



Sistem Prediksi Produksi Kelapa Sawit Berbasis Gradio Menggunakan Algoritma Regresi Linear Berganda

Irfan Jamal Matondang, Elvia Budianita*, Fadhilah Syafria, Iis Afrianty

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹12150113804@students.uin-suska.ac.id, ^{2,*}elvia.budianita@uin-suska.ac.id, ³fadhilah.syafria@uin-suska.ac.id, ⁴iis.afrianty@uin-suska.ac.id

Email Penulis Korespondensi: elvia.budianita@uin-suska.ac.id

Abstrak—Ketidakstabilan produksi kelapa sawit sering menyebabkan ketidaksesuaian antara target dan realisasi produksi, sehingga diperlukan suatu model prediksi yang akurat untuk mendukung perencanaan operasional. Penelitian ini bertujuan untuk mengembangkan model prediksi produksi kelapa sawit serta mengidentifikasi variabel yang paling berpengaruh terhadap hasil produksi sebagai dasar pengambilan keputusan berbasis data. Model dibangun menggunakan metode Regresi Linear Berganda berdasarkan data historis periode 2020–2024 yang terdiri atas 60 data bulanan dengan variabel jumlah pokok, luas lahan, curah hujan, jumlah tandan, dan umur tanaman. Tahapan penelitian meliputi preprocessing data, pemilihan fitur melalui pengujian beberapa variabel independen, pemodelan, serta evaluasi kinerja menggunakan koefisien determinasi (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Square Error (RMSE). Hasil penelitian menunjukkan bahwa hasil pemilihan fitur berupa variabel jumlah pokok, luas lahan, jumlah tandan, dan umur tanaman menghasilkan performa terbaik dengan nilai R^2 sebesar 0,85 pada data training dan 0,81 pada data testing. Nilai MAE masing-masing sebesar 125.307 kg dan 176.984 kg, MSE sebesar 28.870.838.455 kg² dan 52.809.954.662 kg², serta RMSE sebesar 169.914 kg dan 229.804 kg. Berdasarkan nilai koefisien regresi, variabel jumlah tandan memiliki pengaruh paling dominan terhadap produksi kelapa sawit dengan nilai koefisien sebesar 637.720 kg. Model selanjutnya diimplementasikan menggunakan library Python Gradio dalam bentuk antarmuka interaktif guna mendukung efektivitas perencanaan produksi dan meminimalkan risiko ketidaktepatan dalam pengambilan keputusan di lingkungan perkebunan kelapa sawit.

Kata Kunci: Regresi Linear Berganda; Prediksi Produksi; Kelapa Sawit; Evaluasi Model; Gradio

Abstract—Fluctuations in oil palm production often lead to discrepancies between production targets and actual outcomes, thereby necessitating an accurate prediction model to support operational planning. This study aims to develop an oil palm production prediction model and to identify the most influential variables affecting production as a basis for data-driven decision making. The model was developed using the Multiple Linear Regression method based on historical data from 2020–2024, consisting of 60 monthly observations with variables including number of trees, land area, rainfall, number of fruit bunches, and plant age. The research stages comprise data preprocessing, feature selection through testing several independent variables, modeling, and performance evaluation using the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). The results indicate that the selected features number of trees, land area, number of fruit bunches, and plant age produce the best performance, achieving an R^2 value of 0.85 on the training data and 0.81 on the testing data. The MAE values are 125,307 kg and 176,984 kg, the MSE values are 28,870,838,455 kg² and 52,809,954,662 kg², and the RMSE values are 169,914 kg and 229,804 kg, respectively. Based on the regression coefficients, the number of fruit bunches is identified as the most dominant variable influencing oil palm production, with a coefficient value of 637,720 kg. The model is further implemented using the Python Gradio library in the form of an interactive interface to enhance production planning effectiveness and minimize the risk of inaccurate decision making in oil palm plantation environments.

Keywords: Multiple Linear Regression; Production Prediction; Oil Palm; Model Evaluation; Gradio

1. PENDAHULUAN

Kelapa sawit merupakan salah satu komoditas unggulan di sektor perkebunan Indonesia yang memberikan kontribusi besar terhadap perekonomian nasional, baik melalui ekspor maupun pemenuhan kebutuhan industri dalam negeri [1]. Sektor ini sangat penting bagi perekonomian Indonesia, mendatangkan pendapatan besar dan menciptakan lapangan kerja. Salah satu perusahaan besar yang bergerak di bidang perkebunan kelapa sawit adalah PT Perkebunan Nusantara IV (PTPN IV) Afdeling 4 Kebun Tamora, sebuah Badan Usaha Milik Negara (BUMN) yang beroperasi di Provinsi Riau. Sebagai perusahaan yang memiliki peran penting dalam industri perkebunan, PTPN IV dituntut untuk terus meningkatkan produktivitasnya [2]. Upaya tersebut memerlukan pengelolaan operasional dan perencanaan strategis yang efektif [3].

Melalui analisis situasi yang dilakukan, teridentifikasi adanya kesenjangan antara kondisi aktual produksi kelapa sawit dengan kondisi ideal yang diharapkan, sehingga menjadi hambatan dalam peningkatan produksi. Selain itu, berbagai variabel seperti jumlah pokok, luas lahan, tandan, dan umur tanaman turut menyebabkan ketidakstabilan hasil panen, yang menegaskan perlunya prediksi yang lebih akurat [4]. Untuk mengatasi permasalahan tersebut, diperlukan sebuah sistem yang mampu menghasilkan prediksi produksi secara tepat berdasarkan data historis, sehingga dapat mendukung pengambilan keputusan dalam manajemen perkebunan [5]. Oleh karena itu, diperlukan penerapan teknik data mining untuk mengolah data historis dan menemukan pola tersembunyi yang dapat digunakan dalam membangun model prediksi yang andal [6].

Data mining merupakan proses penggalian informasi atau pola yang bermakna dari kumpulan data besar [7]. Salah satu teknik penting dalam data mining adalah prediksi, yaitu proses memperkirakan nilai suatu variabel berdasarkan pola data masa lalu [8]. Untuk mendukung proses prediksi tersebut, salah satu pendekatan yang dapat digunakan adalah algoritma Regresi Linear Berganda. Algoritma ini merupakan metode yang digunakan untuk memodelkan hubungan



antara satu variabel dependen dan dua atau lebih variabel independen [9]. Setelah itu, implementasi sistem prediksi dilakukan menggunakan Gradio sebagai antarmuka interaktif yang memudahkan pengguna dalam memasukkan data, menjalankan model, serta memperoleh hasil prediksi secara real-time [10].

Beberapa penelitian sebelumnya yang berkaitan dengan prediksi produksi kelapa sawit dan regresi linear berganda, yaitu penelitian [3] menggunakan metode Regresi Linier Berganda untuk memprediksi produksi kelapa sawit di Kabupaten Aceh Utara, menghasilkan prediksi yang cukup baik dengan MAPE sebesar 14,28%. Penelitian [4] menerapkan regresi linear berganda untuk memprediksi panen kelapa sawit di PTPN XIII Kebun Rimba Belian, dengan hasil ditunjukkan oleh RMSE 0.0698 dan R^2 sebesar 0.9306. Selanjutnya, penelitian [11] menggunakan metode Regresi Linear Berganda dan analisis time series untuk prediksi jumlah produksi kelapa sawit di PT. Surya Argolika Reksa, dengan hasil metrik error MAE sekitar 35.000 kg dan RMSE sekitar 45.791,95.

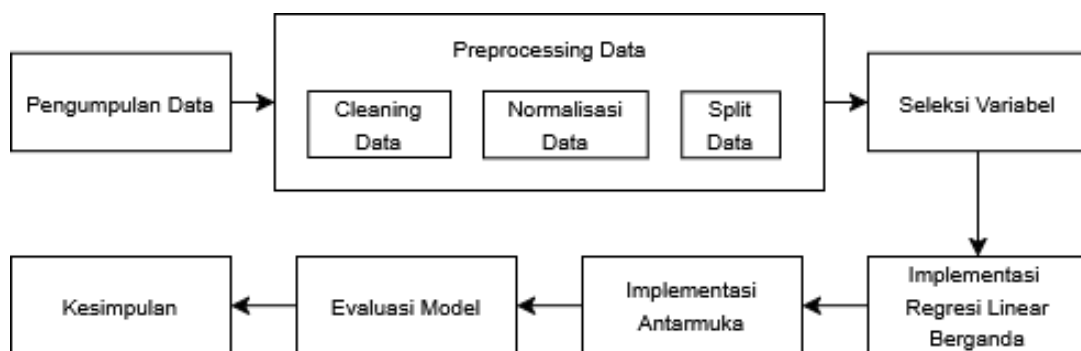
Beberapa penelitian sebelumnya telah mengembangkan model prediksi produksi kelapa sawit, sebagaimana ditunjukkan pada penelitian [3], [4], dan [11]. Namun, sebagian besar penelitian tersebut masih berfokus pada tahap pemodelan dan evaluasi kinerja algoritma tanpa mengintegrasikan model ke dalam sistem antarmuka yang dapat digunakan secara langsung oleh pihak operasional. Padahal, bagi PTPN IV sebagai perusahaan perkebunan, kebutuhan akan sistem prediksi yang tidak hanya akurat tetapi juga mudah diakses dan digunakan secara praktis menjadi sangat penting untuk mendukung perencanaan produksi, serta pengambilan keputusan yang berbasis data.

Kebutuhan terhadap pengembangan antarmuka interaktif terletak pada kemampuannya menjembatani hasil analisis model dengan kebutuhan pengguna non-teknis di lapangan. Tanpa antarmuka, model prediksi hanya dapat dijalankan melalui lingkungan pemrograman tertentu, sehingga kurang efisien dan tidak fleksibel untuk penggunaan operasional harian. Oleh karena itu, penelitian ini tidak hanya membangun model regresi linear berganda, tetapi juga mengintegrasikannya ke dalam antarmuka interaktif berbasis Gradio yang memungkinkan pengguna memasukkan variabel produksi secara langsung dan memperoleh hasil prediksi secara real-time. Penelitian ini terletak pada integrasi antara pengembangan model regresi dan implementasi sistem antarmuka interaktif. Oleh karena itu, penelitian ini tidak hanya membangun model regresi linear berganda dan mengintegrasikannya ke dalam antarmuka interaktif berbasis Gradio, tetapi juga melakukan pengujian terhadap pemilihan fitur melalui pengujian beberapa variabel independen, yaitu jumlah pokok, luas lahan, curah hujan, jumlah tandan, dan umur tanaman.

Penelitian ini diharapkan dapat mengembangkan model prediksi produksi kelapa sawit yang akurat sehingga dapat dijadikan dasar dalam mendukung perencanaan operasional dan pengambilan keputusan di lingkungan PTPN IV, khususnya di Afdeling 4 Kebun Tamora. Dengan memanfaatkan pendekatan data mining dan algoritma regresi linear berganda, penelitian ini juga mengimplementasikan antarmuka sistem prediksi berbasis *Gradio* yang memungkinkan pengguna berinteraksi langsung dengan model melalui tampilan yang sederhana dan mudah digunakan. Dengan demikian, perusahaan diharapkan dapat meningkatkan efektivitas perencanaan produksi serta meminimalkan risiko ketidaktepatan dalam pengambilan keputusan.

2. METODOLOGI PENELITIAN

Berikut pada Gambar 1 merupakan tahapan dari penelitian.



Gambar 1 Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian yang digunakan dalam pengembangan sistem prediksi produksi kelapa sawit. Tahapan penelitian diawali pengumpulan data historis produksi kelapa sawit yang digunakan sebagai dasar dalam proses analisis. Data yang telah dikumpulkan kemudian melalui tahap *preprocessing* yang meliputi *data cleaning* dan normalisasi data untuk memastikan kualitas serta keseragaman skala data. Setelah tahap *preprocessing*, dilakukan tahap seleksi variabel untuk menentukan kombinasi variabel independen yang paling berpengaruh terhadap variabel produksi. Seleksi variabel dilakukan dengan membandingkan nilai koefisien determinasi (R^2) pada data training dan data testing model Regresi Linear Berganda. Variabel terpilih selanjutnya digunakan pada tahap pemodelan menggunakan metode Regresi Linear Berganda. Model yang dihasilkan kemudian dievaluasi menggunakan metrik R^2 , MAE, MSE, dan RMSE untuk menilai tingkat akurasi dan kemampuan generalisasi model. Tahap akhir dari penelitian ini adalah implementasi model ke dalam sistem prediksi berbasis antarmuka interaktif menggunakan Library Python Gradio.



2.1 Pengumpulan Data

Dataset yang digunakan pada penelitian ini diperoleh secara langsung dari PTPN IV Regional 3, Kebun Tamora Afdeling 4. Data yang digunakan merupakan data produksi kelapa sawit bulanan selama periode 2020 hingga 2024, dengan total 60 entri data historis beserta variabel-variabel yang memengaruhinya. Variabel *independen* yang digunakan meliputi luas lahan, jumlah pokok, curah hujan, jumlah tandan, dan umur tanaman, sedangkan variabel *dependen* (target) adalah hasil produksi kelapa sawit. Seluruh data dihimpun dalam format Excel dan merepresentasikan kondisi aktual di lapangan. Berikut disajikan sampel dataset yang digunakan dalam penelitian ini:

Tabel 1. Dataset penelitian

No	jumlah_pokok (X1)	luas_lahan (X2)	Curah_hujan (X3)	Tandan (X4)	umur_tanaman (X5)	Produksi (Y)
1	129109	3016,52	107	217.055	6	2032380
2	129109	2855,93	185	217.785	6	2116340
3	129109	3016,52	152	296.000	6	3088930
4	129109	3096,02	224	274.850	6	2846840
5	129109	2962,41	205	282.077	6	2793090
...
60	127176	3380,64	254	155.622	10	2369220

2.2 Preprocessing Data

Proses preprocessing pada penelitian ini dilakukan untuk memastikan bahwa data yang digunakan telah siap dan layak untuk dianalisis lebih lanjut. Preprocessing data bertujuan untuk mengubah data mentah menjadi bentuk yang lebih terstruktur dan mudah dipahami. Beberapa tahapan yang akan dilakukan di preprocessing adalah sebagai berikut:

- Cleaning Data:** Pada tahap ini dilakukan pemeriksaan data untuk mengidentifikasi adanya nilai yang hilang (*missing values*) dan data duplikat guna memastikan kualitas dataset sebelum tahap analisis lebih lanjut [11].
- Normalisasi Data:** Tahapan selanjutnya adalah penerapan *Standard Scaler*, yaitu metode normalisasi fitur yang digunakan dalam tahap pra-pemrosesan data untuk memastikan setiap variabel memiliki skala yang seragam. Normalisasi ini penting untuk mencegah fitur dengan skala yang lebih besar mendominasi proses pembelajaran model serta untuk meningkatkan kinerja [12]. Persamaan matematis dari proses normalisasi ditunjukkan sebagai berikut:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Pada persamaan 1, z merupakan nilai hasil normalisasi, x adalah nilai data awal, μ menyatakan nilai rata-rata, dan σ adalah standar deviasi. Proses ini menunjukkan di mana setiap nilai fitur dikurangi dengan nilai rata-ratanya dan dibagi dengan standar deviasi, sehingga menghasilkan data dengan rata-rata nol dan deviasi standar satu.

- Pembagian Data (Split Data):**

Pada tahap preprocessing, dataset dibagi menjadi data training dan data testing. Data training digunakan untuk membangun model regresi, sedangkan data testing digunakan untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilatih.

2.3 Seleksi Variabel

Seleksi variabel merupakan tahapan penting dalam proses pemodelan untuk menentukan kombinasi variabel independen yang paling berpengaruh terhadap variabel dependen produksi kelapa sawit. Pada penelitian ini, seleksi variabel dilakukan menggunakan pendekatan *wrapper-based feature selection*, yaitu pemilihan variabel berdasarkan evaluasi kinerja model [13]. Pendekatan ini dilakukan dengan membangun beberapa model regresi menggunakan kombinasi variabel independen yang berbeda, kemudian mengevaluasi masing-masing model berdasarkan nilai koefisien determinasi (R^2) pada data training dan data testing.

2.4 Regresi Linear Berganda

Regresi Linear Berganda adalah model regresi yang melibatkan lebih dari satu variabel bebas (independen) untuk memprediksi variabel terikat (dependen) [14]. Model ini digunakan untuk menganalisis hubungan antara satu variabel dependen dengan beberapa variabel independen yang berskala interval atau rasio [15]. Persamaan umum dari regresi linear berganda adalah:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + \dots + e \quad (2)$$

Keterangan dalam model regresi menunjukkan bahwa Y merupakan variabel terikat atau dependen yang dipengaruhi oleh X_1, X_2 , hingga X_n sebagai variabel bebas atau independen, sedangkan α adalah konstanta (*intercept*) yang menunjukkan nilai Y ketika seluruh variabel bebas bernilai nol, β_1, β_2 , hingga β_n merupakan koefisien regresi (*slope*) yang menggambarkan besarnya pengaruh masing-masing variabel independen terhadap variabel dependen, dan e menyatakan error atau residual yang merepresentasikan selisih antara nilai aktual dan nilai hasil prediksi model.



2.5 Implementasi dan Evaluasi

Dalam penelitian ini, proses implementasi mencakup penerapan algoritma regresi linear berganda untuk membangun model prediksi, serta pengembangan antarmuka menggunakan *Library Python Gradio* sebagai media interaksi pengguna dalam melakukan input data dan memperoleh hasil prediksi. Gradio merupakan *open-source Python package* yang digunakan sebagai alat pembangun antarmuka (*user interface*) untuk model *machine learning* [10] Gradio memungkinkan pembuatan antarmuka grafis interaktif yang dapat diakses melalui peramban web menggunakan sebuah URL. Seluruh proses penelitian diimplementasikan menggunakan bahasa pemrograman *Python* pada platform Google Colab guna memastikan proses pelatihan dan pengujian model berjalan secara efektif dan optimal.

Tahap selanjutnya adalah evaluasi model yang bertujuan untuk menilai kemampuan model dalam memprediksi variabel dependen melalui penggunaan metrik evaluasi yang merepresentasikan tingkat kesalahan dan kemampuan penjelasan variasi data. Berikut merupakan metrik evaluasi yang akan digunakan:

a. Koefisien Determinasi (R^2)

Koefisien determinasi (R^2) adalah ukuran statistik yang menunjukkan seberapa baik model regresi menjelaskan varians dari variabel terikat. Nilai R^2 berkisar antara 0 dan 1, di mana semakin mendekati 1, semakin kuat hubungan antara variabel bebas dan variabel terikat, serta semakin baik model dalam menjelaskan data [16].

$$R^2 = \frac{b_1(\sum x_1y) + b_2(\sum x_2y) + \dots + b_n(\sum x_ny)}{\sum y^2} \quad (3)$$

b. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah ukuran statistik yang digunakan untuk mengukur keakuratan model prediksi dengan menghitung rata-rata dari nilai absolut selisih antara nilai aktual dan nilai prediksi. Semakin kecil nilai MAE, semakin baik model tersebut dalam melakukan prediksi [17].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

c. Mean Squared Error (MSE)

Mean Squared Error (MSE) adalah metrik evaluasi yang digunakan untuk mengukur seberapa akurat sebuah model dalam memprediksi nilai numerik [18]. MSE menghitung rata-rata kuadrat selisih antara nilai prediksi model dengan nilai sebenarnya dari data yang diamati, semakin kecil nilainya semakin baik.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

d. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) adalah ukuran statistik yang digunakan untuk menilai seberapa dekat prediksi model dengan nilai aktual. RMSE dihitung sebagai akar kuadrat dari rata-rata kuadrat dari residual (perbedaan antara nilai prediksi dan nilai aktual) [19]. Semakin kecil nilai RMSE, semakin akurat prediksi model tersebut dalam memodelkan data.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini terdiri dari 60 entri data historis. Dataset tersebut mencakup variabel independen berupa jumlah pokok, luas lahan, curah hujan, jumlah tandan, dan umur tanaman, yang digunakan untuk memodelkan variabel dependen produksi kelapa sawit. Seluruh variabel dipilih berdasarkan relevansinya terhadap proses produksi dan ketersediaan data yang mendukung analisis.

3.2 Preprocessing Data

- Cleaning Data* : Pada tahap ini dilakukan pemeriksaan kualitas data untuk memastikan tidak terdapat data duplikat maupun nilai yang hilang (*missing values*). Berdasarkan hasil pemeriksaan, dataset dinyatakan bersih sehingga proses *cleaning data* dapat diselesaikan tanpa memerlukan penghapusan maupun penambahan data.
- Normalisasi Data: Tahap selanjutnya dilakukan normalisasi data menggunakan metode *Standard Scaler* untuk menyamakan skala setiap fitur sehingga data memiliki nilai rata-rata nol dan deviasi standar satu. Proses ini diterapkan pada variabel independen guna mencegah perbedaan skala antar variabel memengaruhi kinerja model. Hasil dari normalisasi data dapat dilihat pada Tabel 2 berikut:

Tabel 2. Hasil Normalisasi Data



jumlah pokok	luas lahan	Curah hujan	tandan	Umur tanaman
1.221374	-0.802701	-1.097310	0.505634	-1.414214
1.221374	-1.185046	0.029618	0.519919	-1.414214
1.221374	-0.802701	-0.447159	2.050409	-1.414214
1.221374	-0.613421	0.593082	1.636551	-1.414214
1.221374	-0.931530	0.318574	1.777967	-1.414214

Nilai hasil normalisasi menunjukkan posisi data terhadap nilai rata-rata dalam satuan standar deviasi, di mana nilai positif menandakan data berada di atas rata-rata dan nilai negatif berada di bawah rata-rata. Transformasi ini menghilangkan satuan asli data sehingga seluruh variabel berada pada skala yang sebanding, sebagaimana ditunjukkan pada Gambar 2. Kondisi ini menandakan bahwa proses normalisasi telah berhasil dilakukan dan data siap digunakan pada tahap pemodelan selanjutnya.

c. Pembagian Data (Split Data)

Dataset dalam penelitian ini dibagi menjadi dua subset, yaitu data training dan data testing. Data training digunakan untuk melatih model, sedangkan data testing digunakan untuk mengevaluasi kinerja model terhadap data yang belum pernah dilihat sebelumnya. Pembagian dataset tersebut disajikan pada Tabel 3 berikut.

Tabel 3. Pembagian Dataset

No	Pembagian Data	
	Data Training	Data Testing
1	90%	10%
2	80%	20%
3	70%	30%

3.4 Seleksi Variabel

Proses seleksi variabel pada penelitian ini dilakukan menggunakan pendekatan eksperimental berbasis evaluasi dengan metode *wrapper-based feature selection* [13]. Pendekatan ini memilih kombinasi variabel independen berdasarkan evaluasi masing-masing model berdasarkan nilai koefisien determinasi (R^2) pada data training dan data testing.

Seleksi variabel dilakukan dengan membandingkan nilai koefisien determinasi (R^2) pada data training dan data testing dari berbagai kombinasi variabel independen. Nilai R^2 digunakan sebagai indikator utama karena mampu menunjukkan sejauh mana model dapat menjelaskan variasi data produksi sekaligus menggambarkan kemampuan generalisasi model terhadap data yang belum pernah dilatih. Pendekatan ini bertujuan untuk memperoleh kombinasi variabel yang menghasilkan kinerja prediksi optimal serta meminimalkan risiko *overfitting*.

Untuk mendukung evaluasi yang lebih objektif, proses seleksi variabel dilakukan menggunakan tiga skenario pembagian dataset, yaitu 90:10, 80:20, dan 70:30. Setiap skenario digunakan untuk menguji stabilitas performa model terhadap perubahan proporsi data training dan data testing. Kombinasi variabel yang menghasilkan nilai R^2 training dan testing yang tinggi serta relatif seimbang dipilih sebagai model terbaik.

Pendekatan seleksi variabel berbasis evaluasi kinerja model ini telah banyak digunakan pada penelitian regresi dan data mining, khususnya dalam konteks prediksi produksi dan analisis regresi linear berganda, karena mampu memberikan hasil yang lebih representatif terhadap kondisi data aktual.

Tabel 4. Seleksi Variabel dengan Pembagian Dataset 90:10

Seleksi Variabel (Dataset 90:10)						
jumlah pokok	luas lahan	curah hujan	tandan	umur tanaman	R^2 Train	R^2 Test
✓	✓	✓	✓	✓	85%	39%
✓	✓	✓	✓	✗	64%	54%
✓	✓	✓	✗	✓	14%	3%
✓	✓	✗	✓	✓	85%	37%
✓	✗	✓	✓	✓	84%	33%
✗	✓	✓	✓	✓	85%	39%

Berdasarkan hasil pada Tabel 4 dengan pembagian dataset 90:10, sebagian besar kombinasi variabel menghasilkan nilai R^2 training yang tinggi, namun nilai R^2 testing relatif rendah dan bervariasi. Kondisi ini menunjukkan bahwa meskipun model mampu menjelaskan data pelatihan dengan baik, kemampuan generalisasinya masih terbatas. Kombinasi variabel tanpa umur tanaman memberikan nilai R^2 testing tertinggi sebesar 54%, namun selisih antara R^2 training dan testing masih cukup besar, yang mengindikasikan potensi *overfitting*.

Tabel 5. Seleksi Variabel dengan Pembagian Dataset 80:20

Seleksi Variabel (Dataset 80:20)						
jumlah pokok	luas lahan	curah hujan	tandan	umur tanaman	R^2 Train	R^2 Test
✓	✓	✓	✓	✓	84%	77%



Seleksi Variabel (Dataset 80:20)							
jumlah pokok	luas lahan	curah hujan	tandan	umur tanaman	R ² Train	R ² Test	
✓	✓	✓	✓	✗	62%	55%	
✓	✓	✓	✗	✓	14%	24%	
✓	✓	✗	✓	✓	85%	76%	
✓	✗	✓	✓	✓	84%	75%	
✗	✓	✓	✓	✓	85%	77%	

Selanjutnya, hasil pada Tabel 5 dengan pembagian dataset 80:20 menunjukkan peningkatan nilai R² testing secara signifikan, berkisar antara 55% hingga 77%. Beberapa kombinasi variabel, termasuk kombinasi tanpa curah hujan maupun tanpa jumlah pokok, menghasilkan nilai R² testing yang relatif mendekati R² training. Hal ini menandakan bahwa peningkatan proporsi data testing memberikan gambaran kemampuan generalisasi model yang lebih baik.

Tabel 6. Seleksi Variabel dengan Pembagian Dataset 70:30

Seleksi Variabel (Dataset 70:30)							
jumlah pokok	luas lahan	curah hujan	tandan	umur tanaman	R ² Train	R ² Test	
✓	✓	✓	✓	✓	84%	81%	
✓	✓	✓	✓	✗	66%	51%	
✓	✓	✓	✗	✓	14%	3%	
✓	✓	✗	✓	✓	85%	81%	
✓	✗	✓	✓	✓	83%	82%	
✗	✓	✓	✓	✓	84%	82%	

Hasil terbaik diperoleh pada Tabel 6 dengan pembagian dataset 70:30, di mana beberapa kombinasi variabel menunjukkan nilai R² training dan testing yang tinggi serta relatif seimbang. Secara khusus, kombinasi variabel jumlah pokok, luas lahan, jumlah tandan, dan umur tanaman tanpa melibatkan variabel curah hujan menghasilkan nilai R² sebesar 85% pada data training dan 81% pada data testing. Nilai tersebut menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data sekaligus memiliki kemampuan generalisasi yang stabil.

3.5 Implementasi Regresi Linear Berganda

Setelah melalui beberapa tahapan pengolahan dan seleksi variabel yang akan digunakan, langkah selanjutnya adalah melakukan proses pembelajaran menggunakan metode Regresi Linear Berganda. Berdasarkan hasil dari model yang dibangun, diperoleh nilai intercept (β_0) sebesar 2.300.025 dengan koefisien β_1 untuk variabel jumlah pokok sebesar -38.147, koefisien β_2 untuk luas lahan sebesar -68.870, koefisien β_3 untuk tandan sebesar 637.720, dan koefisien β_4 untuk umur tanaman sebesar 381.442. Berikut merupakan hasil persamaan dari model yang telah dibangun:

$$Y = 2.300.025 + (-38.147) X_1 + (-68.870) X_2 + (637.720) X_3 + (381.442) X_4$$

Model regresi linear berganda tersebut selanjutnya diimplementasikan pada data testing, dan sampel hasilnya ditunjukkan pada Tabel 7 berikut:

Tabel 7. Hasil Presiksi

No	Aktual (kg)	Prediksi (kg)
0	2.032.380	2.091.729
5	2.187.220	2.654.081
36	1.828.500	2.065.183
45	2.761.550	2.444.930
13	1.876.150	1.758.944

Hasil pengujian model pada data *testing* menunjukkan bahwa nilai prediksi yang dihasilkan berada pada kisaran yang mendekati nilai aktual. Hal ini mengindikasikan bahwa model regresi linear berganda yang dibangun mampu merepresentasikan hubungan antara variabel independen dan variabel dependen dengan baik. Kesesuaian antara nilai aktual dan nilai prediksi tersebut menunjukkan bahwa model memiliki kinerja yang stabil dalam melakukan estimasi pada data yang belum digunakan dalam proses pelatihan.

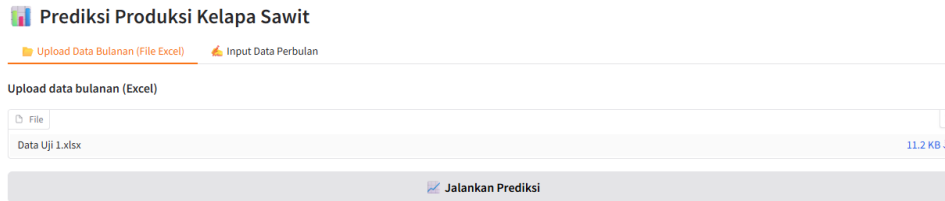
3.6 Implementasi Antarmuka

Pada tahap implementasi antarmuka, dikembangkan antarmuka pengguna menggunakan *Library Python Gradio*. Antarmuka ini dirancang dengan dua halaman utama untuk memudahkan pengguna dalam melakukan prediksi. Setiap halaman pada antarmuka tersebut memiliki fungsi yang berbeda. *Halaman Prediksi Bulanan* berfungsi untuk mengolah data dalam jumlah banyak melalui unggahan berkas Excel, sehingga pengguna dapat memperoleh hasil prediksi secara efisien dan terstruktur. Sementara itu, *Halaman Input Data Perbulan* digunakan untuk melakukan prediksi individual dengan memasukkan nilai variabel secara langsung, yang memungkinkan pengguna melakukan pengujian atau simulasi



prediksi secara cepat. Perancangan fungsi yang terpisah ini bertujuan untuk menyesuaikan kebutuhan pengguna berdasarkan jenis input data yang digunakan.

Fitur pada *Halaman Prediksi Bulanan* menggunakan mekanisme unggah berkas Excel yang terdiri atas beberapa komponen utama. Pertama, pengguna mengunggah data bulanan dalam format Excel melalui komponen unggah berkas yang telah disediakan. Setelah berkas berhasil diunggah, sistem memproses data tersebut dengan menekan tombol *Jalankan Prediksi*, yang berfungsi untuk mengeksekusi model regresi yang telah dibangun.



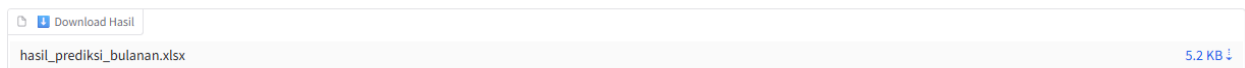
Gambar 2. Tampilan Upload Data Bulanan

Selanjutnya, sistem menampilkan tabel hasil prediksi yang memuat informasi periode, nilai produksi aktual, dan hasil prediksi produksi.

Periode	Produksi_Aktual	Prediksi_Produksi
Bulan-1	null	6.146.917,49
Bulan-2	null	6.153.156,86
Bulan-3	null	6.159.396,22
Bulan-4	null	6.165.635,59
Bulan-5	null	6.171.874,96
Bulan-6	null	6.178.114,33
Bulan-7	null	6.184.353,70
Bulan-8	null	6.190.593,07
Bulan-9	null	6.181.857,95
Bulan-10	null	6.174.370,71
Bulan-11	null	6.165.635,59
Bulan-12	null	6.159.396,22

Gambar 3. Tampilan Hasil Prediksi

Selain itu, sistem menyediakan fitur unduh hasil prediksi dalam format Excel, sehingga pengguna dapat menyimpan dan melakukan analisis lanjutan terhadap hasil yang diperoleh.



Gambar 4. Tampilan Download Hasil

Selanjutnya pada halaman Input Data Perbulan, pengguna mengisi nilai masing-masing variabel independen, seperti jumlah pokok, luas lahan, jumlah tandan, dan umur tanaman. Setelah seluruh data diinput, proses prediksi dijalankan dengan menekan tombol *Prediksi Produksi*, sehingga sistem menghasilkan nilai estimasi produksi.

Gambar 5. Tampilan Input Data Perbulan

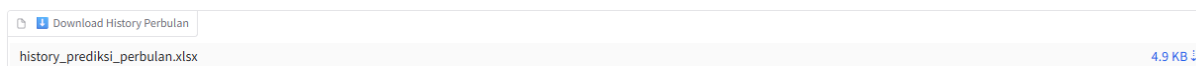
Hasil prediksi ditampilkan dalam bentuk tabel yang memuat periode prediksi dan nilai produksi hasil perhitungan model.



Periode	Produksi_Aktual	Prediksi_Produksi
Perbulan	null	3.332.177,73

Gambar 6. Tampilan Hasil Prediksi Perbulan

Selain itu, sistem juga menyediakan fitur unduh riwayat prediksi dalam format Excel, yang memungkinkan pengguna menyimpan dan meninjau kembali hasil prediksi yang telah dilakukan.



Gambar 7. Tampilan Download Histori Prediksi Perbulan

3.7 Evaluasi Model

Kinerja model dievaluasi dengan menggunakan empat metrik utama, yaitu Koefisien Determinasi (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Square Error (RMSE). Keempat metrik ini digunakan untuk memberikan penilaian yang lebih menyeluruh terhadap performa model. R^2 mengukur seberapa besar variasi data yang dapat dijelaskan oleh model, sedangkan MAE menilai rata-rata kesalahan absolut dari hasil prediksi. MSE dan RMSE digunakan untuk mengukur tingkat kesalahan berbasis kuadrat, dengan RMSE memberikan interpretasi yang lebih intuitif karena menggunakan satuan yang sama dengan data aslinya. Berikut merupakan hasil nilai dari R^2 , MAE, MSE, dan RMSE:

Tabel 8. Hasil Evaluasi

Model	R^2	MAE (kg)	MSE (kg^2)	RMSE (kg)
Training	0,85	125.307	28.870.838.455	169.914
Testing	0,81	176.984	52.809.954.662	229.804

Tabel 8 menyajikan hasil evaluasi kinerja model regresi linear berganda pada data training dan testing menggunakan empat metrik, yaitu Koefisien Determinasi (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Square Error (RMSE). Berdasarkan hasil tersebut, nilai R^2 pada data training sebesar 0,85 menunjukkan bahwa 85% variasi produksi kelapa sawit dapat dijelaskan oleh variabel independen yang digunakan dalam model, sedangkan sisanya sebesar 15% dipengaruhi oleh faktor lain di luar model. Pada data testing, nilai R^2 sebesar 0,81 mengindikasikan bahwa model masih mampu menjelaskan 81% variasi data yang belum pernah dilatih sebelumnya. Nilai ini menunjukkan bahwa model memiliki tingkat akurasi dan konsistensi yang baik.

Nilai MAE pada data training sebesar 125.307 kg menunjukkan bahwa rata-rata selisih absolut antara nilai prediksi dan nilai aktual adalah sekitar 125.307 kg. Pada data testing, MAE meningkat menjadi 176.984 kg, yang berarti terjadi kenaikan rata-rata kesalahan prediksi ketika model diuji pada data baru. Nilai MSE pada data training sebesar 28.870.838.455 kg^2 dan pada data testing sebesar 52.809.954.662 kg^2 menunjukkan besarnya rata-rata kesalahan kuadrat prediksi. Sementara itu, nilai RMSE sebesar 169.914 kg (training) dan 229.804 kg (testing) memberikan gambaran tingkat kesalahan dalam satuan kilogram sehingga lebih mudah diinterpretasikan. Meskipun terjadi peningkatan nilai kesalahan pada data testing, selisihnya masih dalam batas yang wajar. Hal ini menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup baik dan tidak mengalami overfitting secara signifikan.

4. KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sebuah sistem prediksi produksi kelapa sawit menggunakan metode Regresi Linear Berganda berdasarkan data historis periode 2020–2024. Pengembangan sistem diawali dengan tahapan *preprocessing* data yang meliputi *data cleaning*, normalisasi data, serta seleksi variabel untuk memastikan kualitas data yang digunakan dalam proses pemodelan. Berdasarkan hasil seleksi variabel, kombinasi variabel jumlah pokok, luas lahan, jumlah tandan, dan umur tanaman menghasilkan kinerja model prediksi terbaik. Hasil evaluasi menunjukkan bahwa model memiliki nilai koefisien determinasi (R^2) sebesar 0,85 pada data training dan 0,81 pada data testing, yang menandakan bahwa model mampu menjelaskan sebagian besar variasi data produksi dengan baik. Nilai Mean Absolute Error (MAE) sebesar 125.307 kg pada data training dan 176.984 kg pada data testing menunjukkan bahwa rata-rata kesalahan prediksi relatif rendah dan tidak menunjukkan perbedaan yang signifikan. Sementara itu, nilai Mean Squared Error (MSE) dan Root Mean Squared Error (RMSE) pada data testing mengalami peningkatan dibandingkan data training, namun masih berada dalam batas yang wajar, sehingga dapat disimpulkan bahwa model memiliki kemampuan generalisasi yang baik dan tidak mengalami *overfitting* secara signifikan. Model prediksi yang telah dibangun selanjutnya diimplementasikan ke dalam sebuah sistem berbasis antarmuka interaktif menggunakan Library Python Gradio. Sistem ini menyediakan dua fitur utama, yaitu prediksi produksi secara individu melalui input manual dan



prediksi produksi secara bulanan melalui unggahan data dalam format Excel. Dengan adanya antarmuka ini, pengguna dapat melakukan proses prediksi secara lebih mudah, cepat, dan terstruktur, sehingga sistem dapat dimanfaatkan sebagai alat bantu dalam perencanaan produksi dan pengambilan keputusan. Meskipun sistem yang dikembangkan telah menunjukkan kinerja yang baik, penelitian ini masih memiliki beberapa keterbatasan. Keterbatasan tersebut antara lain jumlah data historis yang relatif terbatas, yaitu 60 data bulanan, serta variabel yang digunakan masih berfokus pada faktor internal kebun. Selain itu, metode Regresi Linear Berganda memiliki asumsi hubungan linear antar variabel, sehingga belum sepenuhnya mampu menangkap pola non-linear yang mungkin terdapat pada data produksi kelapa sawit. Sebagai pengembangan ke depan, sistem prediksi ini dapat ditingkatkan dengan menambahkan variabel lain yang berpotensi memengaruhi produksi, seperti data pemupukan, kondisi hama dan penyakit tanaman, serta faktor iklim yang lebih rinci. Selain itu, penelitian selanjutnya dapat mengembangkan sistem dengan membandingkan atau mengintegrasikan metode lain berbasis *machine learning* untuk meningkatkan akurasi dan fleksibilitas model prediksi.

REFERENCES

- [1] D. Yana Armanto, S. Agustian Hudjimartsu, And E. Hermawan, "Identifikasi Perhitungan Pohon Kelapa Sawit Otomatis Dengan Menggunakan Metode Convolutional Neural Network (Cnn)," *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 3, Pp. 2648–2654, May 2024, Doi: 10.36040/Jati.V8i3.9525.
- [2] L. F. Sihaloho, Z. Lubis, And B. Sibuea, "Analisis Faktor-Faktor Penentu Harga Pokok Pengolahan Produksi Di Pabrik Kelapa Sawit Sei Silau Ptpn Iv Regional 1 Palmco," *Jurnal Agrica*, Vol. 18, Pp. 225–237, 2025.
- [3] A. Prasetyo, S. Salahuddin, And A. Amirullah, "Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda," *Jurnal Infomedia*, Vol. 6, No. 2, P. 76, Dec. 2021, Doi: 10.30811/Jim.V6i2.2343.
- [4] H. Hermansyah, A. Abdullah, And P. Y. Utami, "Penerapan Metode Regresi Linier Berganda Untuk Memprediksi Panen Kelapa Sawit," *Progresif: Jurnal Ilmiah Komputer*, Vol. 20, No. 1, P. 540, Mar. 2024, Doi: 10.35889/Progresif.V20i1.1816.
- [5] F. Husaini, I. Permana, M. Afdal, And F. N. Salisah, "Penerapan Algoritma Long Short-Term Memory Untuk Prediksi Produksi Kelapa Sawit," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 4, No. 2, Pp. 366–374, Feb. 2024, Doi: 10.57152/Malcom.V4i2.1187.
- [6] J. A. Adhiva, Mustakim, S. A. Putri, And S. G. Setyorini, "Prediksi Hasil Produksi Kelapa Sawit Menggunakan Model Regresi Pada Pt. Perkebunan Nusantara V," *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (Sntiki) 12*, Pp. 155–162, 2020.
- [7] L. Setiyani, M. Wahidin, D. Awaludin, And S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes: Systematic Review," *Faktor Exacta*, Vol. 13, No. 1, P. 35, Jun. 2020, Doi: 10.30998/Faktorexacta.V13i1.5548.
- [8] D. S. O. Panggabean, E. Buulolo, And N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *Jurikom (Jurnal Riset Komputer)*, Vol. 7, No. 1, P. 56, Feb. 2020, Doi: 10.30865/Jurikom.V7i1.1947.
- [9] M. Adha, E. Utami, And H. Hanafi, "Prediksi Produksi Jagung Menggunakan Algoritma Apriori Dan Regresi Linear Berganda (Studi Kasus : Dinas Pertanian Kabupaten Dompus)," *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, Vol. 7, No. 3, Pp. 803–820, Aug. 2022, Doi: 10.29100/Jipi.V7i3.3139.
- [10] A. Abid, A. Abdalla, A. Abid, D. Khan, A. Alfazan, And J. Zou, "Gradio: Hassle-Free Sharing And Testing Of ML Models In The Wild," 2019. [Online]. Available: Arxiv:1906.02569v1
- [11] U. Rio, "Model Prediksi Jumlah Produksi Kelapa Sawit Menggunakan Regresi Linear Berganda Di Pt. Surya Argolika Reksa," *Betrik*, Vol. 16, No. 02, Pp. 122–130, 2025.
- [12] R. R. Hallan And I. N. Fajri, "Prediksi Harga Rumah Menggunakan Machine Learning Algoritma Regresi Linier," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, Vol. 7, No. 1, Pp. 57–62, Jan. 2025, Doi: 10.47233/Jteksis.V7i1.1732.
- [13] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini, And I. Budiman, "Kombinasi Seleksi Fitur Berbasis Filter Dan Wrapper Menggunakan Naive Bayes Pada Klasifikasi Penyakit Jantung," *Jurnal Teknologi Informasi Dan Ilmu Komputer*, Vol. 10, No. 6, Pp. 1361–1368, 2023.
- [14] M. L. Mu'tashim, T. Muhayat, S. A. Damayanti, H. N. Zaki, And R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," *Informatik : Jurnal Ilmu Komputer*, Vol. 17, No. 3, P. 238, Dec. 2021, Doi: 10.52958/Iftk.V17i3.3635.
- [15] Admin, "Htpps://Skillplus.Web.Id/Multiple-Linear-Regression-Pendahuluan/."
- [16] A. N. Maharadja, I. Maulana, And B. A. Dermawan, "Penerapan Metode Regresi Linear Berganda Untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi," *Journal Of Applied Informatics And Computing*, Vol. 5, No. 1, Pp. 95–102, Jul. 2021, Doi: 10.30871/Jaic.V5i1.3184.
- [17] A. T. Nurani, A. Setiawan, And B. Susanto, "Perbandingan Kinerja Regresi Decision Tree Dan Regresi Linear Berganda Untuk Prediksi Bmi Pada Dataset Asthma," *Jurnal Sains Dan Edukasi Sains*, Vol. 6, No. 1, Pp. 34–43, May 2023, Doi: 10.24246/Juses.V6i1p34-43.
- [18] H. Nuha, "Mean Squared Error (Mse) Dan Penggunaannya," *Available At Ssrn 4420880*, 2023.
- [19] N. Afrilia S, F. Frazna Az-Zahra, And P. Prajoko, "Prediksi Hasil Panen Wortel Menggunakan Algoritma Regresi Linear Berganda," *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 5, Pp. 10255–10262, Sep. 2024, Doi: 10.36040/Jati.V8i5.10954.