

Prediksi Hasil Kelapa Sawit Menggunakan Algoritma Regresi Linier Sederhana Berbasis Data Historis

Risky Diky Syahputra*, Yusuf Ramadhan Nasution

Ilmu Komputer, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1,*}syahputrarisky01@email.com, ramadhannst@uinsu.ad.id,

Email Penulis Korespondensi: syahputrarisky01@email.com*

Submitted: 16/02/2026; Accepted: 09/03/2026; Published: 31/03/2026

Abstrak– Hasil produksi kelapa sawit sering mengalami fluktuasi karena dipengaruhi oleh berbagai faktor operasional dan lingkungan, yang membuat perusahaan kesulitan dalam menyusun rencana produksi yang stabil. Penelitian ini bertujuan membangun model prediksi produksi kelapa sawit menggunakan regresi linier sederhana dengan memanfaatkan data historis produksi. Data yang digunakan mencakup 48 data bulanan dari PTPN III Rantau Prapat selama periode 2021 hingga 2024, dengan jumlah pokok panen (X) sebagai variabel bebas dan produksi Tandan Buah Segar (TBS) (Y) sebagai variabel yang diprediksi. Proses penelitian mencakup tahapan pengumpulan data, pengolahan data, pembentukan model regresi, serta pengujian kinerja model melalui MSE, RMSE, dan MAPE. Skema validasi out of sample digunakan dengan membagi data menjadi data pelatihan (30 data) dan data uji (8 data). Hasil perhitungan menunjukkan MAPE sebesar 13,62%, RMSE 368.724,126, dan MSE 135.957.481.409. Nilai R^2 sebesar 0,8316 menunjukkan hubungan linear yang cukup kuat antara variabel X dan Y. Model yang dihasilkan mudah diterapkan dan dapat digunakan sebagai alat pendukung keputusan dalam perencanaan produksi kelapa sawit, serta sebagai baseline untuk penelitian atau model peramalan yang lebih kompleks di masa depan.

Kata Kunci: Prediksi; Kelapa Sawit; Data Historis; Regresi Linier Sederhana; Produksi TBS

Abstract– Palm oil production often fluctuates due to various operational and environmental factors, making it difficult for companies to develop stable production plans. This study aims to develop a palm oil production prediction model using simple linear regression by utilizing historical production data. The data used includes 48 monthly data from PTPN III Rantau Prapat during the period 2021 to 2024, with the number of harvests (X) as the independent variable and Fresh Fruit Bunch (FFB) production (Y) as the predicted variable. The research process included data collection, data processing, regression model formation, and model performance testing through MSE, RMSE, and MAPE. An out-of-sample validation scheme was used by dividing the data into training data (30 data) and test data (8 data). The calculation results show a MAPE of 13.62%, RMSE of 368,724.126, and MSE of 135,957,481,409. The R^2 value of 0.8316 indicates a fairly strong linear relationship between variables X and Y. The resulting model is easy to apply and can be used as a decision support tool in palm oil production planning, as well as a baseline for future research or more complex forecasting models.

Keywords: Prediction; Oil Palm; historical data; Simple Linier Regression; Fresh Fruit Bunch

1. PENDAHULUAN

Kelapa sawit (*Elaeis guineensis*) merupakan tanaman yang berasal dari kawasan Afrika Barat yang beriklim tropis. Tanaman ini mulai diperkenalkan di Indonesia pada tahun 1848 melalui pemerintah kolonial Hindia Belanda yang membawa empat bibit kelapa sawit, dua berasal dari Mauritius dan dua lainnya dari Hortus Botanicus di Belanda. Bibit-bibit tersebut kemudian ditanam di Kebun Raya Bogor dan pada masa awal hanya dimanfaatkan sebagai tanaman koleksi atau penghias taman. [1]. Kelapa sawit kini menjadi komoditas penting di Indonesia, menyumbang devisa negara melalui ekspor minyak sawit dan menyediakan lapangan kerja bagi jutaan orang di berbagai daerah [2].

Sektor perkebunan serta industri pengolahan kelapa sawit merupakan salah satu penopang penting perekonomian nasional [3]. Komoditas minyak sawit yang diekspor mampu menyumbang devisa negara dalam jumlah signifikan, sekaligus menyediakan kesempatan kerja bagi jutaan tenaga kerja di berbagai daerah [4]. Peningkatan produksi minyak sawit masih sangat memungkinkan melalui penerapan praktik budidaya yang efektif dan berstandar baik. Salah satu kegiatan dalam aspek teknis budidaya yang berpengaruh terhadap pembentukan buah dan hasil minyak kelapa sawit adalah merawat pohon sawit dengan tepat [5]. Hasil utama dari perkebunan kelapa sawit dikenal sebagai Tandan Buah Segar (TBS), yaitu buah yang dipanen langsung dari tanaman dan menjadi bahan baku awal pada proses produksi di sektor hulu industri sawit [6].

Pada produksi kelapa sawit di PTPN III Rantau Prapat menunjukkan bahwa produksi TBS mengalami fluktuasi yang signifikan dan cenderung tidak stabil. Hasil panen yang tidak selalu mengikuti pola produksi yang tetap dapat menurunkan pencapaian perusahaan, atau sebaliknya, ketika hasil panen melebihi kapasitas pengolahan yang tersedia, maka pemanfaatan sumber daya menjadi kurang optimal. Pada produksi kelapa sawit di PTPN III Rantau Prapat, penurunan hasil panen tercatat hingga 15% pada beberapa periode produksi terakhir, yang berdampak pada efisiensi operasional perusahaan.

Peramalan atau prediksi merupakan suatu proses untuk memperkirakan kejadian yang akan datang dengan memanfaatkan informasi yang berasal dari peristiwa pada masa sebelumnya [7] Proses ini dilakukan melalui pengumpulan data historis yang kemudian dianalisis guna mengidentifikasi pola tertentu, sehingga dapat digunakan sebagai dasar dalam menghasilkan estimasi kondisi di masa mendatang [8]. Prediksi produksi kelapa

sawit sangat penting dalam upaya mengurangi risiko ketidakpastian yang terjadi dalam sektor perkebunan dan industri kelapa sawit. Namun, pendekatan konvensional yang umum digunakan dalam proses prediksi masih memiliki keterbatasan dalam mengidentifikasi pola data yang bersifat dinamis dan kompleks [9].

Kemajuan teknologi kecerdasan buatan yang berkembang dengan cepat menyebabkan penggunaan machine learning atau pembelajaran mesin semakin luas dalam berbagai bidang, terutama untuk membantu analisis data serta mengatasi berbagai persoalan [10]. Machine learning merupakan bagian dari ilmu komputer yang mengembangkan algoritma agar komputer dapat mengenali serta mempelajari pola dari data yang tersedia, sehingga pendekatan ini dikenal sebagai konsep pembelajaran berbasis data atau learning from data [11]. Namun, dalam penelitian ini, metode yang digunakan adalah regresi linier sederhana, yang merupakan bagian dari supervised learning berbasis regresi [12]. Dalam kecerdasan buatan, metode regresi linier sederhana dapat digunakan untuk memprediksi hasil produksi berdasarkan variabel yang memengaruhi produksi tersebut. [13]

Berbagai penelitian sebelumnya telah mengembangkan model regresi linier sederhana untuk menganalisis hubungan antara variabel bebas terhadap variabel terikat, namun belum ada penelitian yang secara spesifik mengaplikasikan model ini pada prediksi produksi TBS dengan mempertimbangkan jumlah pokok panen sebagai faktor utama dalam operasional perkebunan kelapa sawit [14]. Penelitian oleh Salsavira (2020) meramalkan *supply* bahan baku menggunakan metode regresi linier dan memperoleh nilai MAPE (*Mean Absolute Percentage Error*) sebesar 23,13% [15]. Penelitian lainnya oleh Moha Lalapa mengimplementasikan metode regresi linier sederhana untuk prediksi harga cabai rawit memperoleh nilai MAPE sebesar 24% [16].

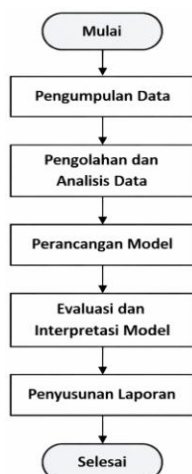
GAP penelitian ini terletak pada belum adanya sistem prediksi produksi kelapa sawit yang memanfaatkan data historis secara optimal untuk menghasilkan estimasi produksi yang akurat. Selain itu, model prediksi yang ada masih berfokus pada jumlah TBS yang masuk ke pabrik tanpa mempertimbangkan jumlah pokok panen yang lebih representatif sebagai faktor utama dalam produksi, yang membuat model tersebut tidak sepenuhnya relevan dengan kondisi operasional di lapangan. Penelitian ini bertujuan untuk membangun dan mengevaluasi model regresi linier sederhana untuk memprediksi produksi TBS di PTPN III Rantau Prapat berdasarkan jumlah pokok panen pada periode 2021–2024.

Berdasarkan permasalahan tersebut diperlukan pembangunan model yang dapat memprediksi produksi kelapa sawit dengan tingkat ketepatan dan efisiensi yang baik. Pelaksanaan penelitian dilakukan melalui langkah-langkah sistematis, dimulai dari pengumpulan data historis, penentuan variabel yang memengaruhi, hingga pengolahan data menggunakan metode regresi linier sederhana.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui pendekatan terstruktur untuk membangun model prediksi hasil panen kelapa sawit menggunakan metode regresi linier sederhana berbasis data historis produksi. Tahap penelitian dimulai dengan studi literatur untuk memperoleh landasan teori serta meninjau penelitian terdahulu yang relevan. Selanjutnya, dilakukan pengumpulan data historis hasil panen dari lokasi penelitian yang kemudian diproses melalui tahap pembersihan dan validasi data agar siap digunakan dalam pemodelan. Data yang telah diproses digunakan dalam perancangan model dengan menentukan variabel independen dan variabel dependen yang sesuai dengan kebutuhan prediksi produksi. Model regresi yang diperoleh kemudian dievaluasi menggunakan metode pengukuran kesalahan untuk mengetahui tingkat akurasi prediksi. Pada tahap akhir, penelitian difokuskan pada proses analisis hasil dan penyusunan laporan guna merangkum seluruh proses dan temuan penelitian. Diagram alur penelitian secara keseluruhan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

a. Pengumpulan Data

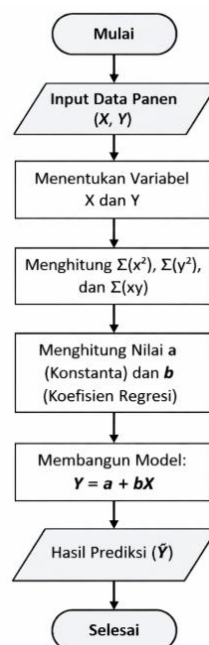
Data penelitian diperoleh melalui pengumpulan catatan produksi kelapa sawit yang terdokumentasi dalam laporan perusahaan PTPN III di wilayah Rantau Prapat selama kurun waktu 2021 sampai dengan 2024. Data yang dikumpulkan mencakup hasil panen TBS (Tandan Buah Segar) yang dicatat setiap bulan. Jumlah data yang dikumpulkan adalah 48 titik data (4 tahun x 12 bulan), yang mencakup jumlah pokok panen dan hasil produksi bulanan. Informasi tambahan juga diperoleh dari sumber sekunder, seperti artikel ilmiah dan laporan resmi yang menyajikan data produksi kelapa sawit dari berbagai wilayah yang relevan untuk mendukung kebutuhan analisis penelitian.

b. Pengolahan dan Analisis Data

Proses analisis dilakukan setelah data yang diperoleh dipastikan kelayakannya untuk digunakan dalam penelitian. Dataset yang tersedia dalam format Excel terlebih dahulu melalui tahap peninjauan dan penyesuaian untuk mengatasi permasalahan data yang tidak lengkap maupun nilai data yang menyimpang dari pola umum (outliers). Data yang telah disaring kemudian disusun kembali dalam bentuk tabel yang siap digunakan dalam proses pemodelan. Selanjutnya, data tersebut diolah menggunakan software Excel untuk menghitung dan menyusun data yang dibutuhkan untuk analisis regresi linier sederhana guna memperoleh hasil analisis yang diperlukan.

c. Perancangan Model

Dalam penelitian ini, proses perancangan model berperan sebagai dasar dalam membangun sistem prediksi produksi kelapa sawit dengan memanfaatkan data historis yang telah tersedia sebelumnya. Kegiatan perancangan mencakup penyiapan serta pengolahan data, penentuan metode analisis yang paling sesuai, yakni regresi linier sederhana yang digunakan untuk menggambarkan hubungan antar variabel, hingga pembentukan model prediksi yang mampu menghasilkan estimasi secara konsisten dan akurat.



Gambar 2. Perancangan Model

Tahapan perancangan dimulai dengan menyiapkan data yang diperlukan, kemudian memasukkan data historis berupa jumlah pokok panen (variabel X) dan produksi kelapa sawit (variabel Y) sebagai bahan analisis. Selanjutnya, dilakukan perhitungan kuadrat pada masing-masing variabel serta hasil perkalian antara variabel X dan Y sebagai dasar dalam analisis regresi. Proses berikutnya dilakukan dengan menghitung nilai konstanta (a) dan koefisien regresi (b) yang diperlukan untuk menyusun model prediksi. Rumus regresi linier yang digunakan adalah:

$$Y = a + bX \quad (1)$$

di mana:

- Y adalah variabel terikat (produksi TBS),
- X adalah variabel bebas (jumlah pokok panen),
- a adalah konstanta,
- b adalah koefisien regresi.

Model yang telah terbentuk kemudian digunakan untuk memperkirakan nilai Y berdasarkan nilai X yang tersedia. Selanjutnya, hasil prediksi ditampilkan dalam bentuk tabel atau grafik untuk memudahkan interpretasi hasil.

d. Evaluasi Dan Interpretasi Model

Evaluasi model merupakan tahapan penting dalam proses analisis statistik untuk menilai kualitas dan keakuratan model yang telah dikembangkan. Beberapa metrik evaluasi yang digunakan dalam penelitian ini adalah:

1. *Mean Square Error* (MSE) digunakan meghitung rata-rata kuadrat perbedaan antara nilai aktual dan nilai prediksi, dengan rumus sebagai berikut:

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

Keterangan:

n = Jumlah Data

y_i = Nilai actual

$\hat{y}_i$ = Nilai Prediksi

2. *Root Mean Square Error* (RMSE) digunakan untuk mengetahui tingkat kesalahan prediksi model, yang dihitung berdasarkan akar dari rata-rata kuadrat selisih antara hasil prediksi dan data aktual :

$$RMSE = \sqrt{MSE} \quad (3)$$

3. *Mean Absolute Percentage Error* (MAPE) mengukur tingkat ketepatan model prediksi dengan melihat rata-rata persentase perbedaan antara data aktual dan hasil prediksi yang dihasilkan model:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

Keterangan:

n = Jumlah Data

y_i = Nilai actual

$\hat{y}_i$ = Nilai Prediksi

e. Penyusunan Laporan

Penyusunan laporan penelitian ini bertujuan memberikan kontribusi ilmiah di bidang ilmu komputer melalui penerapan *machine learning* untuk pemodelan dan prediksi data. Selain menjawab permasalahan yang diteliti, penelitian ini juga memperluas pemahaman mengenai implementasi regresi linier sederhana dalam pengolahan data historis.

3. HASIL DAN PEMBAHASAN

Pada pembahasan ini dijelaskan data mentah yang digunakan dalam pembentukan model, termasuk proses penyajian data tersebut agar dapat diolah melalui perhitungan manual hingga menghasilkan model regresi berdasarkan data yang telah disusun.

3.1 Analisis Data

Dalam penelitian ini, analisis data dilakukan menggunakan metode regresi linier sederhana untuk memprediksi hasil panen kelapa sawit berdasarkan jumlah pokok panen. Tahap pertama mencakup perhitungan nilai perkalian antara variabel prediktor (jumlah pokok panen, X) dan variabel respon (produksi TBS, Y), serta nilai kuadrat dari variabel prediktor (X^2). Variabel X menunjukkan jumlah pokok panen, dan variabel Y menunjukkan produksi TBS (Tandan Buah Segar) yang akan diperkirakan. Selanjutnya, dilakukan perhitungan nilai intersep (a) dan koefisien regresi (b) menggunakan persamaan regresi linier sederhana.

Tabel 1. Data Mentah

NO	BULAN	TAHUN	TOTAL PUPUK (Kg)	CURAH HUJAN (MM)	JUMLAH POKOK	PRODUKSI TBS (Kg)
1	Januari	2021	0	51	672551	1750770
2	Februari	2021	0	136	644372	1523130
3	Maret	2021	64709	149	743475	1857020
...	...	...	...	...	...	...
46	Oktober	2024	0	431	1106065	3193330
47	November	2024	12661	428	1053373	3113080
48	Desember	2024	166872	307	1093069	3212050

3.2 Hasil Analisis

Sebelum model regresi disusun, terlebih dahulu dilakukan proses perhitungan nilai XY dan X² karena kedua komponen tersebut diperlukan dalam penentuan nilai intersep dan koefisien regresi. Pada tahap ini, data yang digunakan sebagai data pelatihan mencakup periode Januari 2021 hingga Juni 2024 dengan jumlah keseluruhan 30 data. Sementara itu, data pada periode Mei 2021 sampai Mei 2024 yang berjumlah 8 data dimanfaatkan sebagai data pengujian untuk mengetahui kinerja model regresi yang telah dibangun. Hasil perhitungan nilai XY dan X² pada data pelatihan kemudian ditampilkan dalam tabel berikut.

Tabel 2. Menghitung Regresi Linier Sederhana

NO	Bulan	Pokok Panen (X)	Produksi TBS (Y)	XY	X ²
1	Januari 2021	672.551	1.750.770	1.177.482.114.270	452.324.847.601
2	Februari 2021	644.372	1.523.130	981.462.324.360	415.215.274.384
3	Maret 2021	743.475	1.857.020	1.380.647.944.500	552.755.075.625
...	...	...	...	...	...
39	Maret 2024	717.719	2.390.410	1.715.642.674.790	515.120.562.961
40	April 2024	627.107	2.499.880	1.567.692.247.160	393.263.189.449
42	Juni 2024	705.661	3.025.980	2.135.316.072.780	497.957.446.921
		ΣX =	ΣY =	ΣXY =	ΣX² =
Total		19.074.020	68.643.410	43.810.891.814.210	12.328.521.319.156

- a. Mencari nilai intersep (*a*)

$$a = \bar{Y} - b\bar{X}$$

$$\bar{Y} = \frac{68.643.410}{30} = 2.288.133,666$$

$$\bar{X} = \frac{19.074.020}{30} = 635.800,666$$

$$a = 2.288.133,666 - 0,8316 \times 635.800,666$$

$$a = 1.759.352,68$$

- b. Mencari nilai koefisien (*b*)

$$b = \frac{n(\Sigma XY) - (\Sigma X)(\Sigma Y)}{n(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{30(43.810.891.814.210) - (19.074.020)(68.643.410)}{30(12.328.521.319.156) - (19.074.020)^2}$$

$$b = \frac{1.314.326.754.426.300 - 1.309.305.775.208.200}{369.855.639.574.680 - 363.818.254.960.400}$$

$$b = \frac{5.020.979.218.100}{6.037.384.614.280}$$

$$b = 0,8316$$

- c. Menentukan persamaan regresi

Dengan menggunakan nilai intersep (*a*) dan koefisien regresi (*b*) yang telah diperoleh, persamaan regresi linier yang dibentuk adalah sebagai berikut:

$$Y = a + bX$$

$$Y = 1.759.352,68 + 0,8316X$$

- d. Menghitung Peramalan

Setelah model regresi terbentuk, digunakan untuk memprediksi produksi TBS pada periode tertentu. Data bulan Mei 2021 digunakan sebagai data uji untuk mengevaluasi kemampuan model dalam memprediksi hasil produksi. Berikut adalah hasil peramalan produksi TBS :

Tabel 3. Hasil Peramalan

No	Bulan	Produksi TBS (Y)
5	Mei 2021	2.329.951,55
7	Juli 2021	2.298. .942,80
15	Maret 2022	2.297.361,01
29	Mei 2023	2.388.547,66
30	Juni 2023	2.378.319,24
34	Oktober 2023	2.333.778,79
38	Februari 2024	2.253.768,63
41	Mei 2024	2.371.622,83

e. Evaluasi Model

Untuk mengukur kinerja model secara kuantitatif, digunakan tiga metrik utama: MSE, RMSE dan MAPE.

1. MSE (*Mean Squared Error*)

MSE digunakan untuk mengetahui seberapa besar penyimpangan hasil prediksi model regresi terhadap nilai sebenarnya, diperoleh dari perhitungan rata-rata kuadrat perbedaan antara nilai aktual dan nilai prediksi. [17].

$$MSE = \frac{\sum y_1^2}{n}$$

$$MSE = \frac{1.087.659.851.272}{8}$$

$$MSE = 135.957.481.409$$

2. RMSE (*Root Mean Squared Error*)

RMSE merupakan ukuran yang digunakan untuk mengetahui tingkat kesalahan prediksi model, yang dihitung berdasarkan akar dari rata-rata kuadrat selisih antara hasil prediksi dan data aktual.

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{135.957.481.409}$$

$$RMSE = 368.724,126$$

3. MAPE (*Mean Absolute Percentage Error*)

MAPE digunakan untuk mengukur tingkat ketepatan model prediksi dengan melihat rata-rata persentase perbedaan antara data aktual dan hasil prediksi yang dihasilkan model. [18].

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$MAPE = \frac{100\%}{8} \times 1,0893$$

$$MAPE = 13,62\%$$

d. Interpretasi Hasil Evaluasi

Nilai MAPE sebesar 13,62% mengindikasikan bahwa model memiliki tingkat akurasi yang baik dalam memprediksi hasil produksi TBS. Berdasarkan klasifikasi akurasi peramalan, nilai MAPE ini tergolong dalam kategori baik.

3.3 Membangun Model Regresi

Tahap awal dalam membangun model regresi dimulai dengan proses *preprocessing* data. Kegiatan ini mencakup pembersihan data, penanganan nilai yang tidak lengkap, serta identifikasi dan penyesuaian data ekstrem atau *outlier*. Hubungan antar variabel dalam data juga dianalisis melalui matriks korelasi guna mengetahui keterkaitan di antara variabel yang digunakan. Setelah proses persiapan data selesai, dataset dimanfaatkan untuk membangun model regresi yang kemudian digunakan dalam proses prediksi terhadap data pengujian. Selanjutnya, performa model dievaluasi dengan menghitung nilai kesalahan menggunakan indikator *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE) dan MAPE (*Mean Absolute Percentage Error*).

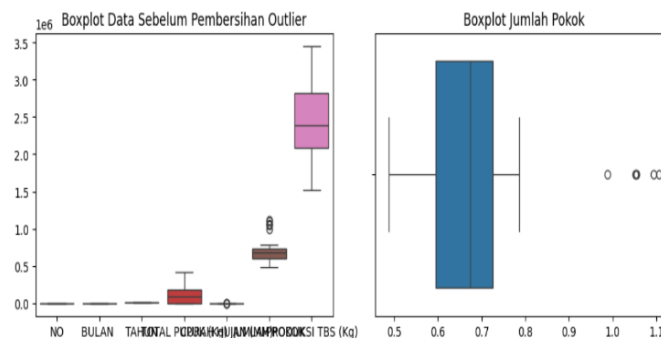
a. Preprocessing Data

Sebelum data digunakan dalam proses pembentukan model, terlebih dahulu dilakukan evaluasi terhadap data mentah yang telah dikumpulkan. Evaluasi ini mencakup pemeriksaan struktur data, identifikasi nilai kosong, pengecekan nilai unik, serta pendeteksian kemungkinan data duplikat. Langkah tersebut bertujuan menjaga kualitas data agar hasil prediksi model menjadi lebih optimal.

```
*** <class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   NO                     48 non-null     int64
1   BULAN                  48 non-null     int64
2   TAHUN                  48 non-null     int64
3   TOTAL PUPUK (Kg)      48 non-null     int64
4   CURAH HUJAN (MM)      48 non-null     int64
5   JUMLAH POKOK          48 non-null     int64
6   PRODUKSI TBS (Kg)     48 non-null     int64
dtypes: int64(7)
```

Gambar 3. Info Data Mentah

Berdasarkan Gambar 3, data penelitian memiliki beberapa fitur dengan tipe numerik, yaitu No, Bulan, Tahun, Total Pupuk, Curah Hujan, Jumlah Pokok, dan Produksi TBS. Selanjutnya, seluruh fitur numerik tersebut dianalisis untuk melihat keberadaan data yang menyimpang atau data *Outlier* yang sebagaimana ditampilkan pada gambar berikut.



Gambar 4. Data Sebelum Dilakukan Pembersihan *Outlier*

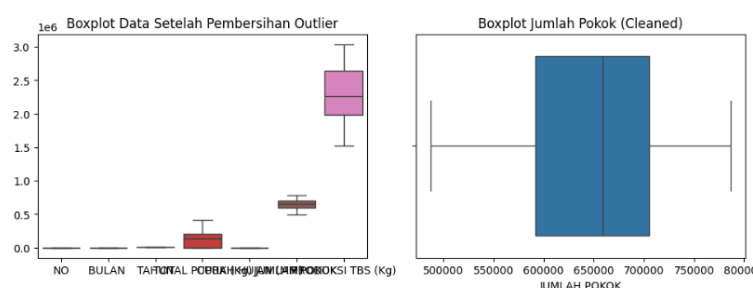
Hasil analisis menunjukkan adanya sejumlah data *Outlier* yang muncul pada beberapa fitur data. Proses identifikasi outlier pada produksi TBS dilakukan dengan metode IQR sesuai penjelasan pada perancangan, dan hasilnya dapat dilihat pada gambar berikut.

	Feature	Q1	Q3	IQR	Lower Bound	Upper Bound	Outlier Count
0	NO	12.75	36.25	23.50	-22.500	71.500	0
1	BULAN	2.75	8.25	5.50	-5.500	16.500	0
2	TAHUN	2021.75	2023.25	1.50	2019.500	2025.500	0
3	TOTAL PUPUK (Kg)	0.00	183852.75	183852.75	-275779.125	459631.875	0
4	CURAH HUJAN (MM)	193.25	336.50	143.25	-21.625	551.375	4
5	JUMLAH POKOK	594657.50	725034.75	130377.25	399091.625	920600.625	6
6	PRODUKSI TBS (Kg)	2081677.50	2823745.00	742067.50	968576.250	3936846.250	0

Gambar 5. Metode IQR

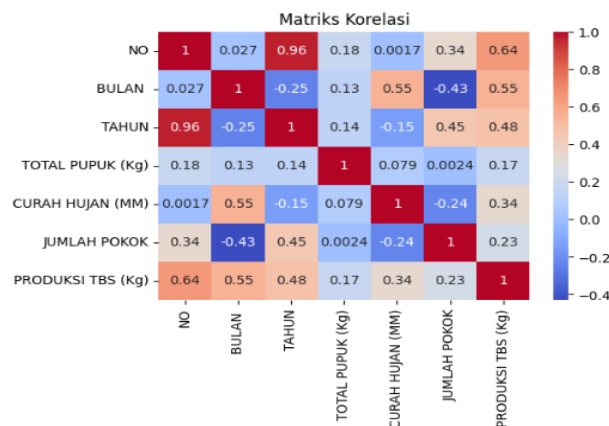
```
*** Total outliers: 10
```

Gambar 6. Jumlah Data *Outlier*



Gambar 7. Data Setelah Dilakukan Pembersihan *Outlier*

Ditemukan sebanyak 10 data yang tergolong *outlier*, sehingga setelah proses pembersihan dilakukan, dataset yang digunakan tersisa 38 record data. Penghilangan data outlier ini turut memengaruhi sebaran nilai pada matriks korelasi yang dapat dilihat pada gambar berikut.



Gambar 8. Matriks Correlation

Berdasarkan hasil yang ditampilkan pada Gambar 8, korelasi antar fitur numerik terlihat cukup tinggi, khususnya pada hubungan antara Pokok Panen dan Produksi TBS. Oleh sebab itu, kedua variabel tersebut dipilih sebagai variabel independen (X) dan dependen (Y) untuk keperluan pemodelan regresi.

3.4 Model Regresi

Model regresi dikembangkan dengan memanfaatkan data pokok panen sebagai variabel X dan produksi TBS sebagai variabel Y sebagaimana terlihat pada gambar berikut.

	JUMLAH POKOK	PRODUKSI TBS (Kg)
26	786188	2402850
16	634954	2166590
18	591132	2479350
9	494083	1952180
17	607597	2236120
13	526751	1579530
21	589887	2828530
10	487480	2156550
36	677129	2373170
0	672551	1750770
27	705810	2156820
5	680123	2080440
12	559740	1766100
1	644372	1523130
41	705661	3025980
23	594709	2832160
2	743475	1857020
32	674966	2630690
39	627107	2499880
3	721308	2186520
38	717719	2390410
25	691000	2084850
34	668526	2822150
11	508291	2082090
24	717715	2288260
20	560335	2878970
22	588148	2965330
8	492506	1944130
15	648864	2056700
30	755893	2646140

Gambar 9. Data Latih Jumlah Pokok dan Produksi TBS

Setelah penentuan data Pokok Panen dan Produksi TBS sebagai data penelitian, proses berikutnya adalah melakukan pemisahan data melalui metode *split data*. Proses ini dilakukan untuk membagi dataset menjadi data pelatihan dan data pengujian sebagai tahap evaluasi model. Dalam penelitian ini, rentang data pelatihan digunakan dari Januari 2021 sampai Juni 2024, sedangkan data pengujian mencakup periode Mei 2021 hingga Mei 2024. Hasil pembagian data tersebut dapat dilihat pada gambar berikut.

	JUMLAH POKOK	PRODUKSI TBS (Kg)
37	594503	1903580
40	736215	2672040
4	686108	1932120
14	646920	1926130
33	690710	3040390
28	756566	2647970
6	648822	2347420
29	744267	2279600

Gambar 10. Data Uji Jumlah Pokok dan Produksi TBS

Sesudah tahapan pemisahan data selesai dan kedua jenis data, yaitu data latih dan data uji, telah ditetapkan, proses berikutnya adalah memperoleh nilai koefisien dan intersep dalam model regresi. Perhitungan tersebut dilakukan dengan bantuan perintah pemrograman yang disajikan pada gambar berikut.

*** Persamaan Regresi Linear Sederhana:

$$Y = 1759352.6753 + 0.8316X$$

Nilai Intersep: 1759352.6753

Nilai Koefisien: 0.8316

Gambar 11. (a) Nilai Koefisien (b) Nilai Intersep

Berdasarkan hasil pengujian, nilai koefisien dan intersep yang diperoleh digunakan sebagai dasar dalam penyusunan model regresi, seperti yang ditunjukkan pada persamaan berikut.

$$Y = a + bX$$

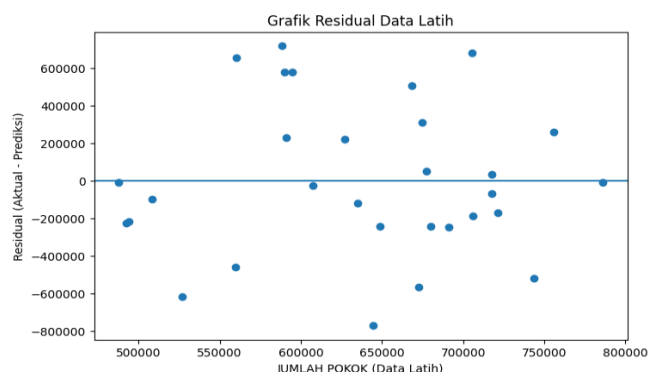
$$Y = 1.759.352,68 + 0,8316X$$

Persamaan regresi yang dihasilkan kemudian digunakan untuk memperkirakan nilai Produksi TBS (Y) pada periode 2025 dengan menjadikan data uji periode Mei 2021–Mei 2024 sebagai dasar analisis

```
*** array([2253768.6332099 , 2371622.83113614, 2329951.55209492,
          2297361.01418469, 2333778.78633659, 2388547.65600628,
          2298942.80460792, 2378319.24359029])
```

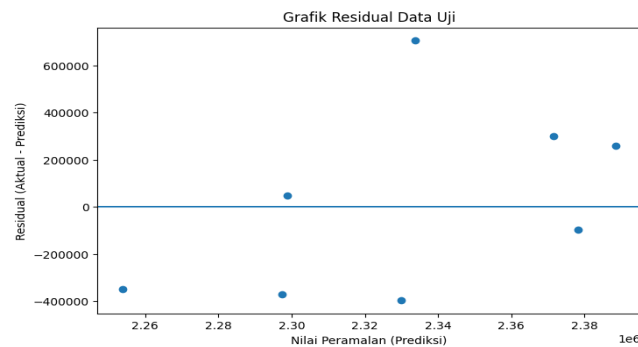
Gambar 12. Hasil Peramalan Produksi TBS Periode 2025

Setelah hasil peramalan diperoleh, model regresi kemudian disajikan dengan menggunakan evaluasi residual seperti yang terlihat pada gambar berikut.



Gambar 13. Grafik Nilai Residual Data Latih

Grafik residual data latih pada Gambar 13 menunjukkan bahwa sebaran residual pola acak di sekitar garis nol, sehingga dapat disimpulkan bahwa model tidak mengalami bias sistematis dalam proses prediksi. Akan tetapi, beberapa titik masih muncul dengan jarak yang cukup jauh dari garis nol, yang mengisyaratkan kemungkinan adanya data *outlier* atau data yang relatif sulit diprediksi oleh model.



Gambar 14. Grafik Nilai Residual Data Uji

Grafik residual data uji pada Gambar 14 menunjukkan bahwa sebagian besar nilai residual berada di sekitar garis nol, yang mengindikasikan bahwa model secara umum mampu memprediksi data di luar pelatihan dengan cukup baik. Namun, masih terdapat beberapa penyimpangan yang cukup besar yang menunjukkan bahwa pada kondisi tertentu prediksi model belum sepenuhnya akurat.

3.5 Evaluasi Model Regresi

Setelah persamaan regresi diperoleh, dilakukan tahap penilaian model untuk mengetahui kemampuan model dalam menghasilkan prediksi yang mendekati nilai sebenarnya. Penilaian ini dilakukan dengan menghitung besarnya selisih antara hasil prediksi dan data aktual guna mengetahui tingkat ketepatan model. Untuk mengukur besarnya kesalahan prediksi tersebut digunakan beberapa parameter evaluasi, yaitu *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), serta *Mean Absolute Percentage Error* (MAPE).

```

***
Hasil Evaluasi Model (Jumlah Pokok):
Mean Absolute Percentage Error (MAPE) : 13.62%
Mean Absolute Error (MAE) : 316612.2956
Mean Squared Error (MSE) : 135957481409.8282
Root Mean Squared Error : 368724.1264

```

Gambar 15. Hasil Evaluasi Model

Berdasarkan hasil pengujian model dalam memprediksi jumlah pokok, diperoleh nilai *Mean Absolute Percentage Error* (MAPE) sebesar 13,62%. Nilai tersebut menunjukkan bahwa rata-rata kesalahan prediksi model berada di sekitar 13,6% dibandingkan nilai aktual, sehingga model dapat dikatakan memiliki tingkat ketepatan yang cukup baik dalam memperkirakan produksi TBS. Sementara itu, nilai *Mean Squared Error* (MSE) tercatat sebesar 1.359.574.814.809,83 dan *Root Mean Squared Error* (RMSE) sebesar 3.687,24, yang menandakan masih adanya selisih yang cukup besar antara hasil prediksi dan data sebenarnya.

3.6 Perbandingan Studi

Hasil penelitian ini menunjukkan bahwa penggunaan regresi linier sederhana dapat memberikan prediksi yang akurat untuk produksi TBS kelapa sawit. Sebagai contoh, penelitian oleh Salsavira (2020) yang menggunakan regresi linier untuk meramalkan supply bahan baku memperoleh MAPE 23,13%, yang lebih tinggi dibandingkan dengan penelitian ini [16]. Hal ini menunjukkan bahwa model yang digunakan dalam penelitian ini lebih efisien dalam memprediksi hasil produksi, meskipun ada kemungkinan untuk meningkatkan akurasi lebih lanjut dengan mempertimbangkan faktor-faktor lain yang lebih dinamis dalam proses produksi kelapa sawit.

3.7 Implikasi dan Keterbatasan

Implikasi dari penelitian ini adalah bahwa model regresi linier sederhana yang dibangun dapat digunakan sebagai alat bantu dalam merencanakan produksi TBS dan meningkatkan efisiensi operasional di perkebunan kelapa sawit. Namun, keterbatasan dari penelitian ini terletak pada penggunaan data yang terbatas hanya dari satu lokasi (PTPN III Rantau Prapat) dan periode yang relatif singkat, yang mungkin tidak sepenuhnya menggambarkan kondisi di daerah lain atau dalam jangka waktu yang lebih panjang. Selain itu, model ini tidak mempertimbangkan

variabel lain yang dapat mempengaruhi hasil produksi seperti kualitas tanah atau pengaruh perubahan iklim yang dapat memberikan dampak signifikan terhadap hasil panen.

4. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, model regresi linier sederhana yang dibangun dalam penelitian ini mampu memprediksi produksi TBS kelapa sawit dengan tingkat akurasi yang memadai. Nilai MAPE sebesar 13,62% mengindikasikan bahwa tingkat penyimpangan prediksi terhadap nilai aktual relatif kecil. Berdasarkan klasifikasi akurasi peramalan, nilai MAPE antara 10%–20% termasuk dalam kategori baik, yang menunjukkan bahwa model regresi yang dibangun cukup efektif dalam memberikan estimasi yang akurat dalam memprediksi hasil produksi TBS. Hal ini sejalan dengan standar umum dalam akurasi peramalan yang sering kali menganggap nilai MAPE di bawah 20% sebagai hasil yang dapat diterima dalam konteks prediksi industri. Penelitian ini dilakukan di PTPN III Rantau Prapat dengan periode data 2021–2024, yang memungkinkan evaluasi model dalam kondisi spesifik wilayah tersebut. Namun, hasil yang diperoleh hanya berlaku untuk lokasi dan periode yang diteliti, sehingga penerapan model pada wilayah lain dengan kondisi lingkungan yang berbeda perlu dievaluasi lebih lanjut. Keterbatasan penelitian ini mencakup penggunaan hanya satu variabel sebagai prediktor, yaitu jumlah pokok panen, yang mungkin tidak sepenuhnya merepresentasikan faktor-faktor lain yang memengaruhi produksi TBS. Selain itu, variabel lain seperti curah hujan, pemupukan, dan faktor iklim belum dipertimbangkan dalam model, yang dapat mempengaruhi hasil panen secara signifikan. Penelitian ini juga belum dilakukan uji regresi berganda, yang dapat membantu mengidentifikasi hubungan yang lebih kompleks antar variabel. Saran untuk penelitian lanjutan adalah penggunaan regresi linier berganda untuk memasukkan lebih banyak variabel yang relevan, seperti curah hujan dan jumlah pupuk, yang dapat meningkatkan akurasi prediksi. Selain itu, penelitian selanjutnya bisa membandingkan model ini dengan metode machine learning lainnya, seperti random forest atau support vector machines, untuk melihat apakah metode tersebut dapat memberikan hasil yang lebih baik dalam memprediksi produksi TBS.

REFERENCES

- [1] Agung Nugroho, “BukuTeknologi Agro Industri Kelapa Sawit,” 2019.
- [2] A. Rifky and S. Anugrah, “SINKRONISASI KEBIJAKAN KELAPA SAWIT BERKELANJUTAN SYNCHRONIZE SUSTAINABLE PALM OIL POLICIES,” 2023. [Online]. Available: <https://jkp.ejournal.unri.ac.idhttps://jkp.ejournal.unri.ac.id>
- [3] D. Poni Egistin, M. Yahdi Rauza, R. Has Ramadhan, and S. Ramadani, “Analisis regresi linier sederhana dan penerapannya,” 2025. [Online]. Available: <https://creativecommons.org/licenses/by/4.0/>
- [4] S. Pd. Si. , M. P. Nuryadi, SE. , M. S. Ak. , CA. , C. Tutut Dewi Astuti, SE. , M. Si. , Ak. , C. Endang Sri Utami, and SE. ,M. Si. ,Ak, C. M. Budiantara, *DASAR-DASAR STATISTIK PENELITIAN*, 1st ed. Bantul: SIBUKU MEDIA, 2017. [Online]. Available: www.sibuku.com
- [5] I. M. Ritonga, I. Zufria, and M. Fakhriza, “ALGORITMA SAW DAN TOPSIS MENENTUKAN BIBIT UNGGUL KELAPA SAWIT,” Aug. 2024. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [6] M. Philipus, P. Raja, and N. Siswanto, “Optimasi Rantai Pasok Tandan Buah Segar Dan Pelepah (Limbah) Kelapa Sawit Pada Hulu Rantai Pasok Industri Kelapa Sawit,” *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT)*, vol. 5, no. 1, pp. 129–140, Mar. 2026.
- [7] C. A. Rahayu, “PREDIKSI PENDERITA DIABETES MENGGUNAKAN METODE NAIVE BAYES,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, pp. 261–266, Aug. 2023, doi: 10.23960/jitet.v11i3.3055.
- [8] R. Nelfi Yolanda, D. Rahmi, A. Kurniati, and S. Yuniati, “Penerapan Metode Triple Exponential Smoothing dalam Peramalan Produksi Buah Nenas di Provinsi Riau,” *Jurnal Teknologi dan Manajemen Industri Terapan (JTMIT)*, vol. 3, no. 1, pp. 1–10, Mar. 2024.
- [9] R. Fijra, M. Navero, J. Jenderal Ahmad Yani, K. I. Seberang Ulu, K. Palembang, and S. Selatan, “Peramalan Produksi CPO (Crude Palm Oil) Pada PT. XYZ,” Jan. 2023.
- [10] Muhammad Asyorori and Wirda Andani, “ANALISIS REGRESI VARIABEL MEDIASI DENGAN METODE KAUSAL STEP,” *Jurnal Buletin Ilmiah Math. Stat. dan Terapannya*, 2023.
- [11] M. Riziq sirfatullah Alfarizi, M. Zidan Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, “PENGUNAAN PYTHON SEBAGAI BAHASA PEMROGRAMAN UNTUK MACHINE LEARNING DAN DEEP LEARNING,” 2023.
- [12] R. Kurniawan, P. B. Wintoro, Y. Mulyani, and M. Komarudin, “IMPLEMENTASI ARSITEKTUR XCEPTION PADA MODEL MACHINE LEARNING KLASIFIKASI SAMPAH ANORGANIK,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 2, Apr. 2023, doi: 10.23960/jitet.v11i2.3034.
- [13] R. Diana, H. Warni, and T. Sutabri, “PENGUNAAN TEKNOLOGI MACHINE LEARNING UNTUK PELAYANAN MONITORING KEGIATAN BELAJAR MENGAJAR PADA SMK BINA SRIWIJAYA PALEMBANG,” *JUTEKIN (Jurnal Teknik Informatika)*, vol. 11, no. 1, pp. 41–50, Jun. 2023, doi: 10.51530/jutekin.v11i1.709.



- [14] A. Setiawan, H. Annur, and S. Melangi, “Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Jumlah Persediaan Pestisida,” *Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer*, vol. 3, no. 1, 2024.
- [15] N. L. Salsavira, D. E. Yulawati, and T. Surabaya, “Peramalan Supply Bahan Baku Menggunakan Metode Regresi Linier dan Exponential Smoothing,” *Jurnal Nusantara Of Engineering*, vol. 06, no. 2, pp. 183–189, Oct. 2023, [Online]. Available: <https://ojs.unpkediri.ac.id/index.php/noe>
- [16] N. Moha Lalapa and W. Yunus, “IMPLEMENTASI METODE REGRESI LINEAR SEDERHANA UNTUK PREDIKSI HARGA CABAI RAWIT,” *Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer*, vol. 2, no. 2, p. 96, 2023.
- [17] D. Arifuddin, K. Kusrini, and K. Kusnawi, “Perbandingan Performansi Algoritma Multiple Linear Regression dan Multi Layer Perceptron Neural Network dalam Memprediksi Penjualan Obat,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 2, pp. 722–737, Apr. 2025, doi: 10.57152/malcom.v5i2.1952.
- [18] I. Armansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, “Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia,” *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 4, pp. 1199–1216, 2024.