

TESIS

**Identifikasi Faktor Yang Mempengaruhi Hasil Panen Padi Menggunakan
Regresi Linear Berganda dan *Algoritma Ensemble***



ANITA JAQUILINE LADO

245313054

MAGISTER INFORMATIKA

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ATMA JAYA YOGYAKARTA

2026

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

IDENTIFIKASI FAKTOR YANG MEMPENGARUHI HASIL PANEN PADI MENGGUNAKAN REGRESI LINEAR
BERGANDA DAN ALGORITMA ENSEMBLE

yang disusun oleh

Anita Jaquiline Lado

245313054

dinyatakan telah memenuhi syarat pada tanggal 02 April 2026

		Keterangan
Dosen Pembimbing 1	: Dra. Ernawati, M.T., Ph.D	Telah Menyetujui
Dosen Pembimbing 2	: Benyamin Langgu Sinaga, S.T., M.Comp.Sc., Ph.D.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Dra. Ernawati, M.T., Ph.D	Telah Menyetujui
Penguji 2	: Prof. Dr. Ir. Alb. Joko Santoso, M.T.	Telah Menyetujui
Penguji 3	: Paulus Mudjihartono, S.T., M.T., Ph.D	Telah Menyetujui

Yogyakarta, 02 April 2026

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Prof. Dr. Ir. Parama Kartika Dewa SP., S.T., M.T., IPU, ASEAN Eng.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

HALAMAN PERNYATAAN TESIS

Saya yang bertanda tangan di bawah ini:

Nama : Anita Jaquiline Lado
NIM : 245313054
Judul : Identifikasi Faktor Yang Mempengaruhi Hasil Panen Padi
Menggunakan Regresi Linear Berganda dan *Algoritma Ensemble*

Menyatakan dengan sesungguhnya bahwa Tesis ini merupakan hasil penelitian, pemikiran dan pemaparan asli saya sendiri. Saya tidak mencantumkan tanpa pengakuan bahan - bahan yang telah dipublikasikan sebelumnya atau ditulis oleh orang lain, atau sebagai bahan yang pernah diajukan untuk gelar atau ijasah pada Universitas Atma Jaya Yogyakarta atau perguruan tinggi lainnya.

Apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik sesuai dengan peraturan yang berlaku di Universitas Atma Jaya Yogyakarta.

Demikian pernyataan ini saya buat.

Yogyakarta, 02 April 2026

Yang membuat pernyataan,

(Anita Jaquiline Lado)

INTISARI

Produktivitas padi merupakan faktor penting dalam mendukung ketahanan pangan, namun variasi hasil panen antarpetani masih cukup tinggi. Perbedaan tersebut dipengaruhi oleh berbagai faktor seperti penggunaan input produksi, karakteristik petani, serta kondisi lahan dan lingkungan. Identifikasi faktor-faktor yang paling berpengaruh terhadap hasil panen menjadi penting untuk mendukung pengambilan keputusan dalam pengelolaan budidaya padi. Namun, hubungan antara berbagai variabel budidaya dan hasil panen seringkali bersifat kompleks sehingga sulit dianalisis menggunakan pendekatan statistik konvensional.

Penelitian ini bertujuan untuk membangun model prediksi hasil panen padi serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap produktivitas menggunakan pendekatan *machine learning*. Data penelitian diperoleh dari kuesioner terhadap 395 petani padi di Kabupaten Nagekeo yang mencakup informasi mengenai karakteristik petani, penggunaan input produksi, serta kondisi lahan. Model yang digunakan dalam penelitian ini meliputi Regresi Linear Berganda, *Random Forest*, *Gradient Boosting Tree* (GBM), dan *Extreme Gradient Boosting* (XGBoost). Evaluasi performa model dilakukan menggunakan metrik RMSE, MAE, dan R^2 , serta uji statistik *Friedman Test* dan *Wilcoxon Signed-Rank Test* untuk menguji signifikansi perbedaan performa antar model.

Hasil penelitian menunjukkan bahwa model *XGBoost* memberikan performa prediksi terbaik dengan nilai RMSE sebesar 0,9255, MAE sebesar 0,6426, dan R^2 sebesar 0,8396 pada data uji. Analisis *feature importance* menunjukkan bahwa terdapat 15 variabel dominan yang memengaruhi hasil panen, dengan variabel paling berpengaruh adalah luas lahan kategori 1–2 hektar, penggunaan benih per hektar, total pupuk per hektar, serta biaya pupuk. Secara agregat, faktor internal budidaya memberikan kontribusi sebesar 55,8%, sedangkan faktor eksternal menyumbang sekitar 44,2% terhadap variasi hasil panen.

Kata kunci: Machine Learning, Prediksi Hasil Panen Padi, Xgboost, Feature Importance, Analisis Data Petani

ABSTRACT

Rice productivity plays an important role in supporting food security; however, variations in rice yield among farmers remain relatively high. These differences are influenced by various factors, including the use of production inputs, farmer characteristics, and land and environmental conditions. Identifying the most influential factors affecting rice yield is important to support decision-making in rice cultivation management. However, the relationships between cultivation variables and yield are often complex, making them difficult to analyze using conventional statistical approaches.

This study aims to develop a predictive model for rice yield and identify the most influential factors affecting productivity using a machine learning approach. The data were collected through questionnaires from 395 rice farmers in Nagekeo Regency, covering information on farmer characteristics, production inputs, and land conditions. The models applied in this study include Multiple Linear Regression, Random Forest, *Gradient Boosted Tree Regression* (GBM), and Extreme Gradient Boosting (XGBoost). Model performance was evaluated using RMSE, MAE, and R^2 metrics, and statistical tests including the Friedman Test and Wilcoxon Signed-Rank Test were conducted to assess the significance of performance differences among models.

The results show that the XGBoost model achieved the best predictive performance with an RMSE of 0.9255, MAE of 0.6426, and R^2 of 0.8396 on the test data. Feature importance analysis identified 15 dominant variables influencing rice yield, with the most influential variables including land size category of 1–2 hectares, seed usage per hectare, total fertilizer per hectare, and fertilizer cost. In aggregate, internal cultivation factors contributed 55.8%, while external factors accounted for approximately 44.2% of the variation in rice yield.

Keywords: Machine Learning, Rice Yield Prediction, Xgboost, Feature Importance, Farmer Data Analysis

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul “Identifikasi Faktor Yang Mempengaruhi Hasil Panen Padi Menggunakan Regresi Linear Berganda dan *Algoritma Ensemble*“ dengan baik. Penulisan tesis ini menjadi syarat wajib bagi mahasiswa pascasarjana Universitas Atma Jaya Yogyakarta untuk memperoleh gelar Magister Komputer (M.Kom.) pada program studi Magister Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta. Penulis berharap, penelitian ini dapat memberi manfaat bagi semua pihak serta dapat dikembangkan lebih lanjut lagi.

Penulis menyadari bahwa dalam pembuatan tesis ini, penulis telah mendapat bantuan dari berbagai pihak baik itu berupa ide, pikiran, bimbingan, tenaga, waktu, dorongan, dan doa dari banyak pihak baik secara langsung maupun tidak langsung. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar besarnya kepada:

1. Tuhan Yang Maha Esa, atas penyertaan-Nya sehingga penelitian ini dapat selesai dengan baik dan lancar.
2. Ibu Tercinta yang telah berpulang, ayah, kakak-kakak, serta seluruh keluarga besar di Kefamenanu maupun di Flores, atas doa, kasih sayang, dukungan moral, dan semangat yang senantiasa diberikan sehingga penulis dapat berada pada tahap ini.
3. Universitas Katolik Widya Mandira Kupang, atas kepercayaan dan dukungan pendanaan melalui program beasiswa yang diberikan kepada penulis selama masa studi, sehingga seluruh rangkaian pendidikan dan penelitian dapat terlaksana dengan baik.
4. Universitas Atma Jaya Yogyakarta, khususnya Fakultas Teknologi Industri dan Program Studi Magister Informatika, atas dukungan fasilitas, layanan akademik, serta suasana pembelajaran yang kondusif selama masa studi penulis.
5. Bapak Dr. Ir. Parama Kartika Dewa SP., S.T., M.T. IPU selaku Dekan Fakultas

Teknologi Industri, Universitas Atma Jaya Yogyakarta yang telah memberikan dukungan selama pengerjaan tesis.

6. Bapak Paulus Mudjihartono, S.T., M.T., Ph.D., selaku Ketua Departemen Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta sekaligus dosen penguji, yang telah meluangkan waktu untuk memberikan arahan, evaluasi, serta masukan konstruktif demi kesempurnaan tesis ini.
7. Ibu Dra. Ernawati, M.T., Ph.D. dan Bapak Benyamin Langgu Sinaga, S.T., M.Com.Sc., Ph.D., selaku dosen pembimbing, yang telah bersedia meluangkan waktu untuk memberikan bimbingan, gagasan, ide, dorongan, koreksi, serta arahan kepada penulis sehingga tesis ini dapat diselesaikan dengan baik.
8. Bapak Prof. Dr. Ir. Albertus Joko Santoso, M.T. dan Bapak Paulus Mudjihartono, S.T., M.T., Ph.D., selaku dosen penguji, yang telah memberikan motivasi, masukan, serta saran yang sangat berharga dalam penyelesaian tesis ini.
9. Seluruh dosen, staf pengajar, dan karyawan Departemen Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta, yang telah memberikan bantuan dan dukungan kepada penulis selama menempuh pendidikan pada Program Studi Magister Informatika.
10. Seluruh responden serta pihak-pihak terkait yang telah memberikan waktu, data, dan dukungan sehingga penelitian ini dapat terlaksana dengan baik.
11. Teman-teman Program Studi Magister Informatika angkatan 2024 dan 2025, yang telah memberikan semangat, kebersamaan, dan dukungan selama masa perkuliahan hingga penyusunan tesis ini.
12. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan, dukungan, dan doa kepada penulis, baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa tesis ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun demi penyempurnaan penelitian ini di masa yang akan datang. Semoga tesis ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan serta menjadi referensi bagi penelitian selanjutnya.

Yogyakarta, 02 April 2026

Anita Jaquiline Lado

DAFTAR ISI

JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN TESIS.....	iii
INTISARI.....	iv
<i>ABSTRACT</i>	v
KATA PENGANTAR	vi
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
LAMPIRAN.....	xv
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Permasalahan	4
1.3 Pertanyaan Penelitian	4
1.4 Tujuan Penelitian	4
1.5 Batasan Permasalahan.....	5
1.6 Keaslian Penelitian.....	6
1.7 Manfaat Penelitian	7
1.8 Sistematika Penulisan	8
BAB II.....	10
TINJAUAN PUSTAKA	10
2.1 Penelitian Terdahulu Terkait Model Prediksi Hasil Panen Padi.....	10
2.2 Penelitian Terdahulu Terkait Faktor-Faktor Yang Mempengaruhi Hasil Panen Padi.....	15
2.3 Kesenjangan Penelitian	16
BAB III	18
LANDASAN TEORI.....	18
3.1 Teori Hasil Panen padi	18
3.2 Faktor – Faktor Yang Mempengaruhi Hasil Panen Padi	18
3.2.1 Faktor Internal Dalam Produksi Padi.....	19

3.2.2	Faktor Eksternal Dalam Produksi Padi	19
3.3	Pertanian Berbasis Data (<i>Data-Driven Agriculture</i>).....	20
3.4	<i>Exploratory Data Analysis</i> (EDA)	21
3.5	Data Preprocessing	22
3.6	Prediksi Hasil Panen Dalam Penelitian Pertanian.....	24
3.7	<i>Machine Learning</i> Untuk Regresi	24
3.8	Algoritma Prediksi	25
3.8.1	Regresi Linear Berganda.....	25
3.8.2	<i>Random Forest</i>	28
3.8.3	<i>Extreme Gradient Boosting (XGBoost)</i>	29
3.8.4	<i>Gradient Boosted Tree Regression</i>	31
3.9	Evaluasi Model.....	33
BAB IV		37
METODOLOGI PENELITIAN		37
4.1	Metodologi	37
4.2	Studi Literatur	39
4.3	Penentuan Variabel Dan Jumlah Responden	39
4.3.1	Penentuan Variabel Penelitian	39
4.3.2	Penentuan Jumlah Responden.....	40
4.4	Pengumpulan Data	41
4.4.1	Data Primer	42
4.4.2	Data Sekunder (Pendukung)	43
4.5	EDA dan Pra-Pemrosesan Data	43
4.5.1	Eksplorasi data awal.....	44
4.5.2	Pra-Pemrosesan Data	45
4.5.3	Eksplorasi Data Setelah Pra-Pemrosesan.....	50
4.6	Pemodelan Prediksi Hasil Panen.....	51
4.6.1	Pembagian Data (<i>Train–Test Split</i>).....	52
4.6.2	Algoritma Yang Digunakan	52
4.6.3	Proses Pelatihan Model	53
4.6.4	Penyesuaian Parameter Model (<i>Hyperparameter Tuning</i>)	54
4.7	Evaluasi Model.....	55
4.8	Pemilihan Model Terbaik.....	56
4.8.1	Kriteria Evaluasi Model	56

4.8.2	Uji Signifikansi Perbandingan Model.....	58
4.9	Analisis <i>Feature Importance</i>	59
4.10	Penyusunan Rekomendasi Berbasis Analisis <i>Feature Importance</i>	60
BAB V.....		62
HASIL DAN PEMBAHASAN.....		62
5.1	Penetapan Variabel Dan Jumlah Responden.....	62
5.1.1	Penetapan Variabel Penelitian	62
5.1.2	Jumlah Responden	63
5.2	Hasil Pengumpulan Data.....	63
5.3	Explorasi Data analysis dan Hasil Pra-Pemrosesan Data	64
5.3.1	Eksplorasi data awal.....	64
5.3.2	Pemeriksaan Dan Seleksi Data Awal.....	67
5.3.3	Transformasi Dan Validasi Variabel.....	68
5.3.4	Penanganan <i>Missing Value</i>	70
5.3.5	Deteksi <i>Outlier</i> dengan <i>IQR Winsorizing</i>	71
5.3.6	<i>Skewness</i> Dan Transformasi <i>Logaritma</i>	76
5.3.7	Pemeriksaan Variabel <i>Near-Zero Variance</i>	80
5.3.8	Normalisasi Data.....	82
5.3.9	EDA Setelah Pra-Preprocessing	83
5.3.9.2	Karakteristik Lahan Dan Sistem Pengairan	84
5.3.9.3	Penggunaan Benih, pupuk dan pestisida.....	86
5.3.9.4	Kondisi Lingkungan dan Gangguan Produksi	89
5.3.9.5	Distribusi Hasil Panen Padi.....	92
5.3.9.6	Ringkasan Hasil Eksplorasi Data	93
5.4	Pemodelan Prediksi Hasil Panen.....	94
5.4.1	Pembagian Data	94
5.4.2	Pelatihan Model	97
5.4.3	<i>Hyperparameter</i> Terbaik.....	99
5.5	Evaluasi Model (Data Uji 20%).....	101
5.5.1	Evaluasi Data Uji	101
5.5.2	Visualisasi Hasil Prediksi Model	103
5.5.3	Distribusi Error Model	105
5.5.4	Uji <i>Friedman</i> dan <i>Wilcoxon</i>	109
5.6	Model Terbaik.....	111

5.7	<i>Feature Importance</i>	113
5.7.1	Analisis Sensitivitas Jumlah Fitur.....	113
5.7.2	Hasil <i>Feature Importance</i>	118
5.7.3	Dominasi Faktor Internal Dan Eksternal.....	121
5.8	Validasi Hasil Model dengan PPL dan Petani	123
5.9	Rekomendasi Prioritas Budidaya Berdasarkan Faktor Dominan.....	125
5.9.1	Prioritas Berbasis Faktor Internal.....	125
5.9.2	Prioritas Berbasis Faktor Eksternal.....	126
5.9.3	Perumusan Strategi Peningkatan Produktivitas	126
5.9.4	Rekomendasi Praktis Bagi Petani	127
5.9.5	Rekomendasi Bagi Penyuluh Pertanian	127
5.9.6	Rekomendasi Pengembangan Model ML	128
BAB VI		129
KESIMPULAN DAN SARAN		129
6.1	Kesimpulan	129
6.2	Saran.....	130
DAFTAR PUSTAKA		132
LAMPIRAN		137

DAFTAR GAMBAR

Gambar 4. 1 Tahapan Dan Metodologi Penelitian.....	38
Gambar 5. 1 Nilai Aktual Vs Prediksi Model.....	104
Gambar 5. 2 Distribusi Rmse Hasil 10-Fold Cross Validation.....	106
Gambar 5. 3 Distribusi Mae Hasil 10-Fold Cross Validation.....	107
Gambar 5. 4 Distribusi R^2 Hasil 10-Fold Cross Validation	108
Gambar 5. 5 Perubahan Rmse Terhadap Jumlah Fitur	115
Gambar 5. 6 Perubahan Mae Terhadap Jumlah Fitur	116
Gambar 5. 7 Perubahan R^2 Terhadap Jumlah Fitur.....	117
Gambar 5. 8 Diagram <i>Feature Importance</i> (Top 15).....	119
Gambar 5. 9 Proporsi Faktor Internal Dan Eksternal.....	122

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu Terkait Model Prediksi	11
Tabel 2. 2 Penelitian Terkait Hasil Panen Padi	15
Tabel 4. 1 Tahap Pra-Pemrosesan.....	48
Tabel 5. 1 Statistik Deskriptif Variabel Numerik.....	65
Tabel 5. 2 Ringkasan Tahap Seleksi Data.....	68
Tabel 5. 3 Ringkasan Perubahan Jumlah Variabel.....	69
Tabel 5. 4 Kisaran Penggunaan Input Berdasarkan Informasi Lapangan	72
Tabel 5. 5 Batas Outlier Berdasarkan Metode Iqr.....	73
Tabel 5. 6 Perbandingan Kisaran Agronomis Dan Batas Statistik (Iqr)	74
Tabel 5. 7 Jumlah Data Yang Mengalami Winsorizing	75
Tabel 5. 8 Skewness Variabel Numerik Setelah Winsorizing	77
Tabel 5. 9 Skewness Variabel Setelah Transformasi Log.....	78
Tabel 5. 10 Karakteristik Responden Petani	84
Tabel 5. 11 Karakteristik Lahan Dan Sistem Pengairan Responden	85
Tabel 5. 12 Statistik <i>Deskriptif</i> Penggunaan Pupuk Dan Pestisida.....	88
Tabel 5. 13 Kondisi Lingkungan Dan Gangguan Produksi	90
Tabel 5. 14 Statistik Hasil Panen Padi	93
Tabel 5. 15 Pembagian Data <i>Train</i> Dan <i>Test</i>	95
Tabel 5. 16 Distribusi Hasil Panen: Data Latih Vs Uji	95
Tabel 5. 17 Ringkasan Statistik Hasil Panen: Data Latih Dan Uji.....	96
Tabel 5. 18 Perbandingan Kinerja Model (<i>10-Fold Cv</i>)	98
Tabel 5. 19 <i>Hyperparameter</i> Terbaik Setiap Model	100
Tabel 5. 20 Kinerja Model Pada Data Uji (20%).....	102
Tabel 5. 21 Hasil Uji <i>Friedman</i> Terhadap Performa Model (Rmse)	109
Tabel 5. 22 Hasil Uji <i>Post-Hoc Wilcoxon</i> (Rmse)	110
Tabel 5. 23 Hasil Uji <i>Post-Hoc Wilcoxon</i> (Mae)	110
Tabel 5. 24 Hasil Uji <i>Post-Hoc Wilcoxon</i> (R^2)	111
Tabel 5. 25 Tabel Hasil Analisis Sensitivitas Jumlah Fitur	114
Tabel 5. 26 Peringkat Feature Importance (Top 15).....	118
Tabel 5. 27 Kontribusi Faktor Internal Dan Eksternal	121

LAMPIRAN

Tabel A 1 Tabel Variabel Data Awal Sebelum Pra-Preprocessing	137
Tabel A 2 Variabel Setelah Pra-Pemrosesan	141
Tabel A 3 Variabel Penting Faktor Feature Importance	146