



ISSN : 2339 - 1871

## BETRIK BESEMAH TEKNOLOGI INFORMASI & KOMPUTER

Editor Office : Pusat Penelitian & Pengabdian Pada Masyarakat  
(PPPM) ITPA

Phone : 0857-9716-9578

email : [betriktpa@itpa.ac.id](mailto:betriktpa@itpa.ac.id)

### Model Prediksi Jumlah Produksi Kelapa Sawit Menggunakan Regresi Linear Berganda Di PT.Surya Argolika Reksa

M Irpan<sup>1</sup>, Unang Rio<sup>2</sup>, Karpen<sup>3</sup>, Hamdani<sup>4</sup>

Program Studi Teknik Informatika, Universitas Sains dan Teknologi Indonesia<sup>1,2,3,4</sup>  
Jalan Purwodadi No. KM. 10, Sidomulyo Barat, Tampan, Kota Pekanbaru, Riau 28294, Indonesia  
Sur-el : [2110031802100@sar.ac.id](mailto:2110031802100@sar.ac.id)<sup>1</sup>, [Unangrio@usti.ac.id](mailto:Unangrio@usti.ac.id)<sup>2</sup>, [karpenkarpen49@gmail.com](mailto:karpenkarpen49@gmail.com)<sup>3</sup>,  
[hamdani1202271@gmail.com](mailto:hamdani1202271@gmail.com)<sup>4</sup>

Penulis Korespondensi: M.Irpan, [2110031802100@sar.ac.id](mailto:2110031802100@sar.ac.id)

**Abstrak:** Perkebunan kelapa sawit merupakan salah satu sektor strategis dalam industri agribisnis di Indonesia. Untuk mendukung efisiensi dan efektivitas produksi, diperlukan model prediksi yang mampu memperkirakan jumlah produksi secara akurat berdasarkan faktor-faktor pendukung. Penelitian ini bertujuan untuk membangun model prediksi jumlah produksi kelapa sawit menggunakan algoritma Regresi Linear Berganda dengan memanfaatkan data variabel luas lahan (Ha), jumlah pohon, dan curah hujan. Data diperoleh dari laporan operasional PT. Surya Argolika Reksa. Evaluasi model dilakukan dengan dua skenario pembagian data, yaitu 80:20 dan 70:30. Hasil evaluasi menunjukkan bahwa pada data uji 80:20 diperoleh nilai MAE sebesar 30.095,68, MSE sebesar 1.533.325.063,46, dan RMSE sebesar 39.151,33. Sementara itu, pada data uji 70:30 diperoleh nilai MAE sebesar 35.455,01, MSE sebesar 2.096.902.404,44, dan RMSE sebesar 45.791,95. Nilai-nilai tersebut menunjukkan tingkat kesalahan prediksi model dalam satuan jumlah produksi kelapa sawit. Penelitian ini memberikan kontribusi dalam mendukung perencanaan produksi yang lebih akurat di sektor perkebunan kelapa sawit berbasis analisis data.

**Kata kunci :** Regresi Linear Berganda, Prediksi Produksi, Kelapa Sawit, Evaluasi Model, MAE, MSE, RMSE.

**Abstract:** Palm oil plantations are one of the strategic sectors in Indonesia's agribusiness industry. To support production efficiency and effectiveness, a predictive model capable of accurately estimating production volume based on supporting factors is required. This study aims to develop a prediction model for palm oil production using the Multiple Linear Regression algorithm by utilizing variables such as land area (Ha), number of trees, and rainfall. The data were obtained from the operational reports of PT. Surya Argolika Reksa. The model evaluation was conducted using two data splitting scenarios: 80:20 and 70:30. The evaluation results show that for the 80:20 test data, the MAE value was 30,095.68, the MSE was 1,533,325,063.46, and the RMSE was 39,151.33. Meanwhile, for the 70:30 test data, the MAE value was 35,455.01, the MSE was 2,096,902,404.44, and the RMSE was 45,791.95. These values indicate the level of prediction error of the model in units of palm oil production. This research contributes to supporting more accurate production planning in the palm oil plantation sector based on data analysis.

**Keywords:** Multiple Linear Regression, Production Prediction, Palm Oil, Model Evaluation, MAE, MSE, RMSE.

Received: 26-07-2025 | Accepted: 28-07-2025 | Published Online: 30-08-2025

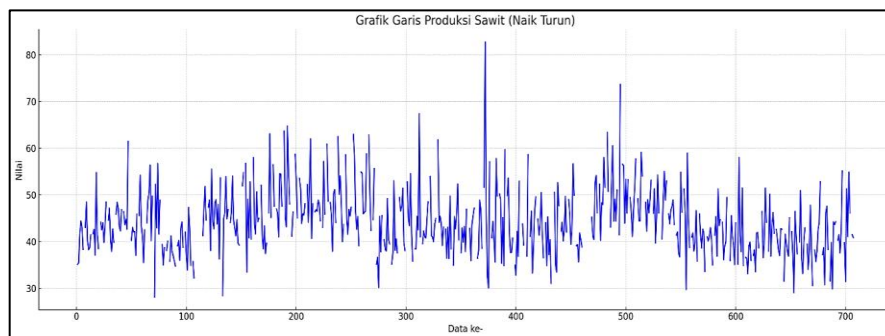
All author: M. Irpan, Unang Rio, Karpen, Hamdani

## 1. PENDAHULUAN

Industri kelapa sawit menjadi salah satu sektor utama dalam perekonomian Indonesia, yang berkontribusi besar terhadap pendapatan nasional, ekspor, dan penyerapan tenaga kerja, menurut[1], sector Kelapa sawit memiliki nilai indeks daya keterkaitan (IDK) di atas satu, menunjukkan bahwa sector ini mampu mendorong pertumbuhan sektor hilirnya dan memberikan efek multiplier terhadap perekonomian nasional.

Dalam oprasional perkebunan, salah satu tantangan utama yang dihadapi adalah ketidakpastian dalam jumlah produksi Kelapa sawit setiap priode. Jumlah produksi yang fluktuatif dapat disebabkan oleh berbagai factor seperti cuaca, tahun tanam, luas lahan, jumlah pohon,, serta factor lingkungan lainnya, penelitian oleh [2] menunjukkan bahwa variabel-variabel tersebut secara signifikan mempengaruhi jumlah produksi Kelapa sawit dikabupaten aceh timur. Ketidak pastian ini hanya berdampak pada proses panen dan pengolahan, tetapi juga berpengaruh terhadap perencanaan anggaran, tenaga kerja, distribusi, dan pemasaran hasil kebun.

PT.Surya Argolika Reksa, sebagai perusahaan yang berfokus pada sektor perkebunan kelapa sawit, juga menghadapi tantangan serupa. Variasi jumlah produksi yang tidak dapat diprediksi secara akurat menyebabkan kesulitan dalam pengelolaan oprasional dan perencanaan strategis perusahaan. Oleh karena itu perlu metode yang tepat untuk memprediksi jumlah produksi Kelapa sawit berdasarkan faktor-faktor yang mempengaruhinya.



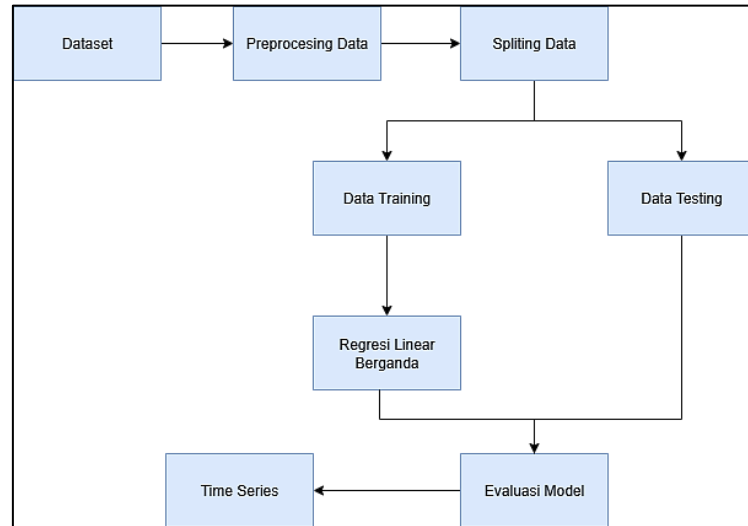
Gambar 1. Grafik Hasil Jumlah Produksi 2023-2024

Metode yang dapat di terapkan untuk tujuan ini adalah metode Regresi Linear Berganda, yang mampu menganalisis hubungan antara jumlah produksi dengan beberapa variabel seperti luas lahan, jumlah pohon, tahun tanam,dan faktor cuaca, dengan bantuan teknologi dan alat analisis berbasis python, proses pengolahan data, pembuatan model prediksi, serta evaluasi kinerja model dapat Dilakukan dengan lebih cepat, akurat, dan efesien[3].

Melalui pengembangan model prediksi jumlah produksi Kelapa sawit ini, diharapkan PT.Surya Argolika Reksa dapat memiliki alat bantu dalam menggambarkan hasil produksi masa depan dengan lebih baik, sehingga mendukung efesien oprasional dan meningkatkan produktivitas perusahaan secara keseluruhan.

## 2. METODOLOGI PENELITIAN

Desain penelitian menyajikan alur sistematis dari setiap tahap yang dilaksanakan sejak awal hingga akhir guna mencapai tujuan studi. diagram ini menggambarkan alur kerja yang logis dan terstruktur dalam membangun model prediksi jumlah produksi menggunakan regresi linear berganda.



Gambar 2. Metodologi Penelitian

### 2.1 Pengumpulan Data

Penelitian ini menggunakan pendekatan kuantitatif dengan metode prediksi berbasis regresi linier berganda. Dataset yang digunakan diperoleh dari PT. Surya Argolika Reksa dan terdiri dari 73 entri data historis harian terkait jumlah produksi kelapa sawit serta variabel-variabel yang memengaruhinya. Variabel independen yang digunakan meliputi luas lahan (Ha), jumlah pohon, dan jumlah curah hujan per tahun, sementara jumlah produksi kelapa sawit digunakan sebagai variabel dependen (target). Data dikumpulkan dalam bentuk file Excel dan mencerminkan kondisi aktual yang terjadi di lapangan.

Berikut adalah 10 sampel data yang di tampilkan 5 data awal dan 5 data akhir, dapat dilihat sebagai berikut:

Tabel 1. Data Luas Lahan, Tahun Tanam, Jumlah Pohon, dan Jumlah Produksi

Inti	no. blok	tahun tanam	luas (Ha)	jumlah pohon	jumlah produksi
1	G 1	2002	22.97	2,940	612.900
1	G 2	2002	12.04	1,649	319.110
1	G 3	2002	26.93	3,555	668.000
1	G 5	2002	8.67	1,144	192.540
1	G 6	2002	19.34	2,533	596.580
....	....	....	....	....	....
2	50	2003	3.65	436	2,561
2	51	2003	31.00	5.700	2,561
2		2003		2.000	2,561
2				500	2,561
2					2,561

Sumber : PT.Surya Argolika Reksa

Tabel 2. Data Cuaca dan Jumlah Produksi Harian

No	Tgl	Curah Hujan(mm)	Suhu(°C)	Kelembapan(%)	Jumlah produksi inti surya
1	01/01/2023	10	22	82	-
2	02/01/2023	18	24	82	35.140
3	03/01/2023	12	24	82	35.480
4	04/01/2023	0	31	83	41.610
5	05/01/2023	0	26	88	44.410
....	....	....	....	....	....
727	27/12/2024	0	26	82	
728	28/12/2024	0	26	82	
729	29/12/2024	0	26	82	
730	30/12/2024	0	26	82	
731	31/12/2024	0	26	82	

Sumber : PT.Surya Argolika Reksa

## 2.2 Splitting Data

Tabel 3. Splitting data

no	Data latih	Data uji
1	80%	20%
2	70%	30%x

## 2.3 Pre-Processing Data

Preprocessing adalah langkah-langkah yang Dilakukan sebelum analisis data untuk memastikan bahwa data siap pakai dan berkualitas tinggi, preprocessing pada dasarnya di terapkan pada data mentah apapun sebelum menggunakan prediksi atau identifikasi apapun. Dalam banyak kasus didunia nyata, data mentah tidak dapat digunakan karna gangguan dan harus diproses terlebih dahulu[4], dalam penelitian ini ada satu tahapan preprocessing yaitu *Cleaning*. Data *cleaning* adalah langkah untuk mengenali dan menangani data yang tidak valid, kurang lengkap, atau tidak di butuhkan dalam kumpulan data.Tujuan utama dari pembersihan data adalah untuk meningkatkan kualitas data, sehingga analisis dan model yang dibangun berdasarkan data tersebut menjadi lebih akurat dan dapat diandalkan[5].

## 2.4 Feature Selection

*Featur selection* adalah tahapan untuk menentukan sekumpulan variabel yang paling berpengaruh terhadap kinerja model, dengan tujuan mengurangi dimensi data, menghindari overfitting, serta meningkatkan efisiensi dan interpretabilitas model[6]. Salah satu metode populer adalah Lasso Regression ,Lasso menggunakan penalti L1 yang mendorong koefisien variabel tertentu menjadi nol, sehingga secara otomatis memilih fitur yang paling penting.

Teknik ini sangat berguna dalam regresi linear berganda yang menggunakan banyak variabel input[7].

## 2.5 Time Series

Deret waktu adalah data yang di rekam secara kronologis dalam jeda waktu yang tetap, seperti harian, bulanan, atau tahunan. Analisis deret waktu bertujuan untuk memahami pola atau tren data berdasarkan waktu, serta melakukan prediksi terhadap nilai di masa depan berdasarkan nilai-nilai historis yang telah dikumpulkan[8].

## 2.6 Pemodelan Regresi Linear Berganda

Regresi linear berganda merupakan metode statistik yang di gunakan untuk menggambarkan hubungan antara satu variabel target dan beberapa variabel predator[6]. Model ini memungkinkan analisis terhadap pengaruh simultan dari beberapa faktor terhadap suatu variabel target. Menurut [9], salah satu keunggulan regresi linear berganda adalah kemampuannya untuk menentukan dan mengukur tingkat kontribusi masing masing variabel bebas terhadap variabel terikat. Namun, model ini juga memiliki kelemahan, seperti sensitif terhadap keberadaan outlier dan memerlukan asumsi-asumsi klasik seperti normalisasi, multikolinearitas rendah, dan homoskedastitas. Dari sisi matematis, regresi linear berganda disajikan dengan persamaan berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (1)$$

## 2.7 Python

Bahasa Pemrograman *Python* adalah bahasa pemrograman tingkat tinggi yang dirancang untuk mempermudah pengembangan perangkat lunak dengan sintaks yang sederhana dan mudah dipahami. *Python* sangat populer di kalangan data scientist, peneliti, dan pengembang aplikasi karena fleksibilitasnya yang tinggi serta ekosistem pustaka yang kaya [10].

## 2.8 Evaluasi Model

Untuk mengetahui akurasi model dalam memprediksi variabel dependen, digunakan metrik evaluasi sebagai alat pengukurannya.

### 1. Mean Absolute Error(MAE)

Mean Absolute Error (MAE) adalah metrik yang mengukur rata-rata kesalahan absolut antara nilai yang diprediksi dan nilai aktual. MAE memberikan gambaran yang jelas tentang seberapa jauh prediksi dari nilai sebenarnya[3]. Rumus Mean Absolute Error(MAE) dinyatakan sebagai berikut :

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

### 2. Mean Square Error(MSE)

Mean Square Error (MSE) merupakan metrik evaluasi yang menghitung rata rata dari kuadrat selisih antara prediksi dan nilai sebenarnya, dimana kesalahan yang berat akan dihukum

lebih berat, sehingga lebih sensitif terhadap outlier[3]. Rumus Mean Square Error(MSE) dinyatakan sebagai berikut :

$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

### 3. Root Mean Square Error(RMSE)

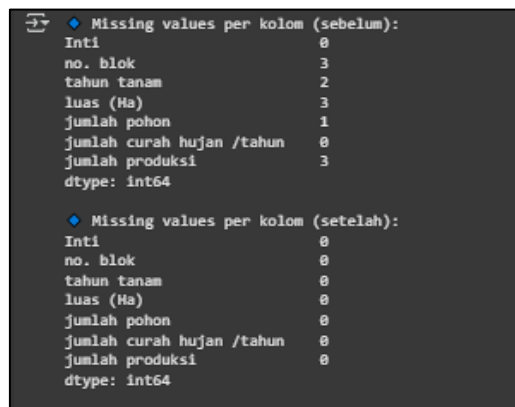
Root Mean Square Error (RMSE) adalah akar kuadrat dari MSE. RMSE memberikan ukuran kesalahan dalam satuan yang sama dengan variabel dependen, sehingga lebih mudah untuk diinterpretasikan[3]. Root Mean Square Error(RSME) dinyatakan sebagai berikut :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Preprocessing Data

#### Cleaning



```
Missing values per kolom (sebelum):
Inti 0
no. blok 3
tahun tanam 2
luas (Ha) 3
jumlah pohon 1
jumlah curah hujan /tahun 0
jumlah produksi 3
dtype: int64

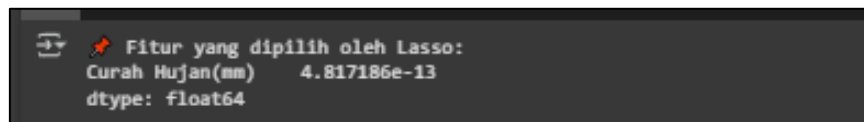
Missing values per kolom (setelah):
Inti 0
no. blok 0
tahun tanam 0
luas (Ha) 0
jumlah pohon 0
jumlah curah hujan /tahun 0
jumlah produksi 0
dtype: int64
```

Gambar 3. Cleaning Data

Berdasarkan gambar di atas sebelum dilakukan pembersihan data, ditemukan adanya nilai kosong (*missing values*) pada beberapa kolom, seperti kolom no. blok sebanyak 3 data, tahun tanam sebanyak 2 data, luas (Ha) sebanyak 3 data, jumlah pohon sebanyak 1 data, dan jumlah produksi sebanyak 3 data.

### 3.2 Feature Engineering

#### Feature Selection

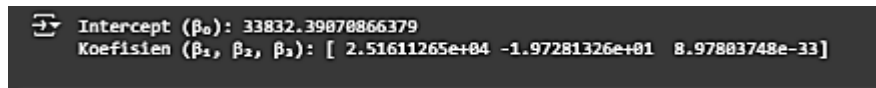


```
Fitur yang dipilih oleh Lasso:
Curah Hujan(mm) 4.817186e-13
dtype: float64
```

Gambar 4. Hasil Selection Regression

Setelah Dilakukan *feature selection* data cuaca menggunakan lasso regression, curah hujan adalah variabel yang paling relevan atau berpengaruh terhadap jumlah produksi sedangkan variabel suhu dan kelembapan tidak di pilih lasso untuk prediksi jumlah produksi.

### 3.3 Hasil Pemodelan Regresi Linear Berganda



Gambar 5. Hasil Persamaan Regresi

Model regresi linear berganda yang dihasilkan memiliki intercept ( $\beta_0$ ) sebesar 33.832,39 dengan koefisien  $\beta_1$  untuk variabel luas lahan sebesar 25.161,13, koefisien  $\beta_2$  untuk jumlah pohon sebesar -19,73, dan koefisien  $\beta_3$  untuk curah hujan sebesar 0,08978.

Hasil Evaluasi Model 80:20:  
 MAE : 35,971.74  
 MSE : 2,148,369,896.29  
 RMSE : 46,350.51

	Luas (Ha)	Jumlah Pohon	Curah Hujan	Aktual	Prediksi
0	9.00	1188.000	2.561	288500	235467.994978
1	22.97	2940.000	2.561	612900	548691.884282
2	9.94	1.322	2.561	280230	291383.359784
3	19.34	2.533	2.561	596580	533592.759491
4	6.34	749.000	2.561	158880	178642.584721

Gambar 6. Hasil Prediksi Spliting Data 80:20

Berdasarkan hasil evaluasi model dengan data splitting 80:20, diperoleh bahwa model regresi linear berganda memiliki *Mean Absolute Error* (MAE) sebesar 35.971,74 yang menunjukkan rata-rata kesalahan prediksi sekitar  $\pm 35$  ribu kg, *Mean Squared Error* (MSE) sebesar 2.148.369.896,29 yang mencerminkan adanya perbedaan kuadrat yang cukup besar antara nilai aktual dan prediksi, serta *Root Mean Squared Error* (RMSE) sebesar 46.350,51.

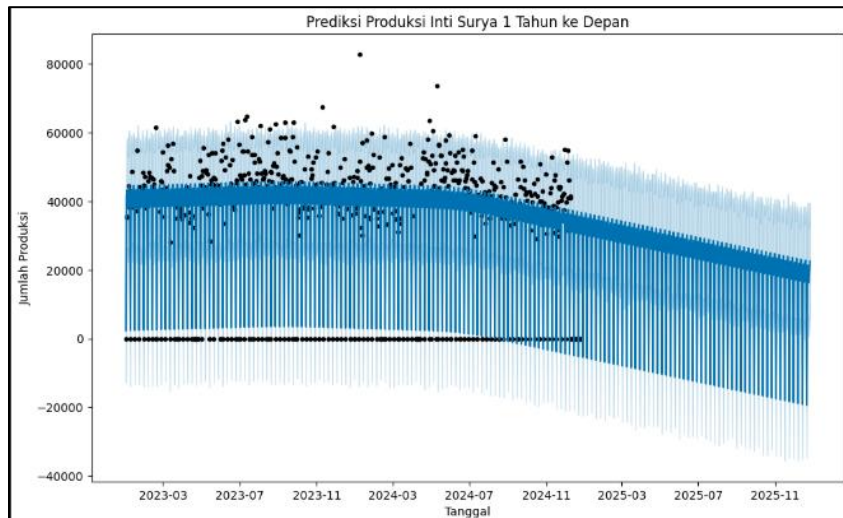
Hasil Evaluasi Model 70:30:  
 MAE : 35,455.01  
 MSE : 2,096,902,404.44  
 RMSE : 45,791.95

	Luas (Ha)	Jumlah Pohon	Curah Hujan	Aktual	Prediksi
0	9.00	1188.000	2.561	288500	234797.342759
1	22.97	2940.000	2.561	612900	544246.848028
2	9.94	1.322	2.561	280230	295134.545216
3	19.34	2.533	2.561	596580	539025.220213
4	6.34	749.000	2.561	158880	179088.397287

Gambar 7. Hasil Prediksi Spliting Data 70:30

Berdasarkan hasil evaluasi model dengan pembagian data 70:30, diperoleh Mean Absolute Error (MAE) sebesar 35.455,01 yang menunjukkan deviasi rata-rata sebesar 35 ribu kg dari nilai produksi aktual, Mean Squared Error (MSE) sebesar 2.096.902.404,44 yang mengindikasikan masih adanya perbedaan cukup besar antara hasil prediksi dan nilai aktual pada beberapa data, serta Root Mean Squared Error (RMSE) sebesar 45.791,95.

### 3.4 Hasil Time Series



Gambar 8. Hasil Prediksi Satu Tahun

Gambar tersebut menunjukkan hasil prediksi jumlah produksi inti sawit selama satu tahun ke depan menggunakan model Prophet, di mana garis biru tebal merepresentasikan nilai prediksi utama (yhat), area biru muda menunjukkan interval kepercayaan prediksi, dan titik-titik hitam merupakan data historis aktual; dari visualisasi terlihat bahwa tren produksi diprediksi menurun secara bertahap mulai pertengahan 2024 hingga akhir 2025, yang mengindikasikan potensi penurunan hasil produksi di masa mendatang jika tidak ada intervensi, serta rentang prediksi yang makin melebar menunjukkan meningkatnya ketidakpastian prediksi seiring waktu, sesuai karakteristik model deret waktu.

## 4. KESIMPULAN

Penelitian ini berhasil membangun model prediksi jumlah produksi kelapa sawit menggunakan algoritma Regresi Linear Berganda berdasarkan tiga variabel prediktor utama, yaitu luas lahan, jumlah pohon, dan curah hujan. Hasil evaluasi model dengan skenario pembagian data 80:20 dan 70:30 menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang masih dapat diterima, dengan nilai MAE masing-masing sebesar 30.095,68 dan 35.455,01; MSE sebesar 1.533.325.063,46 dan 2.096.902.404,44; serta RMSE sebesar 39.151,33 dan 45.791,95. Temuan ini mengindikasikan bahwa model dapat memberikan estimasi awal yang cukup akurat terhadap jumlah produksi kelapa sawit di masa mendatang. Dengan demikian, model ini berpotensi digunakan sebagai alat bantu dalam perencanaan produksi dan pengambilan keputusan di sektor perkebunan kelapa sawit, khususnya dalam memetakan dampak dari perubahan faktor-faktor lingkungan dan agronomi terhadap hasil produksi.

## 5. UCAPAN TERIMA KASIH

Segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala atas limpahan rahmat dan karunia-Nya sehingga penulisan artikel ini dapat terselesaikan dengan baik. Penulis juga menyampaikan apresiasi yang setulusnya kepada Bapak Unang Rio, M.Kom., selaku dosen

pembimbing, atas bimbingan, arahan, serta dukungan yang telah diberikan selama proses penelitian. Terima kasih juga disampaikan kepada Universitas Sains dan Teknologi Indonesia, khususnya Program Studi Teknik Informatika, atas dukungan akademik dan fasilitas yang diberikan selama penyusunan artikel ini.

Penulis juga menyampaikan apresiasi kepada PT. Surya Argolika Reksa atas kerja sama dan izin pemanfaatan data dalam penelitian ini. Penulis juga menyampaikan apresiasi dan terima kasih yang sebesar-besarnya kepada seluruh rekan, kolega, serta pihak-pihak yang telah memberikan dukungan, bantuan, dan motivasi, baik secara langsung maupun tidak langsung, selama proses penelitian dan penulisan jurnal ini berlangsung. Penghargaan yang tinggi juga diberikan kepada editor dan tim reviewer jurnal atas saran dan evaluasi yang konstruktif selama proses publikasi artikel ini.

## DAFTAR RUJUKAN

- [1] T. Ningsih, I. O.Y. Sitompul, and S. D. Siahaan, "Analisa Faktor-Faktor Yang Mempengaruhi Produksi Kelapa Sawit Di Kebun Tanah Raja PT. Bakrie Sumatera Plantations," *J. Agribus. Sci.*, vol. 07, no. 2, pp. 166–174, 2023.
- [2] A. Ridha, "Faktor-Faktor Yang Mempengaruhi Produksi Kelapa Sawit Di Kecamatan Arongan Lambalek Kabupaten Aceh Barat," *Samudra Ekon.*, vol. 2, no. February, p. 13, 2018.
- [3] H. Hermansyah, A. Abdullah, and P. Y. Utami, "Penerapan Metode Regresi Linier Berganda Untuk Memprediksi Panen Kelapa Sawit," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 540, 2024, doi: 10.35889/progresif.v20i1.1816.
- [4] A. Rahman, "Statistics-based data preprocessing methods and machine learning algorithms for big data analysis," *Int. J. Artif. Intell.*, vol. 17, no. 2, pp. 44–65, 2019.
- [5] H. Hanafi, "Data Cleaning dalam Big Data : Review," no. December, pp. 1–5, 2023, [Online]. Available: <https://www.researchgate.net/publication/376758940>
- [6] "View of Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang.pdf."
- [7] M. Robbani, F. Agustiani, and N. Herrhyanto, "Regresi Least Absolute Shrinkage and Selection Operator (Lasso) Pada Kasus Inflasi Di Indonesia Tahun 2014-2017," *J. EurekaMatika*, vol. 7, no. 2, pp. 1–16, 2019,
- [8] R. Mubarak, T. Tursina, and E. E. Pratama, "Prediksi Hasil Tangkapan Ikan Menggunakan Fuzzy Time Series," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, p. 303, 2020, doi: 10.26418/justin.v8i3.39831.
- [9] N. Khan *et al.*, "Prediction of Oil Palm Yield Using Machine Learning in the Perspective of Fluctuating Weather and Soil Moisture Conditions: Evaluation of a Generic Workflow," *Plants*, vol. 11, no. 13, 2022, doi: 10.3390/plants11131697.
- [10] T. T. Teoh and Z. Rong, *Artificial Intelligence with Python*. 2022.