probabilitas kelas C.

$P(X)$ adalah probabilitas fitur X.

Meskipun tergolong sederhana, algoritma ini memiliki kemampuan klasifikasi yang baik. Naïve Bayes telah terbukti efektif dalam berbagai aplikasi, seperti:

1. Klasifikasi Teks

Naïve Bayes sering digunakan untuk klasifikasi dokumen, seperti identifikasi spam atau sentimen analisis.

2. Analisis Risiko

Naïve Bayes dapat digunakan untuk memprediksi risiko, seperti risiko kredit macet atau risiko kesehatan.

3. Sistem Rekomendasi

Naïve Bayes dapat digunakan untuk membangun sistem rekomendasi produk atau layanan.

Berikut keunggulan dari Metode Naïve Bayes :

1. Sederhana dan mudah diimplementasikan.
2. Efisien dalam hal waktu komputasi dan penggunaan memori.
3. Cocok untuk dataset berdimensi tinggi.
4. Dapat menangani data numerik dan kategorikal.

Berikut kelemahan dari Metode Naïve Bayes :

1. Asumsi independensi antar fitur seringkali tidak realistis.
2. Sensitif terhadap data yang noisy atau tidak lengkap.
3. Performa dapat menurun jika terdapat hubungan yang kuat antar fitur.

Naïve Bayes hadir dalam beberapa varian yang dirancang untuk menangani tipe data yang berbeda. Setiap varian memiliki keunggulan tersendiri sesuai dengan karakteristik dataset yang digunakan:

1. Gaussian Naïve Bayes (GNB) digunakan untuk data numerik kontinu. Pada GNB, setiap fitur diasumsikan mengikuti distribusi normal dalam tiap kelas. GNB banyak dipakai pada kasus klasifikasi dengan data cuaca, kesehatan, atau produktivitas pertanian. Misalnya, fitur seperti suhu, kelembapan, dan curah hujan cocok ditangani GNB.
2. Multinomial Naïve Bayes digunakan untuk data berupa hitungan atau frekuensi. Metode ini populer dalam klasifikasi dokumen atau teks, seperti analisis berita dan ulasan produk, di mana fitur berupa frekuensi kata.
3. Bernoulli Naïve Bayes digunakan untuk fitur biner. Cocok untuk aplikasi di mana fitur berupa kehadiran atau ketiadaan unsur tertentu, contohnya deteksi spam SMS berdasarkan kehadiran kata kunci tertentu.
4. Complement Naïve Bayes dikembangkan untuk mengatasi bias pada kelas minoritas, terutama dalam klasifikasi teks dengan data yang sangat tidak seimbang. Model ini cenderung lebih stabil dibanding Multinomial Naïve Bayes pada data yang kelasnya timpang.

Setiap varian Naïve Bayes tetap mengandalkan prinsip probabilistik sederhana, tetapi disesuaikan dengan tipe distribusi data sehingga hasil klasifikasi lebih optimal.

2.2.2 Studi Pustaka Naïve Bayes

1. Attamami et al. (2023) menggunakan Naïve Bayes untuk memprediksi penerima bantuan jaminan kesehatan. Hasilnya, Naïve Bayes mencapai akurasi 92.97%.
2. Wirma (2022) menerapkan Naïve Bayes untuk memprediksi tingkat kepuasan pelayanan dokumen kependudukan dan memperoleh hasil yang baik.
3. Apriyani dan Kurniati (2020) membandingkan Naïve Bayes dan SVM dalam klasifikasi penyakit diabetes melitus. Naïve Bayes menghasilkan akurasi 92.0746%.
4. Anisa dan Unjung (2022) menggunakan Naïve Bayes untuk klasifikasi penyakit diabetes dan mendapatkan akurasi 92%.
5. Susana et al. (2022) menerapkan Naïve Bayes untuk menganalisis hak akses internet siswa dan mendapatkan akurasi sebesar 89.83%.
6. Rachman dan Handayani (2021) menggunakan Naïve Bayes untuk memprediksi tingkat kelancaran pembayaran sewa teras UMKM dan memperoleh akurasi 81.81%.

2.3 Regresi Logistik

2.3.1 Konsep Regresi Logistik

Regresi logistik adalah metode statistik yang digunakan untuk memprediksi probabilitas terjadinya suatu peristiwa biner (dua kelas), dengan menggunakan satu atau lebih variabel prediktor. Berbeda dengan regresi linear yang memodelkan hubungan linier antar variabel, regresi logistik memanfaatkan fungsi logit untuk memetakan nilai input ke rentang 0–1, sehingga hasilnya dapat ditafsirkan sebagai

peluang. Regresi logistik banyak dipilih karena selain dapat digunakan untuk klasifikasi, hasil modelnya mudah diinterpretasikan. Model ini mampu menjelaskan pengaruh masing-masing variabel bebas terhadap peluang suatu peristiwa terjadi dalam bentuk odds ratio.

Kelebihan metode regresi logistik adalah lebih fleksibel dibanding teknik lain yaitu:

- a. **Mengenai Asumsi Normalitas:** Salah satu kelebihan utama regresi logistik adalah tidak adanya asumsi distribusi normal untuk variabel independennya. Dengan kata lain, variabel prediktor tidak harus terdistribusi secara normal, memiliki hubungan linier, ataupun kesamaan varians (homoskedastisitas) antar kelompok.
- b. **Mengenai Jenis Variabel:** Model regresi logistik sangat fleksibel karena dapat menangani berbagai jenis skala data untuk variabel prediktornya secara bersamaan, baik yang berskala kontinu (angka), diskrit (hitungan), maupun dikotomi (kategorikal).
- c. **Mengenai Pola Hubungan:** Regresi logistik merupakan pilihan yang sangat tepat ketika hubungan antara variabel terikat (hasil) dengan variabel bebasnya diperkirakan tidak membentuk garis lurus (bersifat non-linier).

Model yang digunakan pada regresi logistik adalah:

$$\text{Log}(P/1 - P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K \quad (2.2)$$

Dalam model regresi logistik, probabilitas (P) dari sebuah hasil ($Y=1$) diprediksi menggunakan serangkaian variabel independen ($X_1, X_2, \text{dst.}$) dengan koefisien regresi (β) masing-masing. Inti dari metode ini adalah mengubah probabilitas menjadi *log-odds* atau logit, yaitu $\text{Log}(P/1-P)$, yang nilainya merupakan kombinasi linier dari variabel-variabel independen tersebut. Nilai logit ini kemudian dapat ditransformasikan kembali untuk mendapatkan nilai probabilitas. Selain itu, seperti yang dikutip dari Tripena (2023), model ini juga menghasilkan nilai *odds ratio* (rasio peluang) untuk setiap variabel, yang berguna untuk menginterpretasikan pengaruh masing-masing prediktor.

Konsep "peluang" (*odds*) sendiri didefinisikan sebagai rasio antara probabilitas suatu peristiwa akan terjadi dan probabilitas peristiwa itu tidak terjadi. Sementara itu, "rasio peluang" (*odds ratio*) adalah perbandingan antara dua nilai peluang. Menurut Gunawan et al. (2025), *odds ratio* untuk sebuah prediktor dalam regresi logistik mengukur seberapa besar peluang suatu hasil akan berubah untuk setiap kenaikan satu unit pada variabel prediktor tersebut. Jika nilainya lebih besar dari 1, artinya peluang meningkat; jika lebih kecil dari 1, artinya peluang menurun.

Regresi logistik memiliki beberapa bentuk variasi yang disesuaikan dengan karakteristik variabel target (*dependent variable*) serta kebutuhan klasifikasi data:

1. Regresi logistik biner: Jenis ini paling umum digunakan ketika variabel target memiliki dua kelas atau kategori, misalnya "tinggi/rendah", "ya/tidak", atau "sakit/sehat". Model ini menghitung peluang suatu objek masuk ke salah satu dari dua kelas tersebut.

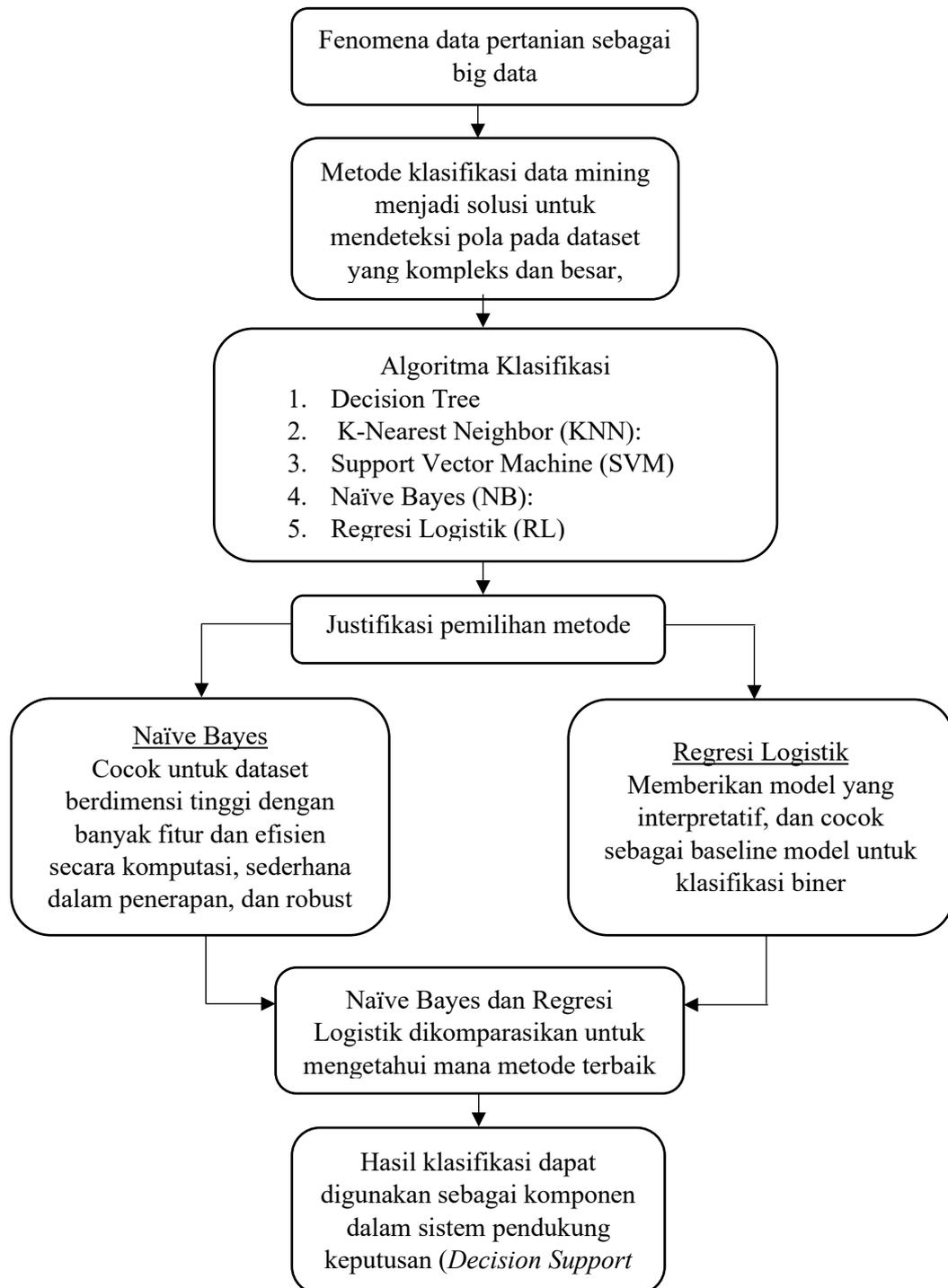
2. Regresi logistik multinomial: Digunakan ketika variabel target memiliki lebih dari dua kategori yang tidak berurutan.
3. Regresi logistik ordinal: Cocok untuk variabel target dengan lebih dari dua kelas yang memiliki urutan alami, seperti “rendah”, “sedang”, “tinggi” pada tingkat kesuburan lahan atau “kurang puas”, “puas”, “sangat puas” pada survei kepuasan petani.

2.3.2 Studi Pustaka Regresi Logistik

1. Wiguna et al. (2022) menggunakan regresi logistik untuk mengklasifikasi kredit macet dalam Koperasi Simpan Pinjam. Hasilnya, model tersebut mencapai akurasi 85%, presisi 82%, dan recall 90%.
2. Lestari et al. (2023) menerapkan regresi logistik untuk memprediksi kejadian diabetes melitus tipe 2. Model tersebut menghasilkan akurasi 78%, presisi 75%, dan recall 83%.
3. Saefullah (2021) menganalisis faktor-faktor yang berhubungan dengan putus sekolah pada siswa SMA menggunakan regresi logistik. Hasilnya menunjukkan akurasi 72%, presisi 70%, dan recall 75%.
4. Firdaus et al. (2022) memprediksi penerimaan beasiswa Bidikmisi menggunakan metode regresi logistik. Model prediksi tersebut mencapai akurasi 80%, presisi 83%, dan recall 76%.
5. Aminah et al. (2023) menganalisis sentimen masyarakat pada penggunaan vaksin COVID-19 di Twitter menggunakan regresi logistik. Penelitian tersebut menunjukkan akurasi sebesar 75%, presisi 73%, dan recall 78%.

2.4 Kerangka Teori

Berikut kerangka teori dalam penelitian ini sebagaimana pada gambar 2.1



Gambar 2.1 Kerangka Teori

Penelitian ini berangkat dari sebuah fenomena yang semakin relevan di era modern, yaitu masifnya data pada sektor pertanian yang kini telah berkembang menjadi "big data". Volume data yang sangat besar, kecepatan pertumbuhannya yang tinggi, serta keragaman jenis data yang mencakup informasi cuaca, kondisi tanah, penggunaan pupuk, hingga hasil panen, menciptakan sebuah tantangan sekaligus peluang. Tantangan utamanya adalah bagaimana mengolah kumpulan data yang begitu besar dan kompleks ini untuk mengekstrak pengetahuan atau pola tersembunyi yang bernilai. Untuk menjawab tantangan tersebut, metode-metode dalam disiplin ilmu data mining hadir sebagai solusi yang efektif. Secara khusus, teknik klasifikasi dalam data mining menjadi pendekatan yang sangat relevan untuk mendeteksi dan mengidentifikasi pola-pola spesifik dalam dataset pertanian yang besar dan rumit tersebut.

Dalam bidang data mining, terdapat beragam algoritma klasifikasi yang telah teruji dan banyak digunakan oleh para peneliti. Beberapa di antaranya yang populer meliputi Decision Tree, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), dan Regresi Logistik (RL). Masing-masing algoritma ini memiliki karakteristik, kelebihan, dan kekurangannya sendiri.

Dari sekian banyak pilihan algoritma, penelitian ini memfokuskan pada dua metode, yaitu Naïve Bayes dan Regresi Logistik. Pemilihan Naïve Bayes didasarkan pada beberapa keunggulannya yang signifikan. Metode ini dikenal sangat cocok untuk menangani dataset berdimensi tinggi yang memiliki banyak sekali fitur atau atribut. Selain itu, Naïve Bayes unggul dalam hal efisiensi komputasi, relatif sederhana dalam proses penerapan, dan memiliki sifat yang

robust atau tangguh terhadap data yang tidak lengkap. Karakteristik ini menjadikannya kandidat yang kuat untuk analisis data pertanian yang seringkali kompleks dan memiliki banyak variabel.

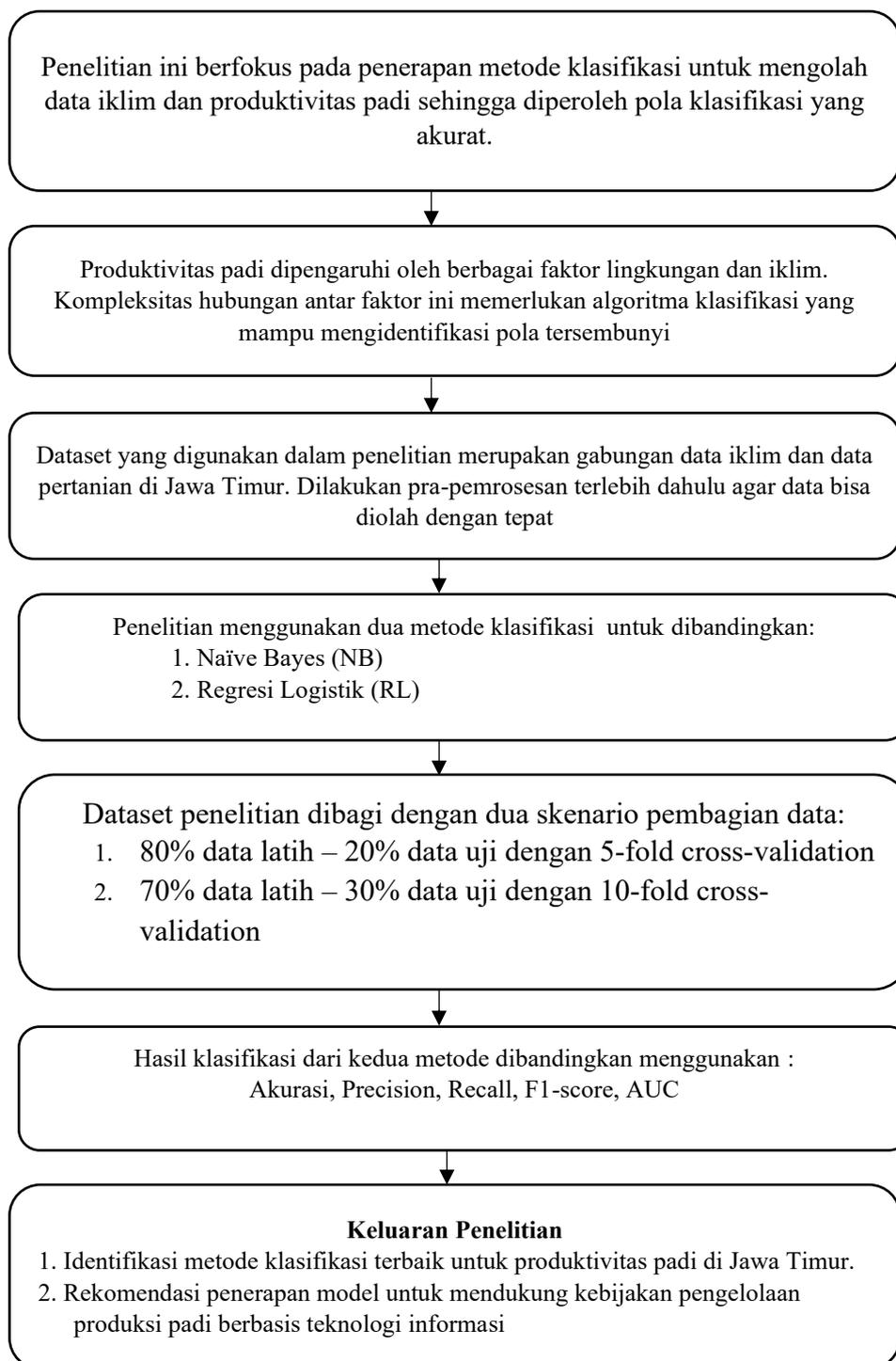
Di sisi lain, Regresi Logistik dipilih sebagai metode pembandingan dengan pertimbangan yang berbeda namun sama pentingnya. Keunggulan utama dari Regresi Logistik adalah kemampuannya untuk menghasilkan model yang sangat interpretatif, artinya hasil dari model ini mudah dipahami dan dijelaskan secara statistik. Hal ini sangat berguna untuk memahami hubungan antara variabel prediktor (seperti cuaca atau jenis tanah) dengan hasil (misalnya, produktivitas panen).

Dengan justifikasi tersebut, inti dari penelitian ini adalah melakukan analisis komparatif atau perbandingan secara langsung antara metode Naïve Bayes dan Regresi Logistik. Tujuan utamanya adalah untuk mengevaluasi dan mengetahui metode mana yang menunjukkan kinerja terbaik dalam melakukan klasifikasi pada dataset pertanian yang digunakan.

Pada akhirnya, hasil dari penelitian ini tidak hanya berhenti pada penentuan model terbaik. Hasil klasifikasi yang akurat, baik dari Naïve Bayes maupun Regresi Logistik, memiliki potensi aplikasi yang lebih luas. Model klasifikasi tersebut dapat diintegrasikan sebagai komponen inti dalam sebuah Sistem Pendukung Keputusan (Decision Support System). Sistem ini nantinya dapat membantu para pemangku kepentingan di sektor pertanian, seperti petani atau pemerintah, dalam membuat keputusan yang lebih baik, lebih cepat, dan berbasis data untuk meningkatkan efisiensi dan produktivitas pertanian.

2.5 Kerangka Konsep

Berikut kerangka konsep dalam penelitian ini sebagaimana pada gambar 2.2



Gambar 2.2 Kerangka Konsep

Penelitian ini dibangun dengan fokus utama pada penerapan metode klasifikasi untuk mengolah data iklim serta data produktivitas padi. Tujuan dari penerapan ini adalah untuk dapat menghasilkan pola klasifikasi yang akurat. Pendekatan ini didasari oleh pemahaman bahwa produktivitas padi sangat dipengaruhi oleh berbagai faktor lingkungan dan iklim yang kompleks. Hubungan yang kompleks antar faktor tersebut memerlukan penggunaan algoritma klasifikasi yang efektif untuk mampu mengidentifikasi pola-pola tersembunyi di dalam data.

Untuk analisisnya, penelitian ini menggunakan dataset gabungan yang mencakup data iklim dan data pertanian dari wilayah Jawa Timur. Dataset tersebut akan melalui tahap pra-pemrosesan terlebih dahulu untuk memastikan data dapat diolah secara tepat dan akurat. Dua metode klasifikasi utama yang dipilih untuk diimplementasikan dan dibandingkan adalah Naïve Bayes (NB) dan Regresi Logistik (RL).

Eksperimen akan dilakukan dengan menggunakan dua skenario pembagian data. Skenario pertama adalah pembagian 80% data latih dan 20% data uji yang divalidasi dengan 5-fold cross-validation. Skenario kedua menggunakan rasio 70% data latih dan 30% data uji yang divalidasi dengan 10-fold cross-validation. Hasil klasifikasi dari kedua metode tersebut selanjutnya akan dibandingkan performanya menggunakan berbagai metrik evaluasi, yaitu Akurasi, Precision, Recall, F1-score, dan AUC.

Pada akhirnya, penelitian ini diharapkan dapat memberikan dua keluaran utama. Pertama adalah identifikasi metode klasifikasi terbaik yang paling sesuai untuk kasus produktivitas padi di Jawa Timur. Kedua, penelitian ini akan

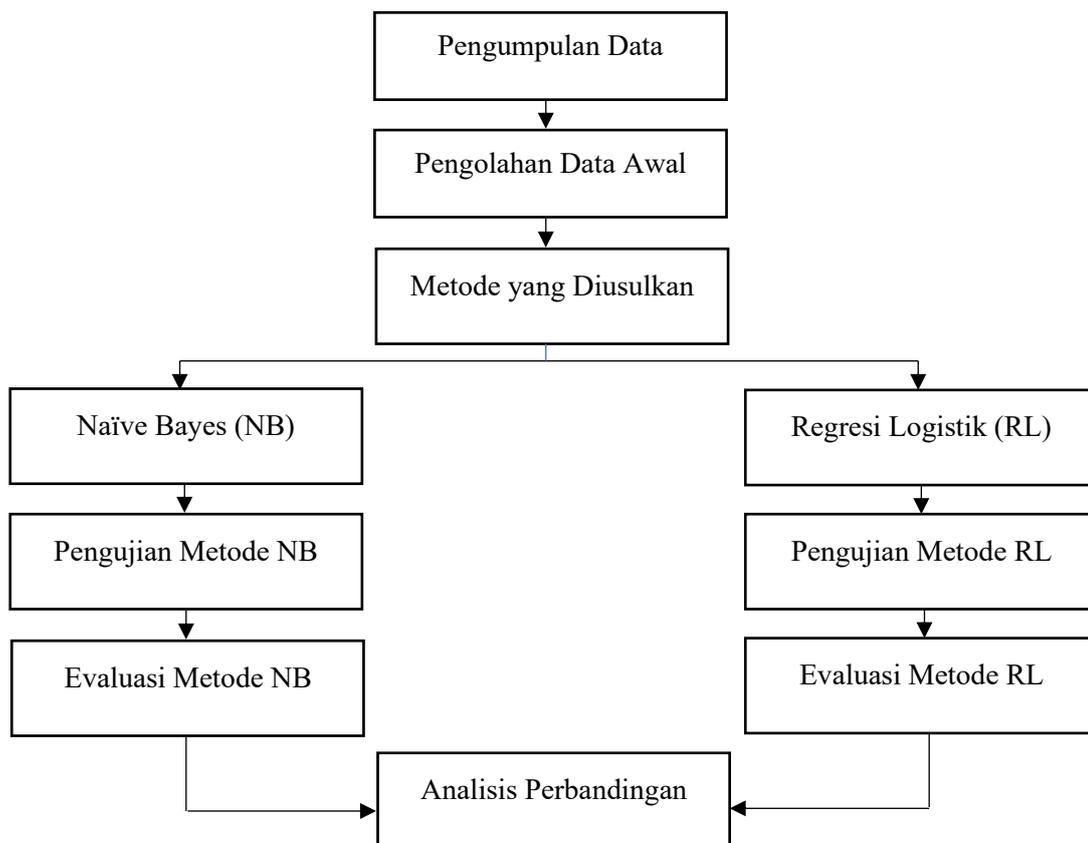
menghasilkan rekomendasi penerapan model tersebut untuk mendukung kebijakan pengelolaan produksi padi yang berbasis pada teknologi informasi.

BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komparatif. Dua metode klasifikasi, yaitu Naïve Bayes (NB) dan Regresi Logistik (RL), akan diimplementasikan untuk mengklasifikasi produktivitas padi di Jawa Timur. Hasil klasifikasi dari kedua metode tersebut akan dibandingkan untuk mengetahui metode mana yang memiliki performa terbaik. Adapun desain penelitian ini terdapat di **gambar 3.1**.



Gambar 3.1 Tahapan Penelitian

Keterangan :

1. Pengumpulan Data

Pengumpulan data merupakan teknik atau metode yang digunakan untuk mengumpulkan data yang akan diteliti. Data penelitian dikumpulkan dan diperoleh dari sumber sekunder, yaitu Badan Pusat Statistik (BPS) Provinsi Jawa Timur.

2. Pengolahan Data Awal

Data yang telah dikumpulkan di awal akan diproses terlebih dahulu sehingga dapat diperoleh data yang lebih valid dan berkualitas. Dalam tahap ini dilakukan beberapa langkah meliputi : pembersihan data (*data cleaning*) dan transformasi data (*data transformation*).

3. Metode yang Diusulkan

Proses dalam tahapan ini adalah penentuan metode yang akan digunakan untuk mengolah data yang telah melalui tahapan pengolahan awal data. Metode yang diusulkan adalah Metode Naïve Bayes dan Regresi Logistik.

4. Pengujian Metode

Pada tahap ini dilakukan pengujian pada hasil klasifikasi dari metode yang diusulkan dan data yang telah diperoleh. Metode akan diuji dengan *k-fold cross-validation*.

5. Evaluasi Metode

Pada tahapan ini dilakukan evaluasi terhadap tingkat akurasi dari masing-masing metode klasifikasi untuk melihat kinerja setiap metode yang

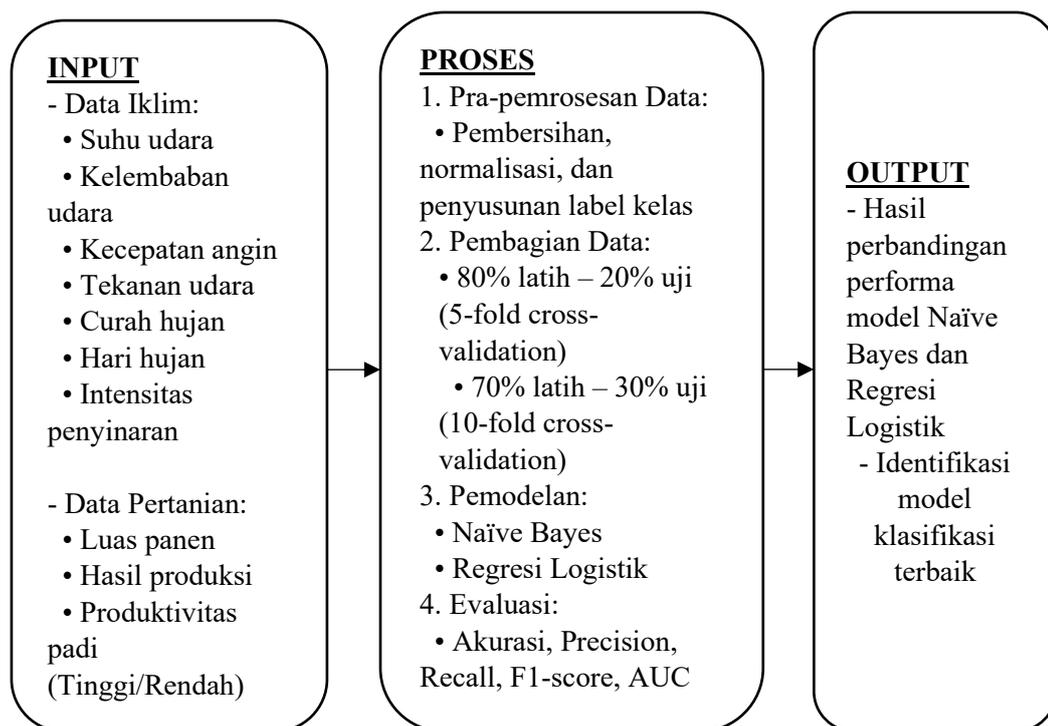
digunakan. Metode yang diusulkan akan dievaluasi dengan *Confusion Matrix* dan Kurva ROC.

6. Analisis Perbandingan

Pada tahapan ini, hasil evaluasi model dari kedua metode akan diperbandingkan, sehingga dapat diketahui metode mana yang memiliki performa yang lebih baik.

3.2 Desain Sistem

Berikut desain sistem penelitian sebagaimana gambar 3.2



Gambar 3.2 Desain Sistem

Desain sistem penelitian ini dirancang dengan kerangka kerja yang terdiri dari tiga tahapan utama: masukan (Input), proses (Proses), dan keluaran (Output). Pada tahap masukan, sistem akan menerima dua jenis data utama. Jenis pertama

adalah Data Iklim, yang mencakup atribut-atribut seperti suhu udara, kelembaban udara, kecepatan angin, tekanan udara, curah hujan, jumlah hari hujan, dan intensitas penyinaran. Jenis kedua adalah Data Pertanian, yang terdiri dari data luas panen, hasil produksi, serta variabel target yaitu produktivitas padi yang telah dikategorikan menjadi dua kelas: "Tinggi" dan "Rendah".

Seluruh data masukan tersebut kemudian akan memasuki tahap Proses yang terdiri dari empat langkah berurutan. Langkah pertama adalah Pra-pemrosesan Data, di mana data akan melalui proses pembersihan, normalisasi, dan penyusunan label kelas agar siap untuk dimodelkan. Langkah kedua adalah Pembagian Data, yang dilakukan melalui dua skenario eksperimen: skenario pertama dengan rasio 80% data latih dan 20% data uji yang divalidasi menggunakan 5-fold cross-validation, dan skenario kedua dengan rasio 70% data latih dan 30% data uji yang menggunakan 10-fold cross-validation. Langkah ketiga adalah Pemodelan, di mana data latih akan digunakan untuk membangun dua model klasifikasi, yaitu model Naïve Bayes dan model Regresi Logistik. Langkah terakhir dalam proses ini adalah Evaluasi, di mana performa kedua model akan diukur dan dibandingkan menggunakan metrik Akurasi, Precision, Recall, F1-score, dan AUC.

Dari keseluruhan proses tersebut, sistem akan menghasilkan dua keluaran (Output) utama. Pertama adalah laporan hasil perbandingan performa antara model Naïve Bayes dan Regresi Logistik berdasarkan metrik evaluasi yang telah ditentukan. Kedua, berdasarkan hasil perbandingan tersebut, akan dilakukan

identifikasi untuk menentukan model klasifikasi mana yang terbaik untuk kasus ini.

3.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang dikumpulkan dan diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur tahun 2022. Data tersebut berupa data bulanan kota atau kabupaten di Jawa Timur. Variabel-variabel yang digunakan dalam penelitian ini meliputi:

- Variabel X (Atribut): suhu udara ($^{\circ}\text{C}$), kelembaban udara (%), kecepatan angin (m/s), tekanan udara (mb), curah hujan (mm), hari hujan (hari), penyinaran (%), luas panen padi (ha).
- Variabel Y (Label): Produktivitas padi (ton/ha).

Sampel data yang dikumpulkan dan diperoleh dapat dilihat pada **tabel 3.1**.

Tabel 3.1 Sampel Data Pertanian

Kabupaten/Kota	Bulan	Suhu	Kelembaban	Kecepatan Angin	Tekanan Udara	Curah Hujan	Hari Hujan	Penyinaran	Luas Panen	Produksi Padi	Produktivitas Padi
Kab. Blitar	Januari	28	76	2,6	977	182	22	4,8	897,4	5895,5	6,57
Kab. Malang	Januari	24	81	2,4	945	216	25	4,4	4233,9	26207,0	6,19
Kab. Tuban	Januari	23	85	1,5	1002	111	18	4,7	791,8	5039,6	6,36
Kab. Jember	Januari	27	81	2,7	1003	335	21	4,1	6812,3	35404,4	5,20
Kab. Sumenep	Januari	28	85	4,8	1010	421	24	4,5	613,8	3403,1	5,54
Kab. Bondowoso	Januari	27	81	2,7	998	335	21	4,1	2918,1	14393,7	4,93
Kab. Situbondo	Januari	27	81	2,7	998	179	25	4,1	1158,1	6141,4	5,30
Kab. Probolinggo	Januari	21	94	1,4	915	324	13	2,0	1205,1	7271,9	5,56
Kab. Pasuruan	Januari	21	94	1,4	915	235	28	2,0	1953,7	10203,2	5,22
Kab. Sidoarjo	Januari	28	97	9,2	1010	377	25	5,4	279,9	1716,7	6,13
Kota Malang	Januari	25	79	2,8	944	216	23	4,0	71,3	453,9	6,37
Kota Probolinggo	Januari	22	91	0,9	911	683	26	1,8	18,9	118,8	6,29
Kota Pasuruan	Januari	27	94	1,4	915	683	28	2,0	38,0	204,7	5,39
Kota Surabaya	Januari	28	82	1,8	1009	265	19	4,4	16,7	98,3	5,89
Kota Batu	Januari	25	79	2,8	944	216	29	4,0	113,8	686,1	6,03

Selengkapnya terdapat di lampiran

3.4 Pra-Pemrosesan Data

Tahap pra-pemrosesan data yang dilakukan meliputi :

1. Pembersihan Data (*Data Cleaning*)

Tahap pembersihan data mencakup identifikasi dan penanganan *missing values* serta data yang tidak konsisten. Data yang hilang diperiksa dan ditangani dengan teknik seperti imputasi atau penghapusan baris, bergantung pada proporsi dan dampaknya terhadap analisis. Data tidak konsisten, misalnya nilai ekstrim atau kesalahan input, dikoreksi atau dibersihkan agar kualitas data terjaga dan hasil pemodelan lebih andal.

2. Transformasi Data (*Data Transformation*)

Transformasi data dilakukan untuk memastikan kesiapan data dalam pemodelan. Data numerik distandarisasi atau dinormalisasi bila diperlukan agar skala seragam dan tidak ada fitur yang mendominasi. Variabel produktivitas padi diubah menjadi kategorik berdasarkan acuan nilai 5,63 ton/ha, sesuai rata-rata produktivitas padi Jawa Timur 2022 (BPS Jatim, 2022), dengan klasifikasi: di atas 5,63 ton/ha sebagai **Tinggi**, dan di bawah atau sama dengan 5,63 ton/ha sebagai **Rendah**. Justifikasi acuan ini dipilih agar klasifikasi mencerminkan kondisi aktual di wilayah penelitian.

3. Seleksi Fitur (*Feature Selection*)

Dalam penelitian ini dilakukan seleksi fitur untuk mencegah overfitting dan kebocoran target (*data leakage*). Fitur produksi padi dan luas lahan dihilangkan karena keduanya membentuk langsung nilai produktivitas padi, sehingga berisiko membuat model belajar secara tidak wajar. Fitur hari hujan juga dikeluarkan karena korelasinya sangat rendah terhadap target produktivitas. Seleksi ini bertujuan menyederhanakan model dan meningkatkan kemampuan generalisasi

Berikut ilustrasi dari identifikasi missing value saat proses data cleaning terdapat di **tabel 3.2**.

Tabel 3.2 Ilustrasi Missing Values

Suhu	Kelembaban	Kecepatan Angin	Tekanan Udara	Curah Hujan	Hari Hujan	Penyinaran	Luas Panen	Produktivitas Padi
26	77	2,5	934	21	2	7,4	133,4	4,68
26	74	2,8	935	103	3	7,6	390,1	4,68
27	70	2,8	935	18	3	7,3	71,7	6,23
25	83	2,3	934	307	13	4,1	24,9	6,23
25	85	2,4	934	584	16	2,6	79,1	6,23
25	84	2,5	933	218	12	3,6	164,2	6,23
28	76	2,6	977	248	23	4,8	6,2	7,64
28	78	2,5	977	264	21	4,3	?	?
28	78	2,6	976	266	27	5,3	30,1	7,64
28	78	2,5	977	155	22	5,4	433,7	7,64
27	75	2,3	978	153	27	5,8	?	?

Data yang tidak ada nilainya dihilangkan sehingga dari 456 record diperoleh 405 record. Dari 405 data, jumlah data kelas produktivitas padi “Tinggi” yakni sebanyak 214 record dan jumlah data kelas produktivitas padi “Rendah” sebanyak 191 record. Tabel pengkategorian setelah dilakukan transformasi data dan seleksi fitur terlihat pada **tabel 3.3**

Tabel 3.3 Hasil Transformasi dan Seleksi Fitur

Suhu	Kelembaban	Kecepatan Angin	Tekanan Udara	Curah Hujan	Penyinaran	Produktivitas Padi
27	83	3,0	929,6	82,3	2,4	Rendah
27	84	2,5	928,5	589,0	2,4	Rendah
26	88	3,6	929,7	364,0	1,1	Tinggi
26	88	2,0	929,2	403,0	1,2	Tinggi
27	84	2,2	929,2	192,0	1,6	Tinggi
27	79	3,6	930,3	176,0	2,1	Tinggi

27	80	2,6	930,4	120,0	3,0	Rendah
26	81	2,2	930,9	167,0	3,9	Rendah
26	78	4,2	931,0	4,0	4,2	Rendah
25	77	5,6	931,5	16,0	4,3	Rendah
26	87	7,1	930,3	30,0	3,5	Tinggi

3.5 Metode yang Diusulkan

Dalam penelitian ini akan dilakukan analisis komparasi menggunakan dua metode klasifikasi. Metode yang diusulkan untuk pengolahan data adalah Gaussian Naïve Bayes (GNB) dan Regresi Logistik (RL). Kombinasi kedua metode ini diharapkan mampu memberikan gambaran menyeluruh mengenai performa model pada data produktivitas padi di Jawa Timur.

3.5.1 Gaussian Naïve Bayes

Metode Gaussian Naïve Bayes dipilih karena sangat cocok untuk menangani fitur numerik seperti suhu, kelembaban udara, kecepatan angin, tekanan udara, curah hujan, dan penyinaran, yang semuanya tersedia dalam dataset produktivitas padi ini. Selain itu, GNB dikenal memiliki kecepatan tinggi dalam proses pelatihan dan prediksi, karena hanya memerlukan perhitungan statistik sederhana berupa rata-rata dan varians. Hal ini menjadikannya efisien terutama untuk eksperimen dengan banyak kombinasi validasi.

3.5.2 Regresi Logistik

Regresi Logistik dipilih karena memberikan interpretasi yang kuat melalui koefisien yang dapat dikonversi menjadi odds ratio, sehingga memudahkan analisis hubungan antara variabel prediktor dan peluang suatu kelas produktivitas padi.

3.6 Hyperparameter Tuning

Hyperparameter tuning merupakan tahap penting dalam penelitian machine learning yang bertujuan untuk mencari konfigurasi parameter terbaik pada model, sehingga diperoleh performa yang optimal.

Pada penelitian ini, proses tuning dilakukan secara sistematis menggunakan teknik *Grid Search* dengan bantuan pustaka *Scikit-learn*. Proses tuning dikombinasikan dengan *k-fold cross validation* untuk memastikan hasil *tuning* tidak *overfitting* terhadap subset data tertentu. *GridSearchCV* atau *Grid Search Cross-Validation* adalah metode yang digunakan dalam machine learning untuk mengoptimalkan hyperparameter model dengan cara yang sistematis dan efisien

3.6.1 Hyperparameter Tuning pada Gaussian Naïve Bayes

Gaussian Naïve Bayes (GNB) memiliki satu hyperparameter penting, yaitu *var_smoothing* yang menentukan nilai kecil yang ditambahkan ke variansi untuk mencegah pembagian dengan nol.

Nilai grid yang diuji sebagaimana dalam tabel 3.4.

Tabel 3.4 Nilai Grid Hyperparameter Gaussian Naïve Bayes

No.	var_smoothing
1	1e-9
2	1e-8
3	1e-7
4	1e-6

3.6.2 Hyperparameter Tuning pada Regresi Logistik

Regresi logistik memiliki beberapa hyperparameter penting yang dituning, yaitu :

- penalty: L1 (Lasso), L2 (Ridge)
- C: Kebalikan dari regularisasi (semakin besar C, semakin lemah regularisasi)
- solver: Algoritma optimasi (liblinear, saga)

Nilai grid yang diuji sebagaimana dalam tabel 3.5.

Tabel 3.5 Nilai Grid Hyperparameter Regresi Logistik

No.	penalty	C	solver
1	L2	0,01	Liblinear
2	L2	0,1	Liblinear
3	L2	1	Liblinear
4	L2	10	Liblinear
5	L1	0,01	Saga
6	L1	0,1	Saga
7	L1	1	Saga
8	L1	10	Saga

3.6.3 Proses Hyperparameter Tuning

Proses tuning pada penelitian ini dilakukan dengan tahapan yang sistematis untuk memperoleh kombinasi nilai hyperparameter yang mampu menghasilkan performa model terbaik secara rata-rata pada berbagai skenario validasi. Langkah pertama adalah menentukan grid nilai hyperparameter, yaitu sekumpulan nilai kandidat yang akan diuji, misalnya untuk Regresi Logistik digunakan grid nilai parameter regularisasi C pada rentang tertentu (contohnya 0.01, 0.1, 1, 10, dan 100), sedangkan untuk Gaussian Naïve Bayes digunakan grid nilai `var_smoothing`

(misalnya $1e-9$, $1e-8$, hingga $1e-6$). Selanjutnya, dilakukan Grid Search yang dikombinasikan dengan teknik *Cross Validation* (dalam hal ini 5-Fold Cross Validation) untuk memastikan setiap kombinasi hyperparameter dievaluasi secara adil di berbagai subset data dan mengurangi potensi bias akibat pembagian data tertentu. Pada setiap kombinasi nilai hyperparameter, model dievaluasi dengan metrik performa utama seperti akurasi dan F1-score, sehingga setiap pasangan hyperparameter akan menghasilkan nilai rata-rata metrik performa dari seluruh lipatan validasi. Terakhir, kombinasi hyperparameter yang menghasilkan nilai metrik performa terbaik secara rata-rata (misalnya akurasi tertinggi atau F1-score tertinggi) dipilih sebagai konfigurasi optimal yang akan digunakan untuk membangun model final pada tahap pengujian akhir, sehingga diharapkan model mampu melakukan generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

3.7 Validasi Data

Validasi data merupakan salah satu tahap penting dalam siklus pembangunan model machine learning yang bertujuan untuk menilai kemampuan model dalam melakukan generalisasi terhadap data baru yang tidak pernah digunakan pada tahap pelatihan. Validasi digunakan untuk mengukur apakah model hanya sekadar menghafal data pelatihan (overfitting) atau benar-benar mampu mempelajari pola-pola penting dari data (Wahyudi et al., 2022). Pada penelitian ini, validasi dilakukan dengan dua pendekatan utama yaitu *hold-out validation* dan *K-Fold Cross Validation (CV)*. Pendekatan ganda ini bertujuan untuk memperoleh

evaluasi performa model yang lebih komprehensif dan robust sebagaimana dianjurkan oleh Harahap et al. (2023).

Hold-out validation digunakan sebagai baseline evaluasi awal untuk memperoleh gambaran performa model secara cepat dan sederhana. Metode ini sangat berguna untuk memberikan hasil awal yang dapat dijadikan acuan dalam mengembangkan model lebih lanjut. Di sisi lain, *K-Fold Cross Validation* dipilih karena mampu memberikan estimasi performa yang lebih akurat dan stabil dibanding hold-out, mengingat seluruh data digunakan secara bergantian sebagai data latih dan data uji pada setiap iterasi. Hal ini memungkinkan model dievaluasi secara menyeluruh terhadap seluruh data yang tersedia, sehingga risiko overfitting dapat ditekan dan hasil evaluasi lebih dapat diandalkan.

Penelitian ini secara khusus merancang berbagai kombinasi skenario validasi, yaitu pembagian data 80:20 dan 70:30 untuk hold-out, serta penggunaan 5-Fold dan 10-Fold pada K-Fold Cross Validation. Tujuannya adalah untuk menguji sensitivitas performa model terhadap berbagai pola pembagian data, sekaligus memastikan hasil penelitian ini tidak bias terhadap skenario validasi tertentu dan benar-benar mampu merepresentasikan kemampuan generalisasi model secara optimal.

3.7.1 Hold-Out Validation

Hold-out validation merupakan metode validasi sederhana yang membagi dataset menjadi dua subset secara acak: data latih dan data uji. Data latih digunakan untuk

membangun model, sementara data uji digunakan untuk mengukur kinerja model terhadap data baru (Han et al., 2012). Dalam penelitian ini, digunakan dua skenario:

- Split 1: 80% data latih, 20% data uji — skenario umum dalam machine learning
- Split 2: 70% data latih, 30% data uji — untuk menguji sensitivitas model terhadap proporsi data latih.

Ilustrasi Hold-Out dengan total data: 405 (214 Tinggi, 191 Rendah)

- 80:20 split

Data latih: 171 Tinggi, 153 Rendah → 324 data

Data uji: 43 Tinggi, 38 Rendah → 81 data

- 70:30 split

Data latih: 150 Tinggi, 134 Rendah → 284 data

Data uji: 64 Tinggi, 57 Rendah → 121 data

3.7.2 K-Fold Cross Validation

K-Fold Cross Validation adalah metode validasi yang membagi dataset ke dalam K lipatan (fold) dan memanfaatkan setiap lipatan sebagai data uji secara bergantian (Han et al., 2012). K-Fold CV dinilai lebih stabil dibanding hold-out karena seluruh data digunakan sebagai data uji dan latih pada siklus berbeda. Dalam penelitian ini digunakan dua konfigurasi:

- 5-Fold CV: seimbang antara akurasi evaluasi dan waktu komputasi.
- 10-Fold CV: memberikan estimasi performa lebih stabil dan variansi lebih rendah.

Proses K-Fold CV:

1. Dataset dibagi menjadi K subset dengan ukuran hampir sama.

2. Pada iterasi ke- i , subset ke- i digunakan sebagai data uji, sisanya data latih.
3. Model dibangun dan dievaluasi.
4. Proses diulang hingga seluruh subset pernah menjadi data uji.
5. Metrik akurasi, presisi, recall, F1-score, dan AUC dihitung rata-rata dari seluruh iterasi.

3.7.3 Kombinasi Eksperimen Validasi

Untuk menjamin kekuatan generalisasi model dan menghindari bias evaluasi akibat pemilihan skenario validasi tertentu, penelitian ini merancang empat kombinasi sebagaimana dalam tabel 3.6.

Tabel 3.6 Kombinasi Eksperimen

No	Metode	Data Split	K-Fold CV
1	Gaussian Naïve Bayes	80:20	5-Fold
2	Gaussian Naïve Bayes	70:30	10-Fold
3	Regresi Logistik	80:20	5-Fold
4	Regresi Logistik	70:30	10-Fold

Setiap kombinasi akan dievaluasi dengan metrik: akurasi, presisi, recall, F1-score, dan AUC-ROC. Evaluasi dilakukan secara berulang untuk memastikan hasil tidak bias terhadap pola data tertentu dan dapat dibandingkan secara adil antara kedua metode (GNB dan Regresi Logistik).

3.8 Evaluasi Metode

3.8.1 Confusion Matrix

Confusion Matrix adalah sebuah tabel visual yang umum dimanfaatkan dalam *supervised learning* untuk mengevaluasi kinerja sebuah model klasifikasi. Alat ini secara efektif menyajikan perbandingan antara hasil yang diprediksi oleh model dengan kelas aktual (yang sebenarnya). Dalam strukturnya, setiap baris pada matriks melambangkan kelas data yang sesungguhnya, sementara setiap kolom melambangkan kelas yang diprediksi oleh model. Sebagai ilustrasi, Tabel 3.7 memberikan contoh penerapan *confusion matrix* untuk kasus klasifikasi dengan dua kelas.

Tabel 3.7 Model Confusion Matrix

		Predited	
		Negative	Positive
Actual	Negative	a	c
	Positive	b	d

Untuk mengevaluasi performa sebuah model klasifikasi, digunakan *Confusion Matrix* yang memuat empat komponen dasar.

True Negative (TN) atau **(a)**: Jumlah data aktual negatif yang diprediksi dengan benar sebagai negatif.

False Positive (FP) atau **(b)**: Jumlah data aktual negatif yang keliru diprediksi sebagai positif. (Dikenal juga sebagai Error Tipe I).

False Negative (FN) atau **(c)**: Jumlah data aktual positif yang keliru diprediksi sebagai negatif. (Dikenal juga sebagai Error Tipe II).

True Positive (TP) atau (d): Jumlah data aktual positif yang diprediksi dengan benar sebagai positif.

Berdasarkan keempat komponen tersebut, beberapa metrik evaluasi standar dapat dihitung untuk mengukur kinerja model:

a. **Akurasi (Accuracy)**

Metrik ini mengukur proporsi keseluruhan prediksi yang benar (baik positif maupun negatif) dibandingkan dengan total data.

b. **Recall (Sensitivitas atau True Positive Rate)**

Recall mengukur kemampuan model untuk "menemukan kembali" semua kasus positif yang ada. Metrik ini menunjukkan proporsi dari seluruh kasus positif aktual yang berhasil diidentifikasi dengan benar oleh model.

c. **Presisi (Precision)**

Presisi mengukur tingkat ketepatan dari prediksi positif yang dibuat. Metrik ini menunjukkan proporsi dari prediksi positif yang ternyata memang benar positif.

d. **Tingkat Positif Salah (False Positive Rate - FPR)**

Metrik ini mengukur proporsi kasus negatif aktual yang salah diklasifikasikan sebagai positif.

e. **Tingkat Negatif Sejati (True Negative Rate - TNR atau Spesifitas)**

Metrik ini adalah kebalikan dari FPR, yaitu mengukur proporsi kasus negatif aktual yang berhasil diidentifikasi dengan benar.

f. Tingkat Negatif Palsu (False Negative Rate - FNR)

Metrik ini mengukur proporsi kasus positif aktual yang gagal diidentifikasi atau salah diklasifikasikan sebagai negatif.

3.8.1 Kurva ROC

Kurva ROC (*Receiver Operating Characteristic*) adalah sebuah grafik yang berfungsi sebagai alat visual untuk mengevaluasi dan membandingkan performa model-model klasifikasi dalam *supervised learning*. Pada dasarnya, kurva ini merupakan representasi dua dimensi dari informasi yang ada dalam *confusion matrix*. Grafik ROC memetakan hubungan antara Tingkat Positif Benar (*True Positive Rate* - TPR) pada sumbu vertikal (Y) dan Tingkat Positif Salah (*False Positive Rate* - FPR) pada sumbu horizontal (X).

Interpretasi dari posisi titik pada grafik ROC sangatlah penting. Titik ideal berada di pojok kiri atas **(0,1)**, yang menandakan sebuah model klasifikasi sempurna karena berhasil mengidentifikasi semua kasus positif ($TPR = 1$) tanpa ada kesalahan dalam memprediksi kasus negatif sebagai positif ($FPR = 0$). Titik lainnya memiliki arti spesifik: titik **(0,0)** merepresentasikan model yang selalu memprediksi semua kasus sebagai negatif, sedangkan titik **(1,1)** adalah model yang selalu memprediksi semua kasus sebagai positif.

Kurva ROC sendiri terbentuk dari kumpulan titik (pasangan nilai FPR dan TPR) yang dihasilkan saat parameter atau ambang batas (*threshold*) dari sebuah model klasifikasi diubah-ubah. Dengan menyesuaikan parameter ini, kita dapat melihat bagaimana model menyeimbangkan kemampuannya dalam mendeteksi

kelas positif (meningkatkan TP) dengan risiko salah mengklasifikasikan kelas negatif (penurunan FP). Model klasifikasi yang tidak memiliki parameter yang bisa diatur (non-parametrik) hanya akan diwakili oleh satu titik tunggal pada grafik ini. Untuk menyederhanakan perbandingan antar model, sering digunakan metrik **Area di Bawah Kurva (Area Under the Curve - AUC)**. Nilai AUC merangkum keseluruhan performa kurva ROC menjadi satu angka tunggal. Kinerja model berdasarkan nilai AUC dapat dikelompokkan sebagai berikut:

- **0.90 – 1.00**: Klasifikasi Sempurna (*Excellent Classification*)
- **0.80 – 0.90**: Klasifikasi Baik (*Good Classification*)
- **0.70 – 0.80**: Klasifikasi Cukup (*Fair Classification*)
- **0.60 – 0.70**: Klasifikasi Buruk (*Poor Classification*)
- **0.50 – 0.60**: Gagal (*Failure*) atau tidak lebih baik dari tebakan acak.

3.9 Novelty Penelitian

Berikut beberapa poin yang menjadi novelty dalam penelitian ini :

1. Meskipun banyak penelitian mengenai prediksi produksi padi, penelitian yang secara spesifik fokus pada klasifikasi produktivitas padi di Jawa Timur masih terbatas.
2. Beberapa penelitian telah membandingkan metode Naïve Bayes dan Regresi Logistik dalam berbagai konteks, namun belum banyak yang diaplikasikan pada kasus klasifikasi produktivitas padi. Penelitian ini akan memberikan informasi berharga tentang performa kedua metode ini dalam memprediksi tingkat produktivitas padi.

3. Penelitian ini menggunakan data terkini dan komprehensif tentang produktivitas padi di Jawa Timur, serta mempertimbangkan faktor-faktor yang relevan, seperti suhu, kelembaban udara, kecepatan angin, tekanan udara, curah hujan, hari hujan dan penyinaran.

4. Interpretasi dan Implikasi Hasil

Penelitian ini dapat memberikan interpretasi yang mendalam tentang hasil analisis, menjelaskan faktor-faktor penting yang mempengaruhi produktivitas padi, dan memberikan rekomendasi yang bermanfaat bagi para stakeholder, seperti petani, pengambil kebijakan, dan peneliti.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Statistik Deskriptif

Sebelum dilakukan pemodelan, dilakukan eksplorasi statistik deskriptif terhadap fitur numerik: Suhu, Kelembaban Udara, Kecepatan Angin, Tekanan Udara, Curah Hujan, dan Penyinaran. Transformasi dengan `StandardScaler` memastikan setiap fitur memiliki rata-rata 0 dan deviasi standar 1, sehingga skala antar fitur seimbang dan tidak ada fitur yang mendominasi. Hal ini penting untuk algoritma berbasis jarak atau yang sensitif terhadap skala. Berikut ringkasan statistiknya setelah normalisasi menggunakan `StandardScaler` sebagaimana dalam tabel 4.1

Tabel 4.1 Statistik Deskriptif Data

Fitur	Mean	Std	Min	25%	50%	75%	Max
Suhu	≈ 0	1	-2,23	-0,58	0,24	0,84	1,66
Kelembaban Udara	≈ 0	1	-2,4	-0,7	0,1	0,51	3
Kecepatan Angin	≈ 0	1	-1,56	-0,6	-0,22	0,41	3,84
Tekanan Udara	≈ 0	1	-1,82	-1,05	0,73	0,97	1,09
Curah Hujan	≈ 0	1	-1,38	-0,7	-0,15	0,55	3,87
Penyinaran	≈ 0	1	-2,35	-0,73	0	0,83	2,21

4.2 Hasil Hyperparameter Tuning

4.2.1 Gaussian Naïve Bayes

Pada Gaussian Naïve Bayes, hanya satu hyperparameter yang dituning, yaitu `var_smoothing`. Nilai grid yang diuji adalah $\{1e-9, 1e-8, 1e-7, 1e-6\}$. Hasil dari proses hyperparameter tuning menunjukkan bahwa konfigurasi model terbaik

ditemukan saat nilai **var_smoothing** adalah $2,31 \times 10^{-6}$ (atau 0.00000231). Dengan menggunakan parameter tersebut, model berhasil mencapai skor performa terbaik (*Best Score*) sebesar 0.6244, atau setara dengan akurasi validasi silang (cross-validation accuracy) sebesar 62,24%.

```
Best parameters: {'var_smoothing': np.float64(2.310129700083158e-06)}
Best score: 0.6244512195121952
```

4.2.2 Regresi Logistik

Untuk memperoleh performa Regresi Logistik yang optimal, dilakukan proses hyperparameter tuning dengan metode *GridSearchCV*. Tuning difokuskan pada tiga parameter kunci: **C** (kekuatan regularisasi), **penalty** (jenis norma regularisasi), dan **solver** (algoritma optimisasi).

Hasil dari proses tuning menunjukkan konfigurasi parameter terbaik sebagai berikut : **C**: 0.1, **penalty**: 'L1', **solver**: 'liblinear'. Dengan menggunakan kombinasi parameter tersebut, model Regresi Logistik mencapai skor performa terbaik (*Best Score*) sebesar 0,6255 yang setara dengan akurasi validasi silang rata-rata 62,55%.

```
Best parameters found: {'C': 0.1, 'penalty': 'l1', 'solver': 'liblinear'}
Best score: 0.6255012531328321
```

4.3 Hasil Eksperimen I

Gaussian Naive Bayes (80:20 Split, 5-Fold Cross-Validation)

Eksperimen pertama menggunakan model Gaussian Naive Bayes (GNB) dengan pembagian data sebesar 80% untuk data latih dan 20% untuk data uji. Pada tahap

validasi, digunakan teknik 5-fold cross-validation untuk memastikan bahwa model diuji pada berbagai subset data dan hasil yang diperoleh tidak bias terhadap distribusi tertentu. Strategi ini bertujuan untuk meningkatkan keandalan hasil evaluasi dan mengurangi kemungkinan overfitting.

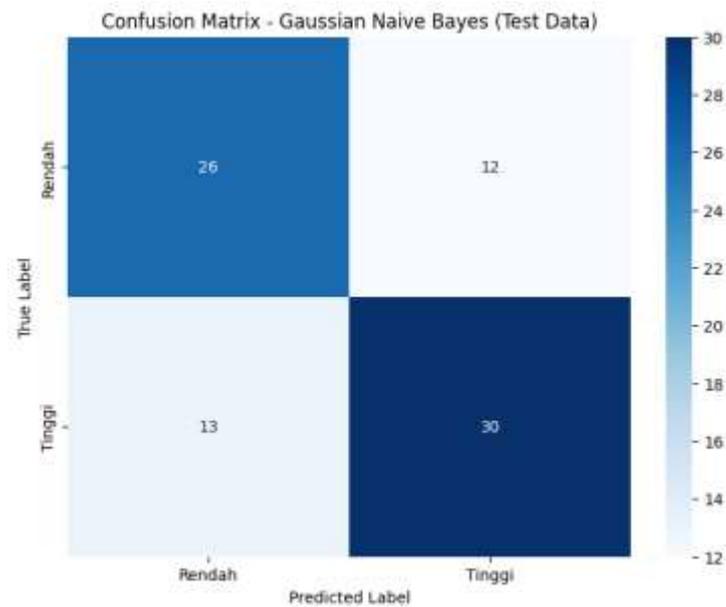
Hasil rata-rata dari evaluasi performa model GNB pada skenario ini adalah sebagaimana dalam tabel 4.2

Tabel 4.2 Hasil Evaluasi Eksperimen 1

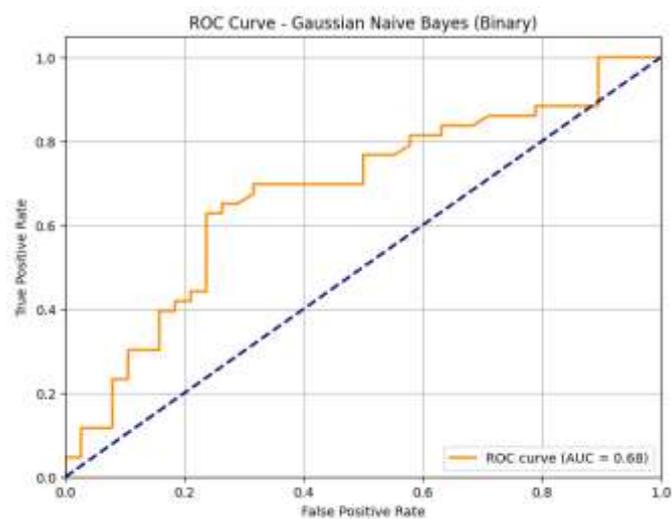
Akurasi	Precision	Recall	F1-Score	AUC
0.69	0.69	0.69	0.69	0.68

Hasil pada skenario ini menunjukkan bahwa model Gaussian Naive Bayes (GNB) mampu memberikan performa yang cukup stabil dan konsisten. Hal ini terlihat dari nilai akurasi, precision, recall, dan F1-score yang identik, yaitu masing-masing 0,69. Nilai-nilai ini mengindikasikan bahwa model tidak memiliki kecenderungan bias dalam memprediksi salah satu kelas (baik kelas Produktivitas Rendah maupun Produktivitas Tinggi). Artinya, model mampu mengidentifikasi kedua kelas dengan kinerja yang relatif seimbang, sehingga tidak terjadi dominasi prediksi pada salah satu kelas.

Selain metrik-metrik tersebut, pada eksperimen ini juga telah dilakukan analisis visual melalui gambar confusion matrix dan kurva ROC (*Receiver Operating Characteristic*) sebagaimana terlihat pada gambar 4.1 dan 4.2.



Gambar 4.1 Confusion Matrix Eksperimen 1



Gambar 4.2 Kurva ROC Eksperimen 1

Nilai AUC sebesar 0,68 mengindikasikan bahwa kemampuan model dalam membedakan antara kelas Tinggi dan Rendah berada pada kategori cukup baik (fair). Meskipun nilai AUC ini belum mencapai kategori baik (biasanya di atas 0.7), namun sudah menunjukkan bahwa model memiliki kemampuan diskriminatif

yang layak untuk kasus ini.

4.4 Hasil Eksperimen II

Gaussian Naive Bayes (70:30 Split, 10-Fold Cross-Validation)

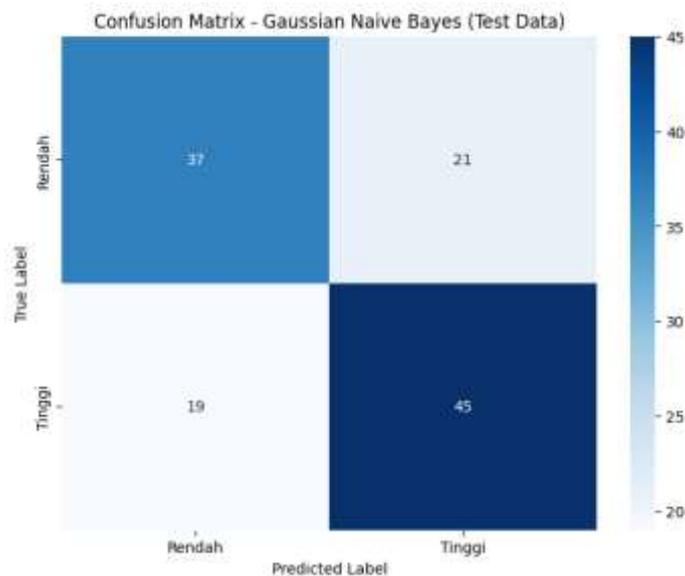
Eksperimen kedua tetap menggunakan model Gaussian Naive Bayes (GNB), namun dengan konfigurasi yang berbeda, yaitu 70% data latih dan 30% data uji. Pada validasi model digunakan strategi 10-fold cross-validation agar model diuji pada lebih banyak kombinasi subset data, sehingga evaluasi performanya menjadi lebih komprehensif dan kuat terhadap variasi data. Hasil rata-rata dari evaluasi performa model GNB pada skenario ini adalah sebagaimana dalam tabel 4.3

Tabel 4.3 Hasil Evaluasi Eksperimen 2

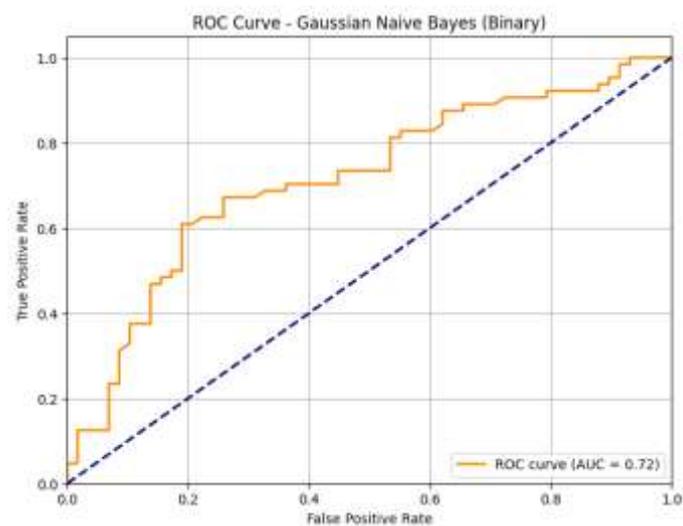
Akurasi	Precision	Recall	F1-Score	AUC
0.66	0.65	0.65	0.65	0.72

Hasil eksperimen kedua menunjukkan adanya sedikit penurunan pada nilai akurasi dan metrik lainnya dibandingkan dengan eksperimen pertama. Nilai akurasi sebesar 0,66 serta precision, recall, dan F1-score masing-masing sebesar 0,65, mengindikasikan bahwa model masih mampu memberikan prediksi yang cukup stabil, meskipun sedikit menurun dibanding konfigurasi sebelumnya (80:20 split). Yang menarik pada skenario ini adalah nilai AUC yang meningkat menjadi 0,72. Nilai ini menunjukkan bahwa kemampuan diskriminatif model dalam membedakan antara kelas *Produktivitas Tinggi* dan *Produktivitas Rendah* mengalami peningkatan yang cukup berarti. Dengan AUC di atas 0,70 model

tersebut masuk dalam kategori baik (good) dalam hal kemampuan memisahkan kedua kelas tersebut, meskipun ketepatan prediksi secara keseluruhan sedikit berkurang. Adapun hasil evaluasi confusion matrix dan kurva ROC terdapat pada gambar 4.3 dan 4.4.



Gambar 4.3 Confusion Matrix Eksperimen 2



Gambar 4.4 Kurva ROC Eksperimen 2

Secara keseluruhan, eksperimen kedua ini memperlihatkan bahwa konfigurasi

70:30 dengan validasi silang 10-fold memberikan peningkatan pada aspek kemampuan diskriminasi antar kelas, meskipun ada sedikit penurunan pada metrik akurasi dan F1-score. Temuan ini menjadi masukan penting untuk pertimbangan dalam pemilihan konfigurasi data latih dan uji pada model GNB di penelitian ini.

4.5 Hasil Eksperimen III

Regresi Logistik (80:20 Split, 5-Fold Cross-Validation)

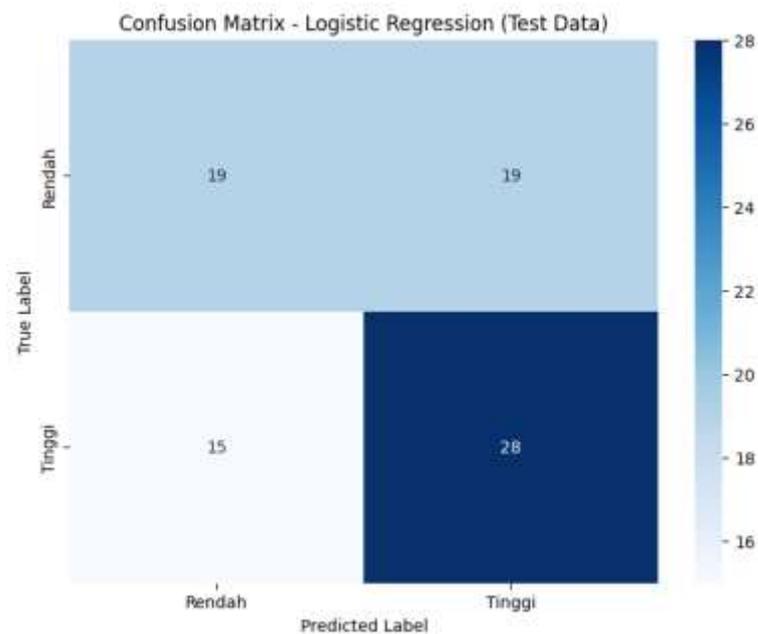
Eksperimen ketiga menggunakan model Regresi Logistik (Logistic Regression / LR) dengan konfigurasi data yang sama seperti pada eksperimen pertama, yaitu 80% data latih dan 20% data uji. Validasi model dilakukan menggunakan 5-fold cross-validation untuk menguji konsistensi performa model terhadap data yang berbeda-beda pada setiap fold. Hasil evaluasi terdapat pada tabel 4.4

Tabel 4.4 Hasil Evaluasi Eksperimen 3

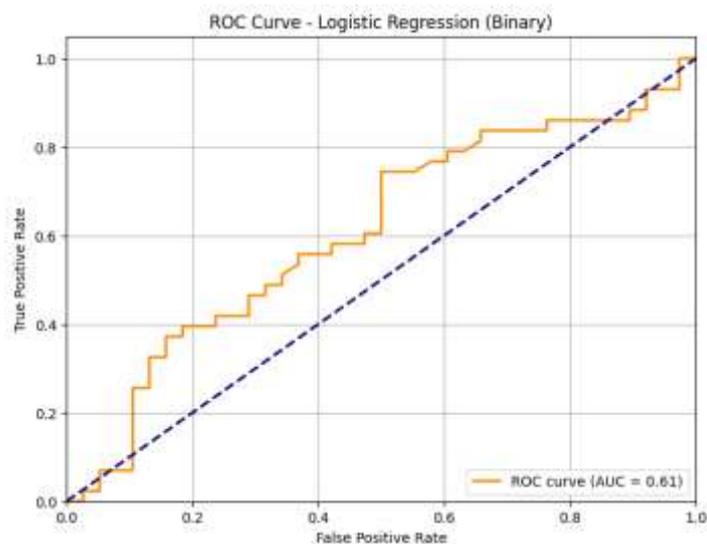
Akurasi	Precision	Recall	F1-Score	AUC
0.58	0.58	0.58	0.57	0.61

Hasil eksperimen ini menunjukkan bahwa model Regresi Logistik pada konfigurasi 80:20 split dan validasi 5-fold memberikan performa yang paling rendah dibandingkan eksperimen-eksperimen sebelumnya (dengan model GNB). Nilai akurasi sebesar 0,58 serta precision dan recall masing-masing sebesar 0,58 menunjukkan bahwa kemampuan model dalam memprediksi kelas *Produktivitas Tinggi* maupun *Produktivitas Rendah* hanya sedikit lebih baik daripada tebakan acak (random guess), yang biasanya berada pada kisaran 0,50. Nilai F1-score

sebesar 0,57 yang sedikit lebih rendah dari precision dan recall mengindikasikan bahwa keseimbangan antara presisi dan sensitivitas model kurang optimal. Hasil evaluasi confusion matrix dan kurva ROC eksperimen 3 terdapat pada gambar 4.5 dan 4.6



Gambar 4.5 Confusion Matrix Eksperimen 3



Gambar 4.6 Kurva ROC Eksperimen 3

Selain itu, nilai AUC sebesar 0,61 meskipun lebih baik dari 0,5 tetap menunjukkan bahwa kemampuan model dalam membedakan kedua kelas masih tergolong lemah dan jauh dari kategori baik. Hal ini menunjukkan bahwa model mengalami kesulitan dalam memisahkan kedua kelas dengan threshold yang optimal. Secara keseluruhan, eksperimen ketiga ini memperlihatkan bahwa model Regresi Logistik dengan konfigurasi ini kurang efektif untuk data dan target prediksi dalam penelitian ini. Hal ini menjadi masukan penting untuk mempertimbangkan penggunaan model lain atau perbaikan lebih lanjut pada tahapan preprocessing dan pemilihan fitur.

4.6 Hasil Eksperimen IV

Regresi Logistik (70:30 Split, 10-Fold Cross-Validation)

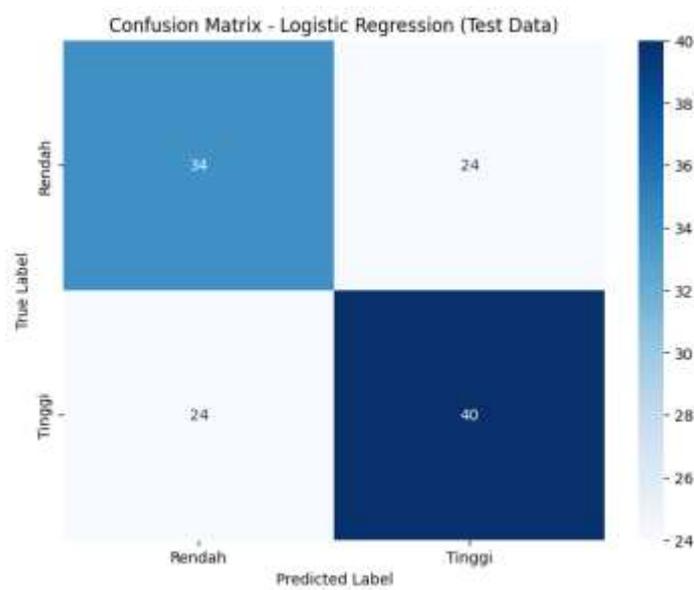
Eksperimen keempat menggunakan model Regresi Logistik dengan konfigurasi 70% data latih dan 30% data uji. Validasi model dilakukan menggunakan 10-fold cross-validation. Strategi ini dipilih untuk memberikan evaluasi performa yang lebih menyeluruh, karena model diuji pada lebih banyak kombinasi subset data, sehingga hasil evaluasi lebih robust dan tidak bias terhadap distribusi data tertentu.

Tabel 4.5 Hasil Evaluasi Eksperimen 4

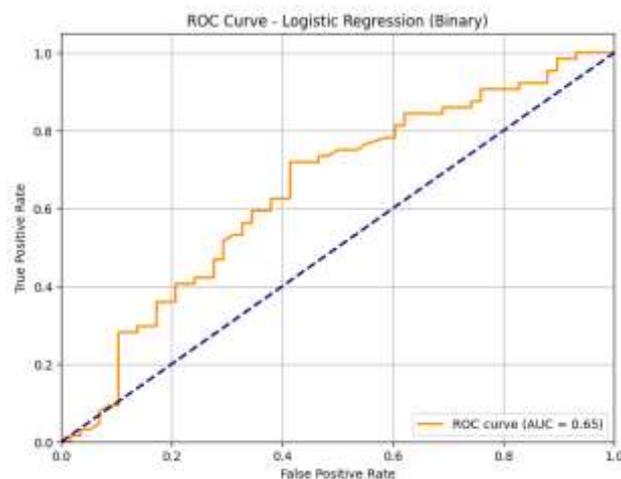
Akurasi	Precision	Recall	F1-Score	AUC
0.61	0.61	0.61	0.61	0.65

Hasil eksperimen ini memperlihatkan adanya peningkatan performa model Regresi Logistik dibandingkan dengan eksperimen sebelumnya (rasio 80:20, 5-

fold CV). Nilai akurasi sebesar 0,61, precision, recall, dan F1-score masing-masing sebesar 0.61 menunjukkan bahwa model memperoleh kemampuan prediksi yang sedikit lebih baik daripada tebakan acak, dan konsistensi metrik ini mengindikasikan bahwa model cukup seimbang dalam memprediksi kedua kelas (*Produktivitas Tinggi* dan *Produktivitas Rendah*). Hasil evaluasi confusion matrix dan kurva ROC terdapat pada gambar 4.7 dan 4.8.



Gambar 4.7 Confusion Matrix Eksperimen 4



Gambar 4.8 Kurva ROC Eksperimen 4

Yang patut dicatat adalah nilai AUC sebesar 0,65 yang menunjukkan adanya peningkatan kemampuan diskriminatif model dibandingkan dengan skenario sebelumnya. AUC ini sudah mulai mendekati kategori cukup baik (fair), meskipun masih di bawah kategori baik (good). Hal ini menunjukkan bahwa penggunaan 10-fold cross-validation membantu model Regresi Logistik menemukan pola yang lebih representatif dalam data, sehingga meningkatkan performanya dalam membedakan kedua kelas. Secara keseluruhan, eksperimen keempat ini memberikan indikasi bahwa strategi validasi silang yang lebih robust (10-fold CV) dapat memberikan manfaat nyata pada model Regresi Logistik, terutama dalam meningkatkan kemampuan diskriminatifnya terhadap kedua kelas target.

4.7 Pembahasan Komparasi Eksperimen

Berdasarkan keempat eksperimen yang telah dilakukan, dapat disusun analisis komparatif terkait performa kedua model yang digunakan, yaitu Gaussian Naïve Bayes (GNB) dan Regresi Logistik (LR), pada dua konfigurasi data dan teknik validasi yang berbeda. Secara umum, hasil yang diperoleh memperlihatkan variasi performa baik pada tingkat akurasi, precision, recall, F1-score, maupun AUC, yang memberikan gambaran menyeluruh tentang kelebihan dan keterbatasan masing-masing model dalam konteks klasifikasi produktivitas padi di Jawa Timur. Pada eksperimen pertama, model GNB dengan rasio 80:20 dan validasi silang 5-fold memberikan hasil yang cukup stabil dengan akurasi, precision, recall, dan F1-score masing-masing sebesar 0,69 serta AUC sebesar **0,68**. Hasil ini mengindikasikan bahwa model memiliki kemampuan prediksi yang seimbang

antara kedua kelas dan mampu membedakan kelas produktivitas tinggi dan rendah pada tingkat yang tergolong cukup baik. Keunggulan pada eksperimen ini terletak pada konsistensi metrik yang identik dan kemampuan diskriminatif model yang cukup solid, meskipun belum masuk kategori baik dalam hal AUC.

Pada **eksperimen kedua**, GNB diuji dengan rasio 70:30 dan validasi silang 10-fold. Terdapat sedikit penurunan pada akurasi dan metrik-metrik lainnya menjadi sekitar **0,65 – 0,66**, namun justru nilai AUC meningkat menjadi 0,72. Peningkatan AUC ini menjadi catatan penting karena menunjukkan bahwa model GNB lebih mampu memisahkan kedua kelas target pada skenario ini dibandingkan pada eksperimen pertama. Hal ini dapat diasumsikan sebagai efek positif dari penggunaan validasi silang yang lebih ketat (10-fold) dan proporsi data uji yang lebih besar, yang memungkinkan model mengeksplorasi pola pada data dengan lebih baik, meskipun hal ini diiringi dengan sedikit penurunan ketepatan prediksi secara keseluruhan.

Berbeda dengan GNB, model Regresi Logistik pada eksperimen ketiga (rasio 80:20, 5-fold CV) memperlihatkan performa yang paling rendah. Akurasi hanya mencapai 0,58, precision dan recall sebesar 0,58, F1-score 0,57, dan AUC 0,61. Nilai-nilai ini mengindikasikan bahwa model LR dalam skenario ini mengalami kesulitan yang cukup besar dalam memisahkan kedua kelas produktivitas. Performa yang mendekati tingkat acak menunjukkan bahwa model cenderung gagal menangkap pola data yang cukup kuat untuk prediksi yang akurat dan diskriminatif pada konfigurasi ini.

Namun, eksperimen keempat memperlihatkan adanya perbaikan kinerja Regresi

Logistik. Dengan konfigurasi 70:30 split dan 10-fold CV, model memperoleh akurasi 0,61, precision, recall, dan F1-score 0,61, serta AUC 0,65. Meskipun masih berada di bawah performa GNB, hasil ini menunjukkan bahwa model Regresi Logistik dapat mengambil manfaat dari strategi validasi silang yang lebih ketat dan proporsi data uji yang lebih besar. AUC yang meningkat menjadi 0,65 menandakan adanya peningkatan kemampuan diskriminatif, meskipun belum mampu mencapai kategori baik dalam memisahkan kedua kelas.

Secara keseluruhan, dari keempat eksperimen ini dapat disimpulkan bahwa model Gaussian Naïve Bayes konsisten memberikan performa yang lebih baik dibandingkan model Regresi Logistik untuk kasus klasifikasi produktivitas padi dengan fitur-fitur numerik yang digunakan.

Model GNB tidak hanya unggul dalam metrik akurasi, precision, recall, dan F1-score, tetapi juga memperlihatkan kemampuan diskriminatif yang lebih baik, terutama pada skenario validasi silang 10-fold dengan rasio 70:30, di mana AUC mencapai angka 0.72. Hal ini menunjukkan bahwa GNB lebih adaptif terhadap pola distribusi data pada dataset yang digunakan. Sementara itu, model Regresi Logistik tampaknya memerlukan pendekatan tambahan, seperti optimasi parameter, pemilihan fitur yang lebih selektif, atau penambahan fitur baru, agar dapat meningkatkan kinerjanya mendekati atau bahkan melampaui performa GNB.

Hasil ini juga menegaskan pentingnya memilih konfigurasi data dan strategi validasi yang tepat dalam pemodelan *machine learning*. Penggunaan validasi silang yang lebih ketat (10-fold) terbukti memberikan manfaat pada kedua model,

meskipun tingkat peningkatannya berbeda-beda. Selain itu, temuan ini menjadi masukan berharga untuk penelitian lanjutan, bahwa eksplorasi terhadap model lain atau kombinasi model (ensemble learning) dapat dipertimbangkan untuk memperoleh performa klasifikasi yang lebih optimal.

4.8 Integrasi Penelitian dengan Islam

Penelitian ini tidak hanya berfokus pada pengembangan ilmu pengetahuan di bidang data mining untuk klasifikasi produktivitas padi, tetapi juga mengandung nilai-nilai integrasi dengan ajaran Islam. Islam memerintahkan umatnya untuk menggunakan akal pikiran dan ilmu pengetahuan dalam memahami fenomena alam sebagai bentuk tadabbur terhadap tanda-tanda kekuasaan Allah SWT.

Hal ini sesuai dengan firman Allah dalam **Q.S. Al-Baqarah ayat 164**:

"Sesungguhnya dalam penciptaan langit dan bumi, silih bergantinya malam dan siang, bahtera yang berlayar di laut membawa apa yang berguna bagi manusia, dan apa yang Allah turunkan dari langit berupa air lalu dengan air itu Dia hiduapkan bumi sesudah mati (kering)-nya dan Dia sebarkan di bumi itu segala jenis hewan, dan berkisarnya angin dan awan yang dikendalikan antara langit dan bumi. Sungguh (terdapat) tanda-tanda (keesaan dan kebesaran Allah) bagi kamu yang memikirkannya."

Ayat ini menegaskan bahwa segala fenomena alam, termasuk dinamika lingkungan dan cuaca yang memengaruhi produktivitas padi, merupakan bagian dari tanda-tanda kebesaran Allah SWT. Melalui penelitian ini, usaha untuk memanfaatkan data cuaca dan faktor lingkungan guna memprediksi produktivitas

padi merupakan bentuk implementasi perintah Allah agar manusia berpikir, meneliti, dan memanfaatkan ilmu pengetahuan untuk kemaslahatan bersama.

Metode ilmiah seperti data mining, khususnya algoritma **Naïve Bayes** dan **Regresi Logistik**, merupakan bentuk ikhtiar untuk memahami pola yang diciptakan Allah SWT dalam sistem alam. Dalam perspektif Islam, hal ini termasuk dalam menjalankan amanah sebagai khalifah di bumi untuk mengelola sumber daya alam dengan ilmu, kebijaksanaan, dan tanggung jawab.

Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan ilmu pengetahuan modern, tetapi juga sebagai bentuk pengabdian dan ketaatan kepada Allah SWT dalam mengelola bumi agar tetap memberikan manfaat sebesar-besarnya bagi umat manusia. Hasil klasifikasi produktivitas padi diharapkan dapat membantu pengambilan kebijakan strategis dalam sektor pertanian, sehingga tercapai keberlanjutan produksi pangan yang sejalan dengan prinsip Islam dalam menjaga keseimbangan dan keberlanjutan alam.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan pada klasifikasi produktivitas padi di Jawa Timur menggunakan dua metode, yaitu Gaussian Naïve Bayes (GNB) dan Regresi Logistik (LR), dengan berbagai konfigurasi pembagian data dan strategi validasi silang, diperoleh beberapa kesimpulan sebagai berikut:

1. Model Gaussian Naïve Bayes secara konsisten memberikan performa lebih baik dibandingkan model Regresi Logistik dalam mengklasifikasikan produktivitas padi menjadi kelas *Tinggi* dan *Rendah*. Hal ini ditunjukkan dengan metrik akurasi, precision, recall, F1-score, dan AUC yang lebih tinggi pada kedua konfigurasi eksperimen GNB.
2. Eksperimen GNB dengan rasio 70:30 dan validasi silang 10-fold merupakan skenario terbaik pada penelitian ini, dengan capaian akurasi sebesar 0,66, precision dan recall 0,65, F1-score 0,65, dan AUC 0,72. Nilai AUC yang mendekati kategori baik menunjukkan kemampuan model dalam membedakan kedua kelas secara cukup efektif.
3. Regresi Logistik menunjukkan performa yang lebih rendah, terutama pada konfigurasi rasio 80:20 dan 5-fold CV, dengan akurasi 0,58 dan AUC 0,61. Meskipun terjadi peningkatan pada konfigurasi 70:30 dengan 10-fold CV (AUC naik menjadi 0,65), performa model ini masih berada di bawah model GNB.

4. Strategi validasi silang 10-fold memberikan kontribusi positif terhadap performa kedua model, terutama pada kemampuan diskriminatif model yang tercermin dari peningkatan nilai AUC. Hal ini menunjukkan pentingnya pemilihan teknik validasi yang tepat dalam pemodelan machine learning.
5. Distribusi data dan proporsi data latih-uji berpengaruh terhadap performa model. Pada penelitian ini, konfigurasi 70:30 dengan validasi silang lebih robust memberikan hasil yang lebih baik pada kedua metode, dibandingkan konfigurasi 80:20

5.2 Saran

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, terdapat beberapa saran untuk penelitian selanjutnya maupun pengembangan sistem klasifikasi produktivitas padi:

1. Perlu dilakukan eksplorasi model machine learning lain seperti Random Forest, Support Vector Machine (SVM), atau Gradient Boosting untuk memperoleh performa klasifikasi yang lebih optimal, terutama dalam meningkatkan nilai AUC dan akurasi secara bersamaan.
2. Perlu dilakukan optimasi hyperparameter pada kedua model (GNB dan LR) yang digunakan. Pada penelitian ini model dijalankan dengan konfigurasi default, sehingga ada peluang peningkatan performa melalui tuning parameter seperti `var_smoothing` pada GNB atau `C` pada LR.
3. Pengayaan fitur dan pemilihan fitur yang lebih relevan diharapkan dapat membantu model dalam menangkap pola yang lebih baik. Penambahan fitur

seperti kualitas tanah, varietas padi, atau penggunaan pupuk dapat menjadi variabel penting yang meningkatkan akurasi dan kemampuan diskriminatif model.

DAFTAR PUSTAKA

- Akmal, K., Faqih, A., & Dikananda, F. (2023). *Perbandingan Metode Algoritma Naïve Bayes dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke*. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1), 467-474.
- Angdresey, A., Sitanayah, L., & Tangka, I. L. H. (2025). *Sentiment Analysis for Political Debates on YouTube Comments using BERT Labeling, Random Oversampling, and Multinomial Naïve Bayes*. JCTA Journal of Computing Theories and Applications, 2(3), 342-354.
- Arifuddin, N. A., Pembangunan, U., Veteran, N., & Machine, S. V. (2023). *Komparasi Naïve Bayes dan Support Vector Machine dalam Klasifikasi Jenis Citrus*. Jurnal Ilmiah Teknik Informatika, 22(2), 409–417.
- Arisandi, R. R. R., Warsito, B., & Hakim, A. R. (2022). *Aplikasi Naïve Bayes Classifier (NBC) pada Klasifikasi Status Gizi Balita Stunting dengan Pengujian K-Fold Cross Validation*. Jurnal GAUSSIAN, 11(1), 130-139.
- Aziz, A. A., & Prasetyo, B. (2025). *Classification of Student Grading Using Naïve Bayes Method with Under-sampling Approach to Handle Imbalance*. Journal of Information System Exploration and Research, 3(1), 39-44.
- Bangun, A. V., & Melanie, R. (2024). *Pemodelan Regresi Logistik Biner tentang Perilaku Pengambilan Keputusan Polifarmasi pada Lansia*. Jurnal Penelitian Kesehatan Suara Forikes, 15(Nomor Khusus Januari), 122–132.

- Burhanuddin, A. (2023). *Analisis Komparatif Inferensi Fuzzy Tsukamoto, Mamdani dan Sugeno Terhadap Produktivitas Padi di Indonesia*. LEDGER: Journal Informatic and Information Technology, 2(1), 48-57.
- Cahyono, B. E., Rahagian, R., & Nugroho, A. T. (2023). *Analisis Produktivitas Padi Berdasarkan Indeks Kekeringan (NDWI dan NDDI) Lahan Sawah Menggunakan Data Citra Sentinel-2A di Kecamatan Ambulu*. Indonesian Journal of Applied Physics (IJAP), 13(1), 88-98.
- Chaniago, N. (2023). *Pengaruh Curah Hujan Terhadap Produksi dan Produktivitas Padi di Kecamatan Percut Sei Tuan Kabupaten Deli Serdang Sumatera Utara*. AGRILAND Jurnal Ilmu Pertanian, 11(3), 130-136.
- Delvika, B., Nurhidayarnis, S., Rinada, P. D., Abror, N., & Hidayat, A. (2022). *Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women*. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 2(2), 68-75.
- Desiani, A., Amran, A., Andriani, Y., Wahyunid, T., & Rizki, F. (2025). *Perbandingan Algoritma Logistic Regression dan Adaptive Boosting (Adaboost) dalam Klasifikasi Penyakit Gagal Jantung*. Jurnal Teknologi Informasi, 19(1), 71-80.
- Fauzan, B. L., Agustin, T., & Mahmudah, A. M. H. (2024). *Prediksi Klasifikasi Kecelakaan Lalu Lintas di Kota Surakarta dengan Menggunakan Metode Regresi Logistik Multinomial*. Sustainable Civil Building Management and Engineering Journal, 1(4), 1-9.

- Febrinita, F., Zaman, W. I., & Puspitasari, W. D. (2024). *Faktor-Faktor yang Mempengaruhi Hasil Belajar Statistika Mahasiswa melalui Pemodelan Regresi Logistik Biner*. Kognitif: Jurnal Riset HOTS Pendidikan Matematika, 4(1), 523-534.
- Gunawan, N. O., & Nasib, S. K. (2025). *Analisis Regresi Logistik Multinomial untuk Menentukan Faktor-Faktor yang Mempengaruhi Jenis Penyakit pada Mahasiswa (Studi Kasus: Mahasiswa Program Studi Statistika, Jurusan Matematika, Universitas Negeri Gorontalo)*. Research Review Jurnal Ilmiah Multidisiplin, 4(1), 72-80.
- Handayani, Y., Hidayat, T., Novitaningrum, D., & Ismail, A. R. (2025). *Perbandingan Algoritma Logistic Regression dan Naïve Bayes Classifier dalam Identifikasi Penyakit Liver*. Journal of Science and Social Research, 8(2), 1435-1440.
- Hasibuan, N. K., Dur, S., & Husein, I. (2022). *Faktor Penyebab Penyakit Diabetes Melitus dengan Metode Regresi Logistik*. G-Tech: Jurnal Teknologi Terapan, 6(2), 257-264.
- Hidayati, L., Agustini, D., Ripai, & Awaludin. (2022). *Pemodelan Produktivitas Padi Menggunakan Regresi Semiparametrik Spline Truncated*. Vol.2 No.3 Agustus 2022, 913-916.
- Keliat, M. R. B., & Ikhsan, M. (2025). *Komparasi Algoritma Support Vector Machine dan Naïve Bayes pada Klasifikasi Jenis Buah Kurma berdasarkan Citra Hue Saturation Value*. Sistemasi: Jurnal Sistem Informasi, 14(1), 470-481.

- Mardiani, E., Rahmansyah, N., Kurniati, I., Matondang, N., Tesalonika, Amelia Zanitha, D., & Romzy, I. (2023). *Membandingkan Algoritma Data Mining Dengan Tools Orange untuk Social Economy*. *Digital Transformation Technology (Digitech)*, 3(2), 686–693.
- Martadiansyah, M. W., Ghufron, A., Hidayah, R. A., Salzabila, D., & Amandas, L. (2025). *Perbandingan Algoritma C4.5 dengan Naive Bayes Untuk Klasifikasi Curah Hujan*. *Jurnal Komputer Antartika*, 3(1), 8-17. <https://doi.org/10.70052/jka.v3i1.648>
- Maulana, A., Ilham, M., Lonang, S., Insyroh, N., Costa, A. D. S. da, Talirongan, F. J. B., Furizal, & Firdaus, A. A. (2025). *Classification of Stunting in Toddlers using Naive Bayes Method and Decision Tree*. *Indonesian Journal of Modern Science and Technology (IJMST)*, 1(1), 28-33.
- Muharrom, M. (2023). *Analisis Penggunaan Orange Data Mining untuk Prediksi Harga USDT/BIDR Binance*. *Bulletin of Information Technology (BIT)*, 4(2), 178–184.
- Mukharyahya, Z. A., Astuti, Y. P., & Cahyani, O. N. (2025). *Perbandingan Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia*. *EDUMATIC: Jurnal Pendidikan Informatika*, 9(1), 119-128. <https://doi.org/10.29408/edumatic.v9i1.29512>
- Muliati, Juliansyah, H., & Rozalina. (2022). *Pengaruh Produksi Dan Produktivitas Padi Terhadap Produk Domestik Bruto Indonesia*. *Jurnal Penelitian Agrisamudra*, 9(2), 90-99.

- Musababa, M. A. (2024). *Impelementasi Algoritma Linear Regression untuk Prediksi Produksi Tanaman Padi di Kabupaten Grobogan*. DSI: Jurnal Data Science Indonesia, 3(1), 1–10.
- Muthmainnah, Rudiman, & Yulianto, F. (2025). *Sentiment Analysis of the Public on the Deployment of Smart Robots in Indonesia Using the Naïve Bayes Method*. JSE Journal of Science and Engineering, 3(2), 85-89.
- Nasien, D., Darwin, R., Cia, A., Winata, A. L., Go, J., M.C, R., Wijaya, R. C., & Los, K. C. (2024). *Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, Dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes*. JURNAL TEKNIK INFORMATIKA, 4(1), 10-17.
- Ningsih, S., Madonsa, M. R., Mahmud, S. L., Djakaria, I., & Nasib, S. K. (2024). *Implementasi Regresi Logistik Biner Stratifikasi Pada Pemodelan Stunting Untuk Anak Balita Di Kabupaten Gorontalo*. Jambura Journal of Probability and Statistics, 5(1), 19-23.
- Pratiwi, I., Sinaga, A., & Asura, M. (2024). *Faktor-Faktor yang Mempengaruhi Kualitas Tidur: Analisis Melalui Regresi Logistik Biner*. PARAMETER JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA, 3(2), 115-126.
- Priambodo, S. A., & Falani, A. Z. (2020). *Pemanfaatan Data Mining untuk Klasterisasi Potensi Produksi Beras di Kabupaten Blitar dengan Menggunakan Metode Fuzzy C-Means*. Jurnal SPIRIT, 12(2), 30–36.

- Purba, T., & Pane, R. (2024). *Analisis Regresi Logistik Biner Pada Penyakit Polycystic Ovary Syndrome (PCOS)*. *Jurnal Pendidikan Inklusif*, 8(12), 152-169.
- Putri, N. B., & Wijayanto, A. W. (2022). *Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing*. *Komputika: Jurnal Sistem Komputer*, 11(1), 59-66.
- Reska, & Khikmah, L. (2025). *Identification of Variables Affecting Employee Performance Using Ordinal Logistic Regression*. *Jurnal Sintak*, 3(2), 147-154. <https://doi.org/10.62375/isintak.v3i2.552>
- Riany, A. F., & Testiana, G. (2023). *Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes*. *SAINTEKOM*, 13(1), 42-54.
- Rokhmah, S., Susilowati, A., & Intan P, M. (2022). *Klasifikasi Data untuk Prediksi Hasil Panen Tanaman Padi di Wilayah Kabupaten Sukoharjo Menggunakan Algoritma C4.5*. *JURTI*, 6(2), 134–140.
- Salsabyla, A. A., & Wulandari, S. P. (2023). *Permodelan Regresi Logistik Biner terhadap Analisis Penderita Penyakit Jantung Koroner di RSUD Dr SOEGIRI Lamongan*. *JURNAL SAINS DAN SENI ITS*, 12(1), 103-110.
- Sani, R. R., Pratiwi, Y. A., Winarno, S., Udayanti, E. D., & Zami, F. A. (2022). *Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Hoax pada Berita Online Indonesia*. *Jurnal Masyarakat Informatika*, 13(2), 85-94.

- Saputra, M. A. W., & Harini, S. (2022). *Java Island Health Profile Clustering using K-Means Data Mining*. Intl. Journal on ICT, 8(1), 1–9.
- Sari, C. A., Sukmawati, A., Aprilli, R. P., Kayaningtias, P. S., & Yudistira, N. (2022). *Perbandingan Metode Naïve Bayes, Support Vector Machine dan Decision Tree Dalam Klasifikasi Konsumsi Obat*. Jurnal Litbang Edusaintech, 3(1), 33-41.
- Sari, F. R., Fitri, F., Putra, A. A., & Permana, D. (2023). *Comparison of Naive Bayes Method and Binary Logistics Regression on Classification of Social Assistance Recipients Program Keluarga Harapan (PKH)*. UNP JOURNAL OF STATISTICS AND DATA SCIENCE, 1(2), 82-89.
- Satria, A., Badri, R. M., & Safitri, I. (2023). *Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning*. Digital Transformation Technology (Digitech), 3(2), 390–398.
- Supriyanto, J., Korespondensi, P., Alita, D., & Isnain, A. R. (2023). *Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring*. Jurnal Informatika dan Rekayasa Perangkat Lunak, 4(1), 74–80.
- Tripenna, A., Maharsi, R., Lianawati, Y., & Setyawan, A. A. (2023). *Analisis Faktor-Faktor Yang Mempengaruhi Kemiskinan Rumah Tangga Di Desa Kotayasa Melalui Pendekatan Regresi Logistik Biner*. JELC: Jurnal Elektro Luceat, 9(2), 42-47.
- Wahyudin, E., Rudin, R. A., Kaslani, & Permana, S. E. (2024). *Penerapan Data Mining Pengelompokan Produktivitas Padi Menggunakan Algoritma K-*

- Means Pada Provinsi Jawa Barat. JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 523-528.
- Widyastuti, E., Hermawan, A., & Avianto, D. (2025). *Klasifikasi Tomat Berdasarkan Varietas Dengan Ekstraksi Fitur RGB Dan Algoritma Naïve Bayes*. *Indonesia Journal Information System (Idealis)*, 8(1), 127-137.
- Wijaya, A. P. (2025). *Perbandingan Algoritma Klasifikasi Random Foresst dengan Naïve Bayes Classifier pada Studi Penyakit Berdasarkan Pola Nutrisi*. *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, 9(1), 65-69.
- Wijaya, F. P., Mawuntu, A. H. P., Muharomah, S., Tumboimbela, M. J., & Langi, F. L. F. G. (2025). *Application of Predictive Models for Tuberculous Meningitis Outcome: A Comparative Analysis of Decision Tree and Logistic Regression Approaches*. *Cermin Dunia Kedokteran*, 52(4), 221-228.

Lampiran 1. Dataset Penelitian

Kabupaten/Kota	Bulan	Suhu	Kelembaban_Udara	Kecepatan_Angin	Tekanan_Udara	Curah_Hujan	Hari_Hujan	Penyinaran	Luas_Panen	Produksi_Padi	Produktivitas_Padi
Kabupaten Pacitan	Januari	27	87	4	930	46	20	3	21	129	6,20
Kabupaten Pacitan	Februari	27	86	2	929	509	21	3	3227	13351	4,14
Kabupaten Pacitan	Maret	27	83	2	929	453	24	4	5315	28860	5,43
Kabupaten Pacitan	April	27	78	4	930	253	17	5	1906	11660	6,12
Kabupaten Pacitan	Mei	27	79	3	930	313	16	3	1016	5066	4,98
Kabupaten Pacitan	Juni	26	80	2	931	245	16	4	1757	8810	5,01
Kabupaten Pacitan	Juli	25	77	4	931	341	4	5	2300	11470	4,99
Kabupaten Pacitan	Agustus	25	76	6	932	150	10	4	563	2807	4,98
Kabupaten Pacitan	September	26	84	7	930	361	13	2	449	2134	4,75
Kabupaten Pacitan	Oktober	27	83	6	930	99	25	2	371	1779	4,79
Kabupaten Pacitan	November	27	83	3	930	82	27	2	826	3984	4,83
Kabupaten Pacitan	Desember	27	84	2	929	589	22	2	80	378	4,74
Kabupaten Ponorogo	Januari	26	88	4	930	364	25	1	1346	7663	5,69
Kabupaten Ponorogo	Februari	26	88	2	929	403	24	1	119	679	5,69
Kabupaten Ponorogo	Maret	27	84	2	929	192	21	2	13146	74835	5,69
Kabupaten Ponorogo	April	27	79	4	930	176	8	2	12980	73881	5,69
Kabupaten Ponorogo	Mei	27	80	3	930	120	6	3	2401	13115	5,46
Kabupaten Ponorogo	Juni	26	81	2	931	167	14	4	2621	14318	5,46
Kabupaten Ponorogo	Juli	26	78	4	931	4	1	4	13697	74823	5,46
Kabupaten Ponorogo	Agustus	25	77	6	932	16	1	4	6645	36299	5,46
Kabupaten Ponorogo	September	26	87	7	930	30	4	4	1443	9197	6,37
Kabupaten Ponorogo	Oktober	27	85	6	930	32	5	3	65	412	6,37
Kabupaten Ponorogo	November	27	85	3	930	374	20	2	1619	10320	6,37
Kabupaten Ponorogo	Desember	27	86	2	929	156	17	1	6883	43872	6,37
Kabupaten Trenggalek	Januari	26	88	4	930	274	16	1	508	2863	5,63
Kabupaten Trenggalek	Februari	26	87	2	929	332	17	1	129	726	5,63
Kabupaten Trenggalek	Maret	27	84	2	929	215	18	2	5101	28038	5,50
Kabupaten Trenggalek	April	27	78	4	930	135	11	2	3306	18316	5,54
Kabupaten Trenggalek	Mei	27	79	3	930	193	13	3	1017	5132	5,05
Kabupaten Trenggalek	Juni	26	80	2	931	316	15	4	545	2751	5,05

Kabupaten Trenggalek	Juli	25	77	4	931	60	7	4	4677	23600	5,05
Kabupaten Trenggalek	Agustus	24	76	6	932	141	10	4	2276	11482	5,05
Kabupaten Trenggalek	September	25	84	7	930	220	11	4	680	3457	5,09
Kabupaten Trenggalek	Oktober	26	84	6	930	757	23	3	859	4390	5,11
Kabupaten Trenggalek	November	26	84	3	930	576	23	3	1676	8629	5,15
Kabupaten Trenggalek	Desember	26	85	2	929	350	18	2	1253	6374	5,09
Kabupaten Tulungagung	Januari	26	0	4	930	217	17	0	2133	11799	5,53
Kabupaten Tulungagung	Februari	27	0	2	929	335	19	0	828	4579	5,53
Kabupaten Tulungagung	Maret	27	0	2	929	275	20	0	4178	23033	5,51
Kabupaten Tulungagung	April	27	0	4	930	91	10	0	10366	57338	5,53
Kabupaten Tulungagung	Mei	27	0	3	930	66	10	0	3032	15383	5,07
Kabupaten Tulungagung	Juni	26	0	2	931	33	5	0	1686	8553	5,07
Kabupaten Tulungagung	Juli	25	0	4	931	5	1	0	2841	14415	5,07
Kabupaten Tulungagung	Agustus	24	0	6	932	7	3	0	7474	37918	5,07
Kabupaten Tulungagung	September	25	0	7	930	25	6	0	2219	10850	4,89
Kabupaten Tulungagung	Oktober	26	0	6	930	441	18	0	2303	11263	4,89
Kabupaten Tulungagung	November	27	0	3	930	200	19	0	1565	7655	4,89
Kabupaten Tulungagung	Desember	27	0	2	929	302	15	0	906	4431	4,89
Kabupaten Blitar	Januari	28	76	3	977	182	22	5	897	5895	6,57
Kabupaten Blitar	Februari	28	78	3	977	328	25	4	907	6031	6,65
Kabupaten Blitar	Maret	28	78	3	976	240	25	5	4589	28150	6,13
Kabupaten Blitar	April	28	78	3	977	266	17	5	10836	76765	7,08
Kabupaten Blitar	Mei	28	76	2	977	253	14	6	6094	32865	5,39
Kabupaten Blitar	Juni	27	75	2	978	140	12	6	1440	7763	5,39
Kabupaten Blitar	Juli	27	72	3	978	11	4	6	1712	9166	5,35
Kabupaten Blitar	Agustus	27	72	3	978	66	9	6	1903	10244	5,38
Kabupaten Blitar	September	27	74	3	979	165	8	6	2208	14800	6,70
Kabupaten Blitar	Oktober	27	81	2	978	578	27	4	1995	13376	6,70
Kabupaten Blitar	November	27	82	2	977	435	27	3	1037	6950	6,70
Kabupaten Blitar	Desember	27	78	2	976	233	22	4	519	3478	6,70
Kabupaten Kediri	Januari	25	84	3	934	498	25	4	630	3983	6,33
Kabupaten Kediri	Februari	25	86	3	933	491	21	3	1019	6444	6,33

Kabupaten Kediri	Maret	25	86	3	933	343	27	4	8129	51325	6,31
Kabupaten Kediri	April	26	83	3	934	400	18	5	6421	40549	6,32
Kabupaten Kediri	Mei	26	83	3	934	310	18	6	923	4447	4,82
Kabupaten Kediri	Juni	25	81	2	934	218	13	7	870	4191	4,82
Kabupaten Kediri	Juli	26	77	3	934	34	3	7	6722	32404	4,82
Kabupaten Kediri	Agustus	26	74	3	935	130	4	8	1769	8528	4,82
Kabupaten Kediri	September	27	70	3	935	47	6	7	824	4829	5,86
Kabupaten Kediri	Oktober	25	83	2	934	399	18	4	155	905	5,86
Kabupaten Kediri	November	25	85	2	934	665	26	3	1070	6269	5,86
Kabupaten Kediri	Desember	25	84	3	933	303	21	4	850	4978	5,86
Kabupaten Malang	Januari	24	81	2	945	216	25	4	4234	26207	6,19
Kabupaten Malang	Februari	24	82	2	945	353	24	4	1677	10378	6,19
Kabupaten Malang	Maret	24	83	3	945	349	26	5	5129	31248	6,09
Kabupaten Malang	April	24	78	3	945	354	22	7	7178	42414	5,91
Kabupaten Malang	Mei	25	77	4	945	75	14	7	5387	30193	5,61
Kabupaten Malang	Juni	23	80	5	946	308	19	7	3756	21053	5,61
Kabupaten Malang	Juli	23	75	3	946	20	4	8	3774	21043	5,58
Kabupaten Malang	Agustus	23	76	4	947	44	9	8	3578	19821	5,54
Kabupaten Malang	September	24	78	9	947	127	12	7	2825	17483	6,19
Kabupaten Malang	Oktober	24	85	3	946	496	24	5	3003	18585	6,19
Kabupaten Malang	November	24	84	2	946	474	27	4	2735	16927	6,19
Kabupaten Malang	Desember	24	79	3	945	282	21	5	2627	16255	6,19
Kabupaten Lumajang	Januari	24	94	3	915	394	17	0	4643	25843	5,57
Kabupaten Lumajang	Februari	24	92	3	914	180	14	0	4335	24131	5,57
Kabupaten Lumajang	Maret	25	91	3	914	423	20	0	5992	33313	5,56
Kabupaten Lumajang	April	26	90	3	915	236	13	0	7109	39021	5,49
Kabupaten Lumajang	Mei	24	90	3	915	185	12	0	4972	26150	5,26
Kabupaten Lumajang	Juni	24	88	3	915	251	15	0	5809	30555	5,26
Kabupaten Lumajang	Juli	24	86	3	915	47	6	0	5029	26449	5,26
Kabupaten Lumajang	Agustus	23	84	3	999	92	7	0	3413	17948	5,26
Kabupaten Lumajang	September	24	80	3	999	116	10	0	2675	15069	5,63
Kabupaten Lumajang	Oktober	24	91	3	999	377	21	0	3925	22117	5,63

Kabupaten Lumajang	November	25	93	3	999	452	24	0	3837	21623	5,64
Kabupaten Lumajang	Desember	24	91	3	999	264	17	0	3302	18610	5,64
Kabupaten Jember	Januari	27	81	3	1003	335	21	4	6812	35404	5,20
Kabupaten Jember	Februari	27	79	3	1003	82	16	5	4287	22279	5,20
Kabupaten Jember	Maret	27	82	3	1002	280	16	6	15570	80843	5,19
Kabupaten Jember	April	28	77	4	1003	73	14	7	31795	165118	5,19
Kabupaten Jember	Mei	28	79	4	1004	41	10	7	9434	47606	5,05
Kabupaten Jember	Juni	27	80	4	1005	149	12	7	5545	27984	5,05
Kabupaten Jember	Juli	26	80	4	1005	16	7	6	9966	50290	5,05
Kabupaten Jember	Agustus	26	78	4	1005	41	9	7	17142	86507	5,05
Kabupaten Jember	September	27	79	4	1006	63	9	8	6964	35458	5,09
Kabupaten Jember	Oktober	26	84	4	1004	293	20	5	4659	23723	5,09
Kabupaten Jember	November	27	83	3	1004	325	16	5	2865	14590	5,09
Kabupaten Jember	Desember	27	79	4	1002	144	16	7	3451	17569	5,09
Kabupaten Banyuwangi	Januari	27	81	3	1003	335	21	4	5334	32500	6,09
Kabupaten Banyuwangi	Februari	27	79	3	1003	82	16	5	2397	14608	6,09
Kabupaten Banyuwangi	Maret	27	82	3	1002	280	16	6	9762	59482	6,09
Kabupaten Banyuwangi	April	28	77	4	1003	73	14	7	13810	84143	6,09
Kabupaten Banyuwangi	Mei	28	79	4	1004	41	10	7	6737	40646	6,03
Kabupaten Banyuwangi	Juni	27	80	4	1004	149	12	7	4200	25340	6,03
Kabupaten Banyuwangi	Juli	26	80	4	1004	16	7	6	4448	26836	6,03
Kabupaten Banyuwangi	Agustus	26	78	4	1005	41	9	7	4170	25156	6,03
Kabupaten Banyuwangi	September	27	79	4	1005	63	9	8	6283	37249	5,93
Kabupaten Banyuwangi	Oktober	26	84	4	1004	293	20	5	10597	62830	5,93
Kabupaten Banyuwangi	November	27	83	3	1004	325	16	5	4959	29402	5,93
Kabupaten Banyuwangi	Desember	27	79	4	1002	144	6	7	4050	24013	5,93
Kabupaten Bondowoso	Januari	27	81	3	998	335	21	4	2918	14394	4,93
Kabupaten Bondowoso	Februari	27	79	3	999	82	16	5	2747	13968	5,08
Kabupaten Bondowoso	Maret	27	82	3	997	280	16	6	7609	37513	4,93
Kabupaten Bondowoso	April	28	77	4	1000	73	14	7	7222	36719	5,08
Kabupaten Bondowoso	Mei	28	79	4	1001	41	10	7	6447	31670	4,91
Kabupaten Bondowoso	Juni	27	80	4	1001	149	12	7	4420	21769	4,93

Kabupaten Bondowoso	Juli	26	80	4	1001	16	7	6	4706	23118	4,91
Kabupaten Bondowoso	Agustus	26	78	4	1002	41	9	7	2749	13507	4,91
Kabupaten Bondowoso	September	27	79	4	1001	63	9	8	2201	10698	4,86
Kabupaten Bondowoso	Oktober	26	84	4	1008	293	20	5	2831	13977	4,94
Kabupaten Bondowoso	November	27	83	3	999	325	16	5	1995	9698	4,86
Kabupaten Bondowoso	Desember	27	79	4	997	144	16	7	2396	11647	4,86
Kabupaten Situbondo	Januari	27	81	3	998	179	25	4	1158	6141	5,30
Kabupaten Situbondo	Februari	27	79	3	999	570	24	5	2193	11665	5,32
Kabupaten Situbondo	Maret	27	82	3	997	265	23	6	6643	35239	5,30
Kabupaten Situbondo	April	28	77	4	1000	430	15	7	4973	26438	5,32
Kabupaten Situbondo	Mei	28	79	4	1001	455	16	7	1503	7540	5,02
Kabupaten Situbondo	Juni	27	80	4	1001	300	14	7	1387	6956	5,02
Kabupaten Situbondo	Juli	26	80	4	1001	187	5	6	3103	15569	5,02
Kabupaten Situbondo	Agustus	26	78	4	1002	73	3	7	2121	10642	5,02
Kabupaten Situbondo	September	27	79	4	1001	465	7	8	609	3519	5,78
Kabupaten Situbondo	Oktober	26	84	4	1008	240	22	5	1139	6597	5,79
Kabupaten Situbondo	November	27	83	3	999	500	21	5	1046	6059	5,79
Kabupaten Situbondo	Desember	27	79	4	997	221	21	7	908	5261	5,79
Kabupaten Probolinggo	Januari	21	94	1	915	324	13	2	1308	7272	5,56
Kabupaten Probolinggo	Februari	21	92	2	914	175	10	2	1245	6922	5,56
Kabupaten Probolinggo	Maret	22	91	2	914	344	15	3	6432	35499	5,52
Kabupaten Probolinggo	April	22	90	2	915	85	5	5	9067	50329	5,55
Kabupaten Probolinggo	Mei	22	90	2	915	169	8	4	4735	25242	5,33
Kabupaten Probolinggo	Juni	22	88	2	915	198	9	6	2467	13152	5,33
Kabupaten Probolinggo	Juli	21	86	2	915	22	1	6	1757	9369	5,33
Kabupaten Probolinggo	Agustus	21	84	2	999	34	2	6	2122	11313	5,33
Kabupaten Probolinggo	September	22	80	2	999	39	3	6	1713	10318	6,02
Kabupaten Probolinggo	Oktober	22	91	1	999	141	8	3	1126	6778	6,02
Kabupaten Probolinggo	November	22	93	1	999	255	13	2	845	5087	6,02
Kabupaten Probolinggo	Desember	22	91	1	999	179	9	2	686	4131	6,02
Kabupaten Pasuruan	Januari	21	94	1	915	235	28	2	1954	10203	5,22
Kabupaten Pasuruan	Februari	21	92	2	914	140	24	2	1901	9892	5,20

Kabupaten Pasuruan	Maret	22	91	2	914	187	29	3	7437	37989	5,11
Kabupaten Pasuruan	April	22	90	2	915	155	22	5	7309	37701	5,16
Kabupaten Pasuruan	Mei	22	90	2	915	124	21	4	5522	29040	5,26
Kabupaten Pasuruan	Juni	22	88	2	915	44	11	6	2177	11450	5,26
Kabupaten Pasuruan	Juli	21	86	2	915	10	8	6	5967	31285	5,24
Kabupaten Pasuruan	Agustus	21	84	2	999	30	5	6	2982	15669	5,25
Kabupaten Pasuruan	September	22	80	2	999	13	4	6	3498	18445	5,27
Kabupaten Pasuruan	Oktober	22	91	1	999	141	24	3	2827	14909	5,27
Kabupaten Pasuruan	November	22	93	1	999	166	24	2	3219	16998	5,28
Kabupaten Pasuruan	Desember	22	91	1	999	141	26	2	2592	13675	5,28
Kabupaten Sidoarjo	Januari	28	97	9	1010	377	25	5	280	1717	6,13
Kabupaten Sidoarjo	Februari	27	98	7	1009	184	25	5	142	869	6,13
Kabupaten Sidoarjo	Maret	28	98	7	1009	393	18	6	1275	7818	6,13
Kabupaten Sidoarjo	April	28	98	7	1009	227	17	6	6040	37044	6,13
Kabupaten Sidoarjo	Mei	28	98	6	1010	461	21	7	6697	38152	5,70
Kabupaten Sidoarjo	Juni	28	97	6	1010	190	18	7	1502	8555	5,70
Kabupaten Sidoarjo	Juli	29	98	8	1010	126	8	9	175	995	5,70
Kabupaten Sidoarjo	Agustus	29	95	9	1011	39	4	9	1346	7669	5,70
Kabupaten Sidoarjo	September	29	95	7	1011	152	3	7	4806	31682	6,59
Kabupaten Sidoarjo	Oktober	28	98	6	1010	260	14	6	4778	31496	6,59
Kabupaten Sidoarjo	November	28	98	6	1010	209	23	3	4066	26803	6,59
Kabupaten Sidoarjo	Desember	28	97	6	1009	205	19	6	264	1740	6,59
Kabupaten Mojokerto	Januari	27	84	5	1008	405	23	5	880	5459	6,21
Kabupaten Mojokerto	Februari	27	84	5	1008	184	23	4	2810	17142	6,10
Kabupaten Mojokerto	Maret	28	85	4	1008	414	19	5	7584	45865	6,05
Kabupaten Mojokerto	April	28	82	3	1008	227	13	6	11315	69623	6,15
Kabupaten Mojokerto	Mei	28	84	4	1009	454	22	5	4934	26522	5,38
Kabupaten Mojokerto	Juni	28	82	3	1009	190	15	7	4209	22588	5,37
Kabupaten Mojokerto	Juli	28	78	5	1009	153	8	8	4477	24031	5,37
Kabupaten Mojokerto	Agustus	28	77	5	1009	11	3	8	5075	27236	5,37
Kabupaten Mojokerto	September	29	76	4	1010	152	1	7	2881	15398	5,34
Kabupaten Mojokerto	Oktober	28	83	4	1009	261	14	5	2509	13415	5,35

Kabupaten Mojokerto	November	28	86	3	1009	216	22	4	942	5034	5,34
Kabupaten Mojokerto	Desember	28	83	5	1007	192	19	4	1780	9515	5,34
Kabupaten Jombang	Januari	27	84	5	1008	405	23	5	994	6587	6,62
Kabupaten Jombang	Februari	27	84	5	1008	184	23	4	192	1271	6,62
Kabupaten Jombang	Maret	28	85	4	1008	414	19	5	7860	52063	6,62
Kabupaten Jombang	April	28	82	3	1008	227	13	6	16175	107143	6,62
Kabupaten Jombang	Mei	28	84	4	1009	454	22	5	5440	30335	5,58
Kabupaten Jombang	Juni	28	82	3	1009	190	15	7	535	2984	5,58
Kabupaten Jombang	Juli	28	78	5	1009	153	8	8	5464	30469	5,58
Kabupaten Jombang	Agustus	28	77	5	1009	11	3	8	8850	49348	5,58
Kabupaten Jombang	September	29	76	4	1010	152	1	7	5975	36268	6,07
Kabupaten Jombang	Oktober	28	83	4	1009	261	14	5	706	4288	6,07
Kabupaten Jombang	November	28	86	3	1009	216	22	4	1069	6489	6,07
Kabupaten Jombang	Desember	28	83	5	1007	192	19	4	1635	9926	6,07
Kabupaten Ngajuk	Januari	25	84	3	934	498	25	4	1119	6549	5,85
Kabupaten Ngajuk	Februari	25	86	3	933	491	21	3	3079	18028	5,85
Kabupaten Ngajuk	Maret	25	86	2	933	343	27	4	18785	109981	5,85
Kabupaten Ngajuk	April	26	83	3	934	400	18	5	11000	64400	5,85
Kabupaten Ngajuk	Mei	26	83	3	934	310	18	6	1828	8352	4,57
Kabupaten Ngajuk	Juni	25	81	2	934	218	13	7	8440	38570	4,57
Kabupaten Ngajuk	Juli	26	77	3	934	34	3	7	14317	65424	4,57
Kabupaten Ngajuk	Agustus	26	74	3	934	130	4	8	3491	15950	4,57
Kabupaten Ngajuk	September	27	70	3	935	47	6	7	1852	11019	5,95
Kabupaten Ngajuk	Oktober	25	83	2	934	399	18	4	1795	10678	5,95
Kabupaten Ngajuk	November	25	85	2	934	665	26	4	2628	15636	5,95
Kabupaten Ngajuk	Desember	25	84	3	933	303	21	4	1998	11889	5,95
Kabupaten Madiun	Januari	28	88	6	934	355	7	7	1705	9822	5,76
Kabupaten Madiun	Februari	27	89	3	933	246	15	5	171	883	5,16
Kabupaten Madiun	Maret	28	88	3	933	429	19	5	13690	79297	5,79
Kabupaten Madiun	April	28	85	4	934	166	10	6	10749	62284	5,79
Kabupaten Madiun	Mei	28	86	4	934	134	10	6	2940	15005	5,10
Kabupaten Madiun	Juni	28	85	4	934	162	11	7	1058	5394	5,10

Kabupaten Madiun	Juli	28	83	5	934	9	1	7	16823	85857	5,10
Kabupaten Madiun	Agustus	28	81	5	934	28	2	6	6924	35340	5,10
Kabupaten Madiun	September	29	78	5	935	59	4	7	2008	11384	5,67
Kabupaten Madiun	Oktober	28	85	7	934	234	14	5	856	4836	5,65
Kabupaten Madiun	November	28	86	5	934	484	19	5	10492	59473	5,67
Kabupaten Madiun	Desember	28	82	4	933	236	13	3	5645	32001	5,67
Kabupaten Magetan	Januari	25	84	3	930	498	26	4	149	955	6,39
Kabupaten Magetan	Februari	25	86	3	929	491	21	3	2214	14151	6,39
Kabupaten Magetan	Maret	25	86	2	929	343	27	4	11266	71998	6,39
Kabupaten Magetan	April	26	83	3	930	400	19	5	3848	24589	6,39
Kabupaten Magetan	Mei	26	83	3	930	310	19	6	790	4481	5,67
Kabupaten Magetan	Juni	25	81	2	931	218	14	7	6448	36570	5,67
Kabupaten Magetan	Juli	26	77	3	931	34	5	7	8134	46137	5,67
Kabupaten Magetan	Agustus	26	74	3	932	130	4	8	1479	8386	5,67
Kabupaten Magetan	September	27	70	3	930	47	6	7	672	4144	6,17
Kabupaten Magetan	Oktober	25	83	2	930	399	18	4	2254	13898	6,17
Kabupaten Magetan	November	25	85	2	930	665	26	4	5060	31206	6,17
Kabupaten Magetan	Desember	25	84	3	929	303	21	4	1652	10185	6,17
Kabupaten Ngawi	Januari	25	84	3	930	498	26	4	845	5184	6,14
Kabupaten Ngawi	Februari	25	86	3	929	491	21	3	6182	37705	6,10
Kabupaten Ngawi	Maret	25	86	2	929	343	27	4	26459	161889	6,12
Kabupaten Ngawi	April	26	83	3	930	400	19	5	11894	72822	6,12
Kabupaten Ngawi	Mei	26	83	3	930	310	19	6	1164	6447	5,54
Kabupaten Ngawi	Juni	25	81	2	931	218	14	7	9527	52579	5,52
Kabupaten Ngawi	Juli	26	77	3	931	34	5	7	26853	148534	5,53
Kabupaten Ngawi	Agustus	26	74	3	932	130	4	8	7598	42047	5,53
Kabupaten Ngawi	September	27	70	3	930	47	6	7	988	5947	6,02
Kabupaten Ngawi	Oktober	25	83	2	930	399	18	4	6058	36290	5,99
Kabupaten Ngawi	November	25	85	2	930	665	26	4	22120	133000	6,01
Kabupaten Ngawi	Desember	25	84	3	929	303	21	4	8899	53495	6,01
Kabupaten Bojonegoro	Januari	23	85	2	1003	474	23	5	1602	9363	5,85
Kabupaten Bojonegoro	Februari	23	86	2	1003	670	24	4	12711	74279	5,84

Kabupaten Bojonegoro	Maret	21	86	1	1002	583	23	7	49075	283043	5,77
Kabupaten Bojonegoro	April	24	84	1	1003	783	15	8	7751	42332	5,46
Kabupaten Bojonegoro	Mei	23	85	1	1003	518	16	8	2481	10979	4,42
Kabupaten Bojonegoro	Juni	23	83	1	1004	261	16	6	23745	105282	4,43
Kabupaten Bojonegoro	Juli	21	81	1	1004	961	8	7	16401	72574	4,42
Kabupaten Bojonegoro	Agustus	23	79	2	1005	768	5	9	2668	11806	4,42
Kabupaten Bojonegoro	September	24	77	1	1003	736	10	8	3339	18335	5,49
Kabupaten Bojonegoro	Oktober	23	80	1	1003	461	20	6	4625	25398	5,49
Kabupaten Bojonegoro	November	23	84	1	1002	344	19	5	5612	30821	5,49
Kabupaten Bojonegoro	Desember	22	83	1	1001	652	18	5	3656	20077	5,49
Kabupaten Lamongan	Januari	23	85	2	1003	474	23	5	5500	31883	5,80
Kabupaten Lamongan	Februari	23	86	2	1003	670	24	4	6434	37298	5,80
Kabupaten Lamongan	Maret	21	86	1	1002	583	23	7	26324	152503	5,79
Kabupaten Lamongan	April	24	84	1	1003	783	15	8	11951	69187	5,79
Kabupaten Lamongan	Mei	23	85	1	1003	518	16	8	2788	14112	5,06
Kabupaten Lamongan	Juni	23	83	1	1004	261	16	6	2690	13616	5,06
Kabupaten Lamongan	Juli	21	81	1	1004	961	8	7	8510	43071	5,06
Kabupaten Lamongan	Agustus	23	79	2	1005	768	5	9	3555	17992	5,06
Kabupaten Lamongan	September	24	77	1	1003	736	10	8	6945	47242	6,80
Kabupaten Lamongan	Oktober	23	80	1	1003	461	20	6	6901	46945	6,80
Kabupaten Lamongan	November	23	84	1	1002	344	19	5	2892	19676	6,80
Kabupaten Lamongan	Desember	22	83	1	1001	652	18	5	796	5414	6,80
Kabupaten Tuban	Januari	23	85	2	1003	111	18	5	792	5040	6,36
Kabupaten Tuban	Februari	23	86	2	1003	247	20	4	18456	117228	6,35
Kabupaten Tuban	Maret	21	86	1	1002	165	16	7	39721	252152	6,35
Kabupaten Tuban	April	24	84	1	1003	132	10	8	4669	29709	6,36
Kabupaten Tuban	Mei	23	85	1	1003	138	15	8	6151	34231	5,57
Kabupaten Tuban	Juni	23	83	1	1004	113	19	6	24522	136534	5,57
Kabupaten Tuban	Juli	21	81	1	1004	127	12	7	16276	90624	5,57
Kabupaten Tuban	Agustus	23	79	2	1005	46	7	9	13591	75716	5,57
Kabupaten Tuban	September	24	77	1	1003	48	8	8	10242	62340	6,09
Kabupaten Tuban	Oktober	23	80	1	1003	172	19	6	9540	58047	6,08

Kabupaten Tuban	November	23	84	1	1002	171	20	5	6036	36736	6,09
Kabupaten Tuban	Desember	22	83	1	1001	296	18	5	908	5526	6,09
Kabupaten Gresik	Januari	28	85	5	1009	301	28	4	95	619	6,50
Kabupaten Gresik	Februari	28	85	4	1008	281	25	4	15730	102195	6,50
Kabupaten Gresik	Maret	28	84	3	1008	124	23	7	13961	90704	6,50
Kabupaten Gresik	April	28	83	2	1008	136	28	8	2150	13965	6,50
Kabupaten Gresik	Mei	28	84	3	1008	228	27	5	2181	14177	6,50
Kabupaten Gresik	Juni	28	83	4	1009	310	28	5	16544	107551	6,50
Kabupaten Gresik	Juli	28	80	6	1009	144	21	7	4256	27669	6,50
Kabupaten Gresik	Agustus	28	78	6	1009	23	22	9	2254	14655	6,50
Kabupaten Gresik	September	28	79	4	1010	67	21	7	1053	6837	6,49
Kabupaten Gresik	Oktober	28	84	2	1009	237	25	5	3522	22863	6,49
Kabupaten Gresik	November	28	86	2	1008	211	27	4	1279	8304	6,49
Kabupaten Gresik	Desember	27	84	3	1007	576	29	5	262	1703	6,49
Kabupaten Bangkalan	Januari	28	82	2	1009	265	19	4	312	1539	4,93
Kabupaten Bangkalan	Februari	27	82	2	1008	427	23	4	5174	25496	4,93
Kabupaten Bangkalan	Maret	28	82	2	1008	360	18	5	17243	85040	4,93
Kabupaten Bangkalan	April	29	78	2	1008	99	7	6	2537	12486	4,92
Kabupaten Bangkalan	Mei	29	79	3	1009	105	11	5	972	4600	4,73
Kabupaten Bangkalan	Juni	28	78	3	1009	98	15	6	5542	26237	4,73
Kabupaten Bangkalan	Juli	29	73	4	1009	12	4	8	4921	23280	4,73
Kabupaten Bangkalan	Agustus	29	71	4	1009	1	2	8	2056	9726	4,73
Kabupaten Bangkalan	September	30	69	4	1010	1	1	7	752	3727	4,95
Kabupaten Bangkalan	Oktober	28	79	2	1009	288	17	4	1041	5158	4,95
Kabupaten Bangkalan	November	28	83	2	1009	423	22	3	168	832	4,95
Kabupaten Bangkalan	Desember	28	81	3	1007	190	18	4	0	0	0,00
Kabupaten Sampang	Januari	28	85	5	1003	344	28	4	0	0	0,00
Kabupaten Sampang	Februari	27	87	4	1004	439	28	4	5340	26771	5,01
Kabupaten Sampang	Maret	28	86	3	1004	274	22	5	19627	97789	4,98
Kabupaten Sampang	April	28	84	2	1005	172	27	7	366	1875	5,12
Kabupaten Sampang	Mei	28	85	3	1005	237	27	6	168	802	4,77
Kabupaten Sampang	Juni	28	82	4	1001	147	26	7	8436	40657	4,82

Kabupaten Sampang	Juli	28	79	6	1001	45	27	8	822	3921	4,77
Kabupaten Sampang	Agustus	28	77	6	1006	21	25	9	0	0	0,00
Kabupaten Sampang	September	29	75	5	1005	3	27	9	0	0	0,00
Kabupaten Sampang	Oktober	28	81	3	1005	153	25	6	122	679	5,57
Kabupaten Sampang	November	28	86	2	1004	280	26	5	0	0	0,00
Kabupaten Sampang	Desember	28	83	4	1002	104	28	5	0	0	0,00
Kabupaten Pamekasan	Januari	28	85	8	1009	344	28	4	0	0	0,00
Kabupaten Pamekasan	Februari	27	87	9	1008	439	28	4	1145	5355	4,68
Kabupaten Pamekasan	Maret	28	86	9	1008	274	22	5	17004	90219	5,31
Kabupaten Pamekasan	April	28	84	8	1008	172	27	7	990	5129	5,18
Kabupaten Pamekasan	Mei	28	85	8	1009	237	27	6	15	76	5,06
Kabupaten Pamekasan	Juni	28	82	8	1009	147	26	7	284	1405	4,95
Kabupaten Pamekasan	Juli	28	79	8	1009	45	27	8	228	1116	4,90
Kabupaten Pamekasan	Agustus	28	77	8	1010	21	25	9	89	433	4,89
Kabupaten Pamekasan	September	29	75	8	1010	3	27	9	742	3916	5,28
Kabupaten Pamekasan	Oktober	28	81	8	1009	153	25	6	0	0	0,00
Kabupaten Pamekasan	November	28	86	9	1009	280	26	5	0	0	0,00
Kabupaten Pamekasan	Desember	28	83	8	1008	104	28	5	0	0	0,00
Kabupaten Sumenep	Januari	28	85	5	1010	421	24	4	614	3403	5,54
Kabupaten Sumenep	Februari	27	88	4	1009	398	23	5	3603	20342	5,65
Kabupaten Sumenep	Maret	28	87	3	1009	258	23	6	20283	110794	5,46
Kabupaten Sumenep	April	28	85	2	1010	172	14	8	5833	32538	5,58
Kabupaten Sumenep	Mei	29	84	3	1010	220	16	7	460	2511	5,46
Kabupaten Sumenep	Juni	28	82	4	1010	114	16	8	3000	16358	5,45
Kabupaten Sumenep	Juli	28	79	6	1011	39	9	8	3284	17740	5,40
Kabupaten Sumenep	Agustus	28	76	6	1011	21	8	9	2070	11321	5,47
Kabupaten Sumenep	September	29	74	5	1012	3	2	9	853	4999	5,86
Kabupaten Sumenep	Oktober	29	80	3	1011	145	12	7	1141	6530	5,72
Kabupaten Sumenep	November	28	86	2	1010	280	25	5	611	3554	5,82
Kabupaten Sumenep	Desember	28	81	4	1009	184	20	6	84	491	5,87
Kota Kediri	Januari	25	84	3	934	291	18	4	107	623	5,80
Kota Kediri	Februari	25	86	3	933	302	18	3	154	892	5,80

Kota Kediri	Maret	25	86	3	933	291	21	4	176	1021	5,80
Kota Kediri	April	26	83	3	934	316	15	5	413	2394	5,80
Kota Kediri	Mei	26	83	3	934	118	16	6	132	619	4,68
Kota Kediri	Juni	25	81	2	934	133	12	7	18	84	4,68
Kota Kediri	Juli	26	77	3	934	21	2	7	133	624	4,68
Kota Kediri	Agustus	26	74	3	935	103	3	8	390	1824	4,68
Kota Kediri	September	27	70	3	935	18	3	7	72	446	6,23
Kota Kediri	Oktober	25	83	2	934	307	13	4	25	155	6,23
Kota Kediri	November	25	85	2	934	584	16	3	79	493	6,23
Kota Kediri	Desember	25	84	3	933	218	12	4	164	1023	6,23
Kota Blitar	Januari	28	76	3	977	248	23	5	6	47	7,64
Kota Blitar	Februari	28	78	3	977	264	21	4	0	0	0,00
Kota Blitar	Maret	28	78	3	976	266	27	5	30	230	7,64
Kota Blitar	April	28	78	3	977	155	22	5	434	3315	7,64
Kota Blitar	Mei	28	76	2	977	135	26	6	209	1316	6,30
Kota Blitar	Juni	27	75	2	978	153	27	6	0	0	0,00
Kota Blitar	Juli	27	72	3	978	9	2	6	6	39	6,30
Kota Blitar	Agustus	27	72	3	978	98	16	6	25	156	6,30
Kota Blitar	September	27	74	3	979	294	26	6	19	109	5,87
Kota Blitar	Oktober	27	81	2	978	362	24	4	0	0	0,00
Kota Blitar	November	27	82	2	977	387	21	3	0	0	0,00
Kota Blitar	Desember	27	78	2	976	367	24	4	0	0	0,00
Kota Malang	Januari	25	79	3	944	216	23	4	71	454	6,36
Kota Malang	Februari	25	81	3	944	353	22	4	214	1364	6,36
Kota Malang	Maret	25	81	3	944	349	24	4	232	1476	6,36
Kota Malang	April	25	77	3	944	355	21	6	31	196	6,36
Kota Malang	Mei	26	76	3	945	75	11	7	129	781	6,03
Kota Malang	Juni	25	78	3	945	308	16	6	148	893	6,03
Kota Malang	Juli	25	73	4	945	20	3	7	137	828	6,03
Kota Malang	Agustus	25	74	4	945	44	8	7	157	950	6,03
Kota Malang	September	25	77	4	946	127	12	7	156	1002	6,41
Kota Malang	Oktober	24	83	3	945	496	24	4	56	358	6,41

Kota Malang	November	25	84	3	945	474	27	4	173	1110	6,41
Kota Malang	Desember	25	78	4	943	245	20	5	212	1358	6,41
Kota Probolinggo	Januari	22	91	1	911	683	26	2	19	119	6,30
Kota Probolinggo	Februari	22	90	1	912	266	21	2	0	0	0,00
Kota Probolinggo	Maret	23	90	1	911	465	25	2	155	977	6,30
Kota Probolinggo	April	23	88	1	912	389	18	4	569	3586	6,30
Kota Probolinggo	Mei	23	88	1	913	358	18	4	470	2832	6,02
Kota Probolinggo	Juni	23	85	1	906	156	12	4	49	297	6,02
Kota Probolinggo	Juli	23	84	1	913	28	6	5	0	0	0,00
Kota Probolinggo	Agustus	23	78	2	999	93	5	5	0	0	0,00
Kota Probolinggo	September	24	76	2	999	36	4	5	15	71	4,78
Kota Probolinggo	Oktober	23	89	1	999	309	20	2	46	221	4,78
Kota Probolinggo	November	22	91	1	999	407	17	2	0	0	0,00
Kota Probolinggo	Desember	22	89	1	999	348	19	2	0	0	0,00
Kota Pasuruan	Januari	21	94	1	915	683	28	2	38	205	5,39
Kota Pasuruan	Februari	21	92	2	914	266	24	2	97	521	5,39
Kota Pasuruan	Maret	22	91	2	914	465	27	3	107	577	5,39
Kota Pasuruan	April	22	90	2	915	389	23	5	94	505	5,39
Kota Pasuruan	Mei	22	90	2	915	358	26	4	140	691	4,93
Kota Pasuruan	Juni	22	88	2	915	156	23	6	230	1133	4,93
Kota Pasuruan	Juli	21	86	2	915	28	22	6	100	491	4,93
Kota Pasuruan	Agustus	21	84	2	999	93	21	6	94	461	4,93
Kota Pasuruan	September	22	80	2	999	36	27	6	104	571	5,47
Kota Pasuruan	Oktober	22	91	1	999	309	24	3	157	857	5,47
Kota Pasuruan	November	22	93	1	999	407	19	2	185	1011	5,47
Kota Pasuruan	Desember	22	91	1	999	348	23	2	49	271	5,47
Kota Mojokerto	Januari	27	84	5	1008	405	28	5	0	0	0,00
Kota Mojokerto	Februari	27	84	5	1008	184	26	4	30	181	5,99
Kota Mojokerto	Maret	28	85	4	1008	414	30	5	0	0	0,00
Kota Mojokerto	April	28	82	3	1008	227	26	6	111	665	5,99
Kota Mojokerto	Mei	28	84	4	1009	454	30	5	158	890	5,62
Kota Mojokerto	Juni	28	82	3	1009	190	27	7	10	57	5,62

Kota Mojokerto	Juli	28	78	5	1009	153	30	8	57	321	5,62
Kota Mojokerto	Agustus	28	77	5	1009	11	29	8	103	578	5,62
Kota Mojokerto	September	29	76	4	1010	152	27	7	132	849	6,44
Kota Mojokerto	Oktober	28	83	4	1009	261	25	5	111	716	6,44
Kota Mojokerto	November	28	86	3	1009	216	28	4	30	194	6,44
Kota Mojokerto	Desember	28	83	5	1007	192	28	4	90	583	6,44
Kota Madiun	Januari	25	84	3	934	498	25	4	0	0	0,00
Kota Madiun	Februari	25	86	3	933	491	21	3	0	0	0,00
Kota Madiun	Maret	25	86	2	933	343	27	4	506	2932	5,80
Kota Madiun	April	26	83	3	934	400	18	5	282	1634	5,80
Kota Madiun	Mei	26	83	3	934	310	18	6	0	0	0,00
Kota Madiun	Juni	25	81	2	934	218	13	7	0	0	0,00
Kota Madiun	Juli	26	77	3	934	34	3	7	712	3051	4,28
Kota Madiun	Agustus	26	74	3	934	130	4	8	75	322	4,28
Kota Madiun	September	27	70	3	935	47	6	7	0	0	0,00
Kota Madiun	Oktober	25	83	2	934	399	18	4	15	82	5,67
Kota Madiun	November	25	85	2	934	665	26	4	499	2829	5,67
Kota Madiun	Desember	25	84	3	933	303	21	4	107	604	5,67
Kota Surabaya	Januari	28	82	2	1009	265	19	4	17	98	5,90
Kota Surabaya	Februari	27	82	2	1008	427	23	4	167	982	5,90
Kota Surabaya	Maret	28	82	2	1008	360	18	5	579	3417	5,90
Kota Surabaya	April	29	78	2	1008	99	7	6	147	866	5,90
Kota Surabaya	Mei	29	79	3	1009	105	11	5	46	208	4,54
Kota Surabaya	Juni	28	78	3	1009	98	15	6	33	148	4,54
Kota Surabaya	Juli	29	73	4	1009	12	4	8	127	577	4,54
Kota Surabaya	Agustus	29	71	4	1009	1	2	8	163	742	4,54
Kota Surabaya	September	30	69	4	1010	1	1	7	82	520	6,35
Kota Surabaya	Oktober	28	79	2	1009	288	17	4	16	101	6,35
Kota Surabaya	November	28	83	2	1009	423	22	3	32	203	6,35
Kota Surabaya	Desember	28	81	3	1007	190	18	4	40	255	6,35
Kota Batu	Januari	25	79	3	944	216	29	4	114	686	6,03
Kota Batu	Februari	25	81	3	944	353	27	4	121	728	6,03

Kota Batu	Maret	25	81	3	944	349	29	4	51	310	6,03
Kota Batu	April	25	77	3	944	355	29	6	60	360	6,03
Kota Batu	Mei	26	76	3	945	75	28	7	90	551	6,11
Kota Batu	Juni	25	78	3	945	308	27	6	145	885	6,11
Kota Batu	Juli	25	73	4	945	20	29	7	103	628	6,11
Kota Batu	Agustus	25	74	4	945	44	27	7	114	696	6,11
Kota Batu	September	25	77	4	946	127	29	7	49	355	7,32
Kota Batu	Oktober	24	83	3	945	496	30	4	60	437	7,32
Kota Batu	November	25	84	3	945	474	30	4	26	193	7,32
Kota Batu	Desember	25	78	3	943	245	31	5	128	934	7,32