

OPTIMASI ARSITEKTUR ANN MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI PRODUKSI PERTANIAN BERBASIS DATA SINTETIS CTGAN

MUHAMMAD FIRLAN MAULANA



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Perpustakaan IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI, SUMBER INFORMASI, PENGGUNAAN AI, DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Optimasi Arsitektur ANN menggunakan Algoritma Genetika untuk Prediksi Produksi Pertanian Berbasis Data Sintetis CTGAN” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan buatan ChatGPT, Gemini, dan Claude untuk membantu memperkaya pemahaman serta menganalisis data. Setelah menggunakan alat/layanan tersebut, penulis meninjau dan menyunting konten sesuai kebutuhan serta bertanggung jawab penuh atas isi karya tugas akhir ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Muhammad Firlan Maulana
G1401221042



ABSTRAK

MUHAMMAD FIRLAN MAULANA. Optimasi Arsitektur ANN Menggunakan Algoritma Genetika untuk Prediksi Produksi Pertanian Berbasis Data Sintetis CTGAN. Dibimbing oleh RAHMA ANISA dan KUSMAN SADIK.

Keterbatasan amatan pada pemodelan prediksi produksi pertanian rentan memicu *overfitting* dan menyulitkan pembelajaran pola interaksi kompleks. Untuk mengatasinya, *Conditional Tabular Generative Adversarial Network* (CTGAN) diterapkan sebagai metode sintesis data, dipadukan dengan *Artificial Neural Network* (ANN) yang pengoptimalan *hyperparameter*-nya dilakukan oleh *Genetic Algorithm* (GA) guna memodelkan hubungan nonlinier. Penelitian ini bertujuan membentuk sintetis menggunakan CTGAN, membangun model prediksi produksi padi menggunakan ANN-GA berbasis data sintetis dan membandingkannya dengan regresi linier berganda, serta mengidentifikasi pengaruh peubah penjelas terhadap produksi padi menggunakan kerangka *Leave-One-Out Cross-Validation* (LOOCV). Dari tujuh skenario arsitektur CTGAN yang diuji, skenario dengan *hyperparameter* seluruhnya bernilai besar paling sering menghasilkan data sintetis terbaik berdasarkan nilai *JS Distance* dan selisih korelasi Spearman. Model regresi linier berganda menghasilkan MAPE LOOCV sebesar 13,37%, lebih rendah dibandingkan ANN-GA sebesar 19,95%. Namun model regresi hanya dapat mempertahankan dua peubah penjelas dan tidak selalu dapat memenuhi asumsi parametrik, sehingga inferensi pengaruh peubah penjelas pada kasus ini cenderung terbatas. Di sisi lain, model ANN-GA hanya menghasilkan satu titik pencilan pada sebaran APE dan mengakomodasi seluruh peubah penjelas yang digunakan. Analisis *Permutation Feature Importance* (PFI) pada ANN-GA mengidentifikasi dampak perubahan iklim, pengendalian organisme pengganggu tanaman, dan *Normalized Difference Vegetation Index* (NDVI) sebagai peubah paling berpengaruh terhadap produksi padi.

Kata kunci: algoritma genetika, ANN, CTGAN, data sintetis, LOOCV, produksi padi



ABSTRACT

MUHAMMAD FIRLAN MAULANA. Optimization of Artificial Neural Network Architecture Using Genetic Algorithm for Agricultural Production Prediction Based on Synthetic Data Generated by CTGAN. Supervised by RAHMA ANISA and KUSMAN SADIK.

Limited observations in agricultural production predictive modeling are highly susceptible to triggering overfitting and complicate the learning of complex interaction patterns. To address this, a Conditional Tabular Generative Adversarial Network (CTGAN) was applied for data synthesis, combined with an Artificial Neural Network (ANN) optimized by a Genetic Algorithm (GA) to model nonlinear relationships. This study aims to generate synthetic data using CTGAN, develop a synthetic data-based ANN-GA rice production prediction model to compare against multiple linear regression, and identify explanatory variables' effects using a Leave-One-Out Cross-Validation (LOOCV) framework. Among seven tested CTGAN architectures, the scenario with exclusively large hyperparameters most frequently produced the best synthetic data based on JS Distance and Spearman correlation differences. The multiple linear regression yielded a LOOCV MAPE of 13.37%, lower than the ANN-GA's 19.95%. However, the regression model only retained two explanatory variables and could not consistently fulfill parametric assumptions, thereby limiting inference. Conversely, the ANN-GA model produced only one outlier in the APE distribution and accommodated all explanatory variables. Permutation Feature Importance (PFI) analysis on the ANN-GA identified climate change impacts, plant pest control, and the Normalized Difference Vegetation Index (NDVI) as the most influential variables on rice production.

Keywords: ANN, CTGAN, genetic algorithm, LOOCV, rice production, synthetic data



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

OPTIMASI ARSITEKTUR ANN MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI PRODUKSI PERTANIAN BERBASIS DATA SINTETIS CTGAN

MUHAMMAD FIRLAN MAULANA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:

1 Prof. Dr. Ir. I Made Sumertajaya, M.Si

Perpustakaan IPB University



Judul Skripsi : Optimasi Arsitektur ANN menggunakan Algoritma Genetika
untuk Prediksi Produksi Pertanian Berbasis Data Sintetis CTGAN
Nama : Muhammad Firlan Maulana
NIM : G1401221042

Disetujui oleh

Pembimbing 1:

Rahma Anisa, S.Stat., M.Si., M.Act.Sc.

Pembimbing 2:

Dr. Kusman Sadik, S.Si., M.Si.

Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:

Dr. Bagus Sartono, S.Si., M.Si.

NIP 197804112005011002

Tanggal Ujian:
23 Juli 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Oktober 2025 sampai bulan Juni 2026 ini ialah optimasi pemodelan prediktif, dengan judul “Optimasi Arsitektur ANN menggunakan Algoritma Genetika untuk Prediksi Produksi Pertanian Berbasis Data Sintetis CTGAN”.

Penulis menyadari bahwa penyusunan karya ilmiah ini dapat diselesaikan berkat doa, perhatian, bimbingan, dan dukungan dari berbagai pihak. Dengan segala kerendahan hati, penulis mengucapkan terima kasih kepada:

1. Ibu Lira, Shafira, Rafa Ndut, Bapak, Tante Niken, dan Om Farid selaku keluarga penulis yang telah memberikan doa, dukungan, semangat, dan kasih sayang selama masa perkuliahan penulis.
2. Ibu Rahma Anisa, S.Stat., M.Si., M.Act.Sc. dan Bapak Dr. Kusman Sadik, S.Si., M.Si. selaku komisi pembimbing yang telah memberikan bimbingan, arahan, bantuan, serta membagikan ilmunya selama penyusunan karya ilmiah ini.
3. Ibu Laily Nissa Atul Mualifah, S.Si., M.Si. selaku moderator kolokium, Ibu Aulia Rizki Firdawanti, S.Stat., M.Si selaku moderator seminar hasil, dan Bapak Prof. Dr. Ir. I Made Sumertajaya, M.Si selaku penguji luar komisi yang turut memberikan saran untuk karya ilmiah ini.
4. Seluruh dosen dan tenaga kependidikan Program Studi Statistika dan Sains Data yang telah memberikan ilmu pengetahuan, wawasan, bimbingan, serta dukungan administratif selama penulis menempuh perkuliahan dan menyusun karya ilmiah ini.
5. Prasmanan, Suprapti, Roro Jonggrang, Intern Comel, Keluarga Lawle, dan teman-teman TSA IPB 2025 yang telah menemani penulis selama menempuh masa perkuliahan.
6. Qonita, Deu, Arief, Oci, Eca, Deswita, Yayas, dan Iyan yang telah memberikan dukungan, doa, kasih sayang, waktu, dan bantuan dalam segala bentuk untuk penulis dalam proses penulisan karya ilmiah ini.

Bogor, Agustus 2026

Muhammad Firlan Maulana



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Regresi Linier Berganda	3
2.2 <i>Conditional Tabular Generative Adversial Network</i> (CTGAN)	9
2.3 <i>Jensen-Shannon Distance (JS Distance)</i>	11
2.4 <i>Artificial Neural Network</i> (ANN)	11
2.5 Algoritma Genetika (<i>Genetic Algorithm</i> (GA))	14
2.7 <i>Leave One-Out Cross Validation</i> (LOOCV)	16
2.8 <i>Mean Absolute Percentage Error</i> (MAPE)	17
III METODE	18
3.1 Bahan dan Data	18
3.2 Prosedur Analisis	19
VI HASIL DAN PEMBAHASAN	24
4.1 Eksplorasi dan Pra-pemrosesan Data	24
4.2 Pemodelan Regresi Linier Berganda	27
4.3 Pembentukan Data Sintetis CTGAN	33
4.4 Pemodelan ANN dengan Optimasi GA	37
4.5 Perbandingan Model ANN-GA dan Regresi Linier Berganda	42
V SIMPULAN DAN SARAN	44
DAFTAR PUSTAKA	45
LAMPIRAN	50

DAFTAR TABEL

1	<i>Hyperparameter</i> ANN	13
2	Peubah yang digunakan dalam penelitian	18
3	Skenario konfigurasi arsitektur CTGAN	20
4	Pengaturan <i>hyperparameter</i> model ANN dan GA	21
5	GVIF terkoreksi dan korelasi Spearman tiap peubah penjelas	29
6	Pemilihan jumlah peubah penjelas terbaik berdasarkan SelectKBest	30
7	ANOVA model regresi linier berganda dengan transformasi logaritma natural	31
8	Ringkasan hasil pemodelan regresi linier berganda pada <i>fold</i> 18	31
9	Hasil uji asumsi klasik sisaan model regresi linier berganda dengan transformasi logaritma natural	31
10	Hasil diagnostik amatan model regresi berganda <i>fold</i> 18 dengan transformasi logaritma natural	32
11	DFBETAS model regresi linier berganda <i>fold</i> 18 dengan transformasi logaritma natural	32
12	Ringkasan hasil pembangkitan data sintesis di setiap <i>fold</i> LOOCV	34
13	Tingkat kemiripan distribusi marginal per peubah berdasarkan <i>JS Distance</i>	34
14	Korelasi Spearman antarpeubah numerik pada data asli dan data sintesis <i>fold</i> 7	37
15	Ringkasan frekuensi <i>hyperparameter</i> ANN terpilih	38
16	Peringkat 5 teratas peubah terpenting dalam model ANN-GA <i>fold</i> 5	41
17	Total skor pembobotan dan kemunculan peubah di peringkat 5 teratas tiap <i>fold</i>	41

DAFTAR GAMBAR

1	Arsitektur CTGAN (Asimopoulos <i>et al.</i> 2023)	10
2	Arsitektur ANN (Bre <i>et al.</i> 2018)	12
3	Diagram alir <i>genetic algorithm</i> (Dieterle 2003)	14
4	Diagram batang distribusi peubah kategorik (a) X1, (b) X8, (c) X9, (d) X10, (e) X11, (f) X12	24
5	Histogram peubah penjelas numerik (a) X2, (b) X3, (c) X4, (d) X5, (e) X6, (f) X7, (g) X13, (h) X14, (i) Y	25
6	<i>Scatter plot</i> hubungan peubah respons Y dengan peubah penjelas (a) X2, (b) X3, (c) X4, (d) X5, (e) X6, (f) X7, (g) X13, (h) X14	26
7	<i>Boxplot</i> hubungan peubah respons Y dengan peubah penjelas kategorik (a) X1, (b) X8, (c) X9, (d) X10, (e) X11, (f) X12	27
8	<i>Heatmap</i> korelasi Spearman antar peubah penjelas numerik dan peubah respons	27
9	Histogram setelah dilakukan agregasi peubah dosis pupuk bagi peubah numerik (a) X15, (b) X14, (c) X13	28
10	<i>Scatter plot</i> hubungan peubah respons Y dengan peubah penjelas X15	28
11	Fluktuasi nilai APE regresi linier berganda pada tiap <i>fold</i>	33



12	Perbandingan distribusi peubah kategorik (a) X_1 , (b) X_8 , (c) X_9 , (d) X_{10} , (e) X_{11} , (f) X_{12} pada data asli dan data sintetis <i>fold 7</i>	35
13	Perbandingan distribusi peubah numerik (a) X_{15} , (b) X_{13} , (c) X_{14} , dan (d) Y pada data asli dan data sintetis <i>fold 7</i>	36
14	Perbandingan visualisasi <i>heatmap</i> korelasi Spearman pada (a) data asli dan (b) data sintetis <i>fold 7</i>	36
15	Fluktuasi nilai APE model ANN-GA pada tiap <i>fold</i>	39
16	Peubah penjelas paling berpengaruh berdasarkan PFI	40
17	<i>Boxplot</i> distribusi nilai APE ANN-GA dan regresi linier berganda	43

DAFTAR LAMPIRAN

1	Hasil prediksi regresi linier berganda dengan transformasi logaritma natural dan nilai APE pada tiap <i>fold</i>	50
2	Hasil diagnosis sisaan model regresi linier berganda tanpa transformasi	51
3	Hasil diagnosis sisaan model regresi linier berganda dengan transformasi logaritma natural	52
4	Nilai <i>JS Distance</i> , selisih Spearman, dan skor komposit dari skenario terbaik pada tiap <i>fold</i>	53
5	Kromosom terbaik <i>output</i> ANN-GA di setiap <i>fold</i>	54
6	Hasil prediksi ANN-GA dan nilai APE pada tiap <i>fold</i>	55