

**EVALUASI PERFORMA ALGORITMA MACHINE
LEARNING UNTUK PREDIKSI PRODUKTIVITAS
TANAMAN PADI DI JAWA TENGAH**

SKRIPSI SARJANA KOMPUTER

Disusun Oleh:

GUING TRI SUHATMAJO

207064516083



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA**

UNIVERSITAS NASIONAL

2025

HALAMAN PENGESAHAN

TUGAS AKHIR

EVALUASI PERFORMA ALGORITMA MACHINE LEARNING UNTUK
PREDIKSI PRODUKTIVITAS TANAMAN PADI DI JAWA TENGAH



Guing Tri Suhatmajo

207064516083

Dosen Pembimbing 1



Ratih Titi Komala Sari, ST., MM., MMSI.

Dosen Pembimbing 2



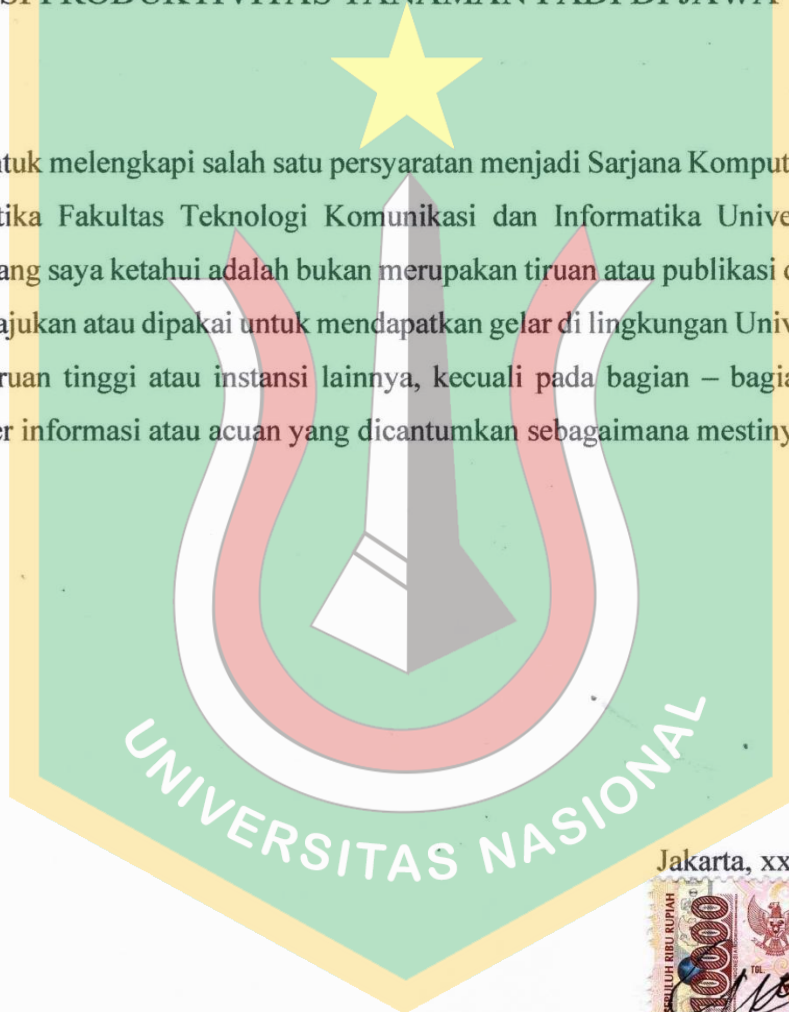
Rima Tamara Aldisa, S.Kom., M.Kom

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

EVALUASI PERFORMA ALGORITMA MACHINE LEARNING UNTUK PREDIKSI PRODUKTIVITAS TANAMAN PADI DI JAWA TENGAH

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.



Jakarta, xxxxxxxxxx



Guing Tri Suhatmajo

207064516083

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

**“EVALUASI PERFORMA ALGORITMA MACHINE LEARNING UNTUK
PREDIKSI PRODUKTIVITAS TANAMAN PADI DI JAWA TENGAH”**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2021-2022 pada tanggal (isi tanggal) Tahun 2022



Dosen Pembimbing 1

Ratih Titi Komala Sari, S.T.,
M.M., MMSI
NIDN: 0301038302

Ketua Program Studi

Ratih Titi Komala Sari, S.T.,
M.M., MMSI
NIDN: 0301038302

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Guing Tri Suhatmajo

NPM : 2070645166083

Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika

Program Studi : Informatika

Tanggal Sidang : 25 Februari 2026

JUDUL DALAM BAHASA INDONESIA :

“EVALUASI PERFORMA ALGORITMA MACHINE LEARNING UNTUK
PREDIKSI PRODUKTIVITAS TANAMAN PADI DI JAWA TENGAH”

JUDUL DALAM BAHASA INGGRIS :

“EVALUATION OF MACHINE LEARNING ALGORITHM PERFORMANCE FOR
RICE PRODUCTIVITY PREDICTION IN CENTRAL JAVA”

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 3 Maret 2026	TGL : 3 Maret 2026	TGL : 3 Maret 2026
		

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Guing Tri Suhatmajo
NPM : 207064516083
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 25 Februari 2026

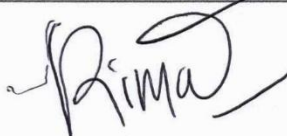
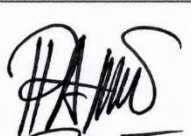
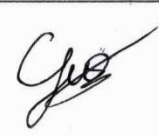
JUDUL DALAM BAHASA INDONESIA :

“EVALUASI PERFORMA ALGORITMA MACHINE LEARNING
UNTUK PREDIKSI PRODUKTIVITAS TANAMAN PADI DI JAWA
TENGAH”

JUDUL DALAM BAHASA INGGRIS :

“EVALUATION OF MACHINE LEARNING ALGORITHM PERFORMANCE FOR
RICE PRODUCTIVITY PREDICTION IN CENTRAL JAVA”

TANDA TANGAN DAN TANGGAL

Pembimbing 2	Ka. Prodi	Mahasiswa
TGL : 3 Maret 2026	TGL : 3 Maret 2026	TGL : 3 Maret 2026
		

Abstrak

Penelitian ini membandingkan performa tiga algoritma *machine learning* yakni Linear Regression, Random Forest, dan XGBoost dalam memprediksi produktivitas padi di 29 kabupaten Jawa Tengah periode 2004–2024 (609 observasi panel). Metodologi mengikuti kerangka *Knowledge Discovery in Database* (KDD) dengan evaluasi menggunakan *TimeSeriesSplit 5-fold* dan metrik MAE, RMSE, dan R^2 . Hasil penelitian menunjukkan Random Forest sebagai model terbaik dengan Test $R^2 = 0,6021$, MAE = 2,43 ku/ha, dan MAPE = 4,31%, mengungguli XGBoost ($R^2 = 0,5707$) dan Linear Regression yang gagal total. Analisis feature importance mengungkap karakteristik tetap per kabupaten sebagai prediktor paling dominan (51% total importance), dengan *district_Sukoharjo* (26%) sebagai yang terkuat, diikuti luas panen (11%) dan tren temporal tahun (9%). Penelitian ini mengonfirmasi superioritas ensemble methods dalam memprediksi produktivitas pertanian dan menghasilkan model dengan akurasi excellent yang dapat diimplementasikan sebagai early warning system untuk mendukung ketahanan pangan di Jawa Tengah.

Kata Kunci: prediksi produktivitas padi, Random Forest, XGBoost, ensemble methods, Jawa Tengah



Abstrack

This study compares the performance of three machine learning algorithms its Linear Regression, Random Forest, and XGBoost in predicting rice productivity across 29 districts in Central Java for the period 2004–2024 (609 panel observations). The methodology follows the Knowledge Discovery in Database (KDD) framework with evaluation using 5-fold TimeSeriesSplit and MAE, RMSE, and R^2 metrics. Results show Random Forest as the best model with Test $R^2 = 0.6021$, MAE = 2.43 ku/ha, and MAPE = 4.31%, outperforming XGBoost ($R^2 = 0.5707$) and Linear Regression which completely failed. Feature importance analysis reveals district fixed characteristics as the most dominant predictor (51% of total importance), with district_Sukoharjo (26%) as the strongest, followed by harvested area (11%) and year trend (9%). This study confirms the superiority of ensemble methods in predicting agricultural productivity and produces a model with excellent accuracy that can be implemented as an early warning system to support food security in Central Java.

Keywords: rice productivity prediction, Random Forest, XGBoost, ensemble methods, Central Java



DAFTAR ISI

BAB I PENDAHULUAN	10
1.1 Latar Belakang	10
1.2 Identifikasi Masalah	13
1.3 Tujuan Penelitian.....	14
1.4 Manfaat penelitian.....	14
1.5 Ruang Lingkup Penelitian.....	15
1.6 Batasan Masalah.....	16
1.7 Sistematika Penulisan	17
1.8 Mata Kuliah Pendukung.....	18
BAB II TINJAUAN PUSTAKA	20
2.1 Penelitian Terdahulu.....	20
2.2 Landasan Teori	31
2.2.1 Padi.....	31
2.2.2 Produktivitas	32
2.2.3 Faktor-faktor yang Mempengaruhi Produktivitas Padi.....	33
2.2.4 Machine Learning	35
2.2.5 Linear Regression.....	36
2.2.6 Random Forest (RF).....	37
2.2.7 Extreme Gradient Boost (XGBoost)	38
2.3 Kerangka Pemikiran.....	38
2.4 Hipotesis Penelitian.....	40
BAB III METEDOLOGI PENELITIAN	41
3.1 Jenis dan Rancangan Penelitian	41
3.2 Lokasi dan Waktu Penelitian.....	41
3.3 Data dan Sumber Data	42
3.4 Variabel Penelitian	44
3.4.1 Variabel Terikat (Y).....	45
3.4.2 Variabel Bebas (X) - Kandidat Prediktor	45
3.5 Populasi dan Sampel	48
3.5.1 Populasi.....	48

3.5.2	Sampel.....	48
3.6	Proses Penelitian	51
3.6.1	Selection (Pemilihan Data)	53
3.6.2	Preprocessing (Pra-pemrosesan Data).....	53
3.6.3	Data Transformation (Transformasi Data)	54
3.6.4	Pemodelan Machine Learning (Data Mining)	57
3.6.5	Intepretation dan Evaluation	59
3.7	Teknik Pengumpulan Data	61
3.8	Instrumen Penelitian.....	62
3.8.1	Bahasa Pemrograman.....	62
3.8.2	Lingkungan Pengembangan	63
3.8.3	Pustaka Python (Library)	63
3.9	Teknik Analisis Data	66
3.9.1	Pra-pemrosesan Data (Preprocessing)	66
3.9.2	Pembagian Data	68
3.9.3	Pemodelan Machine Learning.....	69
3.9.4	Evaluasi Model.....	74
3.9.5	Perbandingan Model	75
BAB IV	PEMBAHASAN.....	78
4.1	Karakteristik Awal Dataset.....	78
4.2	Penilaian Kualitas Data dan Preprocessing.....	80
4.2.1	Missing Values dan Duplikasi	81
4.2.2	Deteksi Dan Penanganan Outliers.....	81
4.2.3	Analisis Skewness dan Transformasi	82
4.3	Eksplorasi Data (EDA) dan Pola yang Ditemukan	83
4.3.1	Tren Temporal Produktivitas.....	83
4.3.2	Variasi Spasial Antar Kabupaten	84
4.3.3	Analisis Korelasi	84
4.4	Feature Engineering dan Persiapan Pemodelan	86
4.4.1	Temporal Train-Test Split	86
4.4.2	Iterasi Feature Engineering dan Justifikasi Pemilihan Versi Final	87

4.4.3	Fitur Final yang digunakan	88
4.4.4	Standarisasi	90
4.5	Implementasi Model dan Validasi Silang (CrossValidation)	90
4.5.1	Strategi Validasi: TimeSeriesSplit.....	90
4.5.2	Hyperparameter Tuning	91
4.5.3	Hasil Cross-Validation: Iterasi Awal	92
4.5.4	Hasil Cross-Validation: Konfigurasi Final	92
4.6	Evaluasi Model pada Data Uji	95
4.7	Analisis Residual Model Terbaik	97
4.7.1	Residual vs Predicted	97
4.7.2	Distribusi Residual dan Normalitas	98
4.7.3	Residual terhadap Waktu.....	98
4.8	Analisis Feature Importance.....	99
4.9	Diskusi dan Interpretasi Komperhensif.....	102
4.9.1	Jawaban Terhadap Pertanyaan Penelitian	102
4.9.2	Konfirmasi Hipotesis	102
4.9.3	Keterbatasan Penelitian.....	102
4.9.4	Implikasi Praktis	103
4.10	Ringkasan Bab IV	103
BAB V PENUTUP.....		105
5.1	Kesimpulan	105
5.2	Saran.....	107
5.2.1	Saran Untuk Penilitan Selanjutnya.....	107
5.2.2	Saran untuk Praktisi dan Stakeholder	108

Daftar Tabel

Tabel 4.1 Karakteristik Dataset	77
Tabel 4.2 Deteksi Outliers	80
Tabel 4.3 Korelasi Fitur Kandidat terhadap Produktivitas Padi (Pooled, N=609) 84	
Tabel 4.4 Pasangan Fitur dengan Multikolinearitas Tinggi ($ r > 0,8$).....	85
Tabel 4.5 Perbandingan Performa Antar Versi Feature Engineering	87
Tabel 4.6 Fitur Iklim dan Pertanian yang Digunakan dalam Pemodelan.....	88
Tabel 4.7 Grid dan Hasil Tuning Hyperparameter Random Forest	90
Tabel 4.8 Grid dan Hasil Tuning Hyperparameter XGBoost.....	90
Tabel 4.9 Hasil Cross-Validation Iterasi Awal (Feature Engineering v1, 6 Fitur). 91	
Tabel 4.10 Hasil Cross-Validation Konfigurasi Final (Feature Engineering v3, 41 Fitur).....	92
tabel 4.11 Performa Model pada Test Set.....	94



Daftar Gambar

Gambar II.1 Kerangka Berpikir.....	38
Gambar III.1 Diagram Alur Penelitian (Research Diagram)	51
Gambar IV.1 Distribusi Analisis Curah Hujan dan Suhu	79
Gambar IV.2 Distribusi Analisis Suhu dan kelembapan	79
Gambar IV.3 Distribusi Analisis Radiasi Matahari, Luas Panen, dan Produktivitas	79
Gambar IV.4 Rata-rata Produktivitas Padi di Semua Kabupaten.....	82
Gambar IV.5 Rentang Minimal dan Maksimal Produktivitas Padi.....	83
Gambar IV.6 Perbandingan Metriks MAE.....	93
Gambar IV.7 Perbandingan RMSE antar Model.....	93
Gambar IV.8 Komparasi Metriks R^2 untuk Setiap Model.....	94
Gambar IV.9 Linear Regression yang Tidak Dapat Menangkap Pola	95
Gambar IV.10 Random Forest model Terbaik Pada Test	96
Gambar IV.11 XGBoost Model Terbaik Kedua Pada Test.....	96
Gambar IV.12 Analisis Residual Model Terbaik: (a) Residual vs Predicted, (b) Distribusi Residual, (c) Q-Q Plot, (d) Residual vs Waktu (Notebook 06)	98
Gambar IV.13 Feature Importance Pada Model Terbaik (Random Forest)	100