

# Penerapan Algoritme *Linear Regression* untuk Prediksi Hasil Panen Tanaman Padi

Heru Wahyu Herwanto<sup>1</sup>, Triyanna Widiyaningtyas<sup>2</sup>, Poppy Indriana<sup>3</sup>

**Abstract**—Rice yields are very influential in meeting the basic food needs of rice. Because the needs of rice are always rising, it is necessary to predict crop yields to estimate the future planting to meet the basic food needs. The method used in this paper is linear regression algorithm, which can predict the yield of rice plants. The steps in this research are as follows: (1) data collection through surveys to farmers in Lamongan by giving questionnaires to respondents; (2) pre-processing the data, which is data cleaning; (3) applying linear regression to determine the strength of the relationship between one dependent or independent variable and a set of independent or independent variables; and (4) results of the validation. Testing accuracy is carried out by measuring Root Mean Squared Error (RMSE). The average value of accuracy of the RMSE is 0.432. This indicates that the variation of values produced by a forecast model is close to accurate, and results in the compatibility of the Multiple Linear Regression Model, with a reliability level of 94.51%.

**Intisari**—Hasil panen padi sangat berpengaruh dalam pemenuhan kebutuhan beras. Dikarenakan kebutuhan beras tiap tahunnya selalu mengalami peningkatan, diperlukan prediksi hasil panen untuk memperkirakan penanaman yang akan dilakukan sehingga dapat mencukupi pemenuhan kebutuhan pokok pangan. Metode yang digunakan dalam makalah ini adalah algoritme *linear regression* yang dapat melakukan prediksi terhadap hasil panen tanaman padi. Tahapan yang dilakukan adalah sebagai berikut: (1) pengumpulan data melalui survei kepada petani yang ada di Lamongan dengan memberikan angket kepada responden; (2) *pre-processing* data ini dengan *cleaning* data; (3) penerapan *linear regression* untuk menentukan kekuatan hubungan antara satu variabel dependen (tak bebas) dan serangkaian variabel independen (bebas); dan (4) validasi hasil. Pegujian akurasi dilakukan dengan mengukur *Root Mean Squared Error* (RMSE). Nilai rata-rata akurasi RMSE yang dihasilkan, sebesar 0,432, menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati akurat, serta menghasilkan kecocokan model *multiple linear regression* dengan tingkat keandalan sebesar 94,51%.

**Kata Kunci**—Algoritme *Linear Regression*, *Tanaman Padi*, *Prediksi*, *Hasil Panen*.

## I. PENDAHULUAN

Padi merupakan bahan dasar dari beras yang masuk ke dalam bahan pangan pokok terbesar di dunia. Padi termasuk dalam suku padi-padian (*poaceae*) [1]. Produksi padi dunia saat ini

menempati urutan ketiga setelah jagung dan gandum. Hal ini berlaku pada Benua Asia, yang merupakan tempat tinggal dari para petani yang memproduksi sekitar 90% dari total produksi beras dunia [2]. Indonesia sendiri masuk dalam peringkat ke-3 produsen beras terbesar dunia, setelah China dan India. Provinsi Jawa Timur masuk peringkat ke-4 sebagai provinsi penghasil beras terbesar, setelah Provinsi Sumatera Selatan, Jawa Barat, dan Jawa Tengah. Padi merupakan tanaman pangan utama di Indonesia dikarenakan sebagian besar penduduknya memilih beras sebagai makanan pokok sehari-hari. Padi merupakan komoditas pangan yang berperan penting dalam kehidupan ekonomi masyarakat Indonesia dalam memenuhi kebutuhan asupan karbohidrat yang mengenyangkan untuk makanan pokok sehari-hari [3]. Di wilayah ASEAN, beras sangat cocok dibudidayakan karena wilayah ASEAN memiliki iklim yang hangat dan curah hujan yang tinggi [4]. Berdasarkan data dari Badan Pusat Statistik (BPS) pada tahun 2018, Indonesia memproduksi beras sebesar 32,4 juta ton, dengan target 48 juta ton hingga akhir tahun oleh Kementerian Pertanian. Perkiraan konsumsi beras hingga akhir tahun mencapai 29,6 juta ton. Selisih 2,85 juta ton tersebut diperkirakan tidak akan cukup untuk memenuhi kebutuhan.

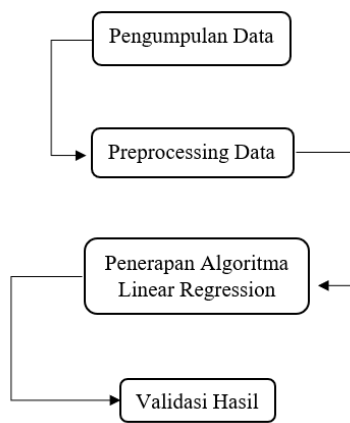
Hasil panen padi sangat berpengaruh dalam pemenuhan kebutuhan pangan pokok beras. Dikarenakan kebutuhan beras tiap tahunnya selalu mengalami peningkatan, diperlukan prediksi hasil panen untuk memperkirakan penanaman yang akan dilakukan, sehingga dapat mencukupi pemenuhan kebutuhan beras. Untuk memperoleh hasil panen, diperlukan informasi mengenai luas lahan, jumlah bibit yang akan di tanam, dan pupuk yang digunakan sebagai penunjang kebutuhan tanaman padi.

Luas lahan yang digunakan untuk kegiatan menanam padi sangat berpengaruh terhadap hasil panen yang diperoleh. Apabila luas lahan yang digunakan semakin besar, maka potensi hasil panen juga semakin besar [4]. Lahan yang baik untuk pertumbuhan tanaman padi adalah tanah sawah yang mengandung fraksi pasir, debu, dan lempung dalam perbandingan tertentu, serta diperlukan air dalam jumlah yang cukup.

Jumlah bibit padi yang akan ditanam tergantung pada luas lahan yang akan digunakan sebagai media tanam. Apabila luas lahan semakin besar, maka jumlah bibit padi yang akan digunakan juga besar atau banyak. Semakin besar jumlah bibit padi yang ditanam, jumlah hasil panen yang dapat diperoleh juga semakin besar. Bibit yang bermutu merupakan salah satu komponen teknologi yang penting untuk meningkatkan produksi dan pendapatan usaha tani [5]. Penggunaan pupuk juga berpengaruh dalam hasil produksi tanaman padi. Pupuk merupakan faktor yang memengaruhi produktivitas hasil panen tanaman padi. Namun, ketepatan bibit padi yang dipilih juga

<sup>1,2</sup> Jurusan Teknik Elektro Fakultas Teknik Universitas Negeri Malang, Jl. Semarang No.5 Sumbersari, Kec. Lowokwaru, Kota Malang, Jawa Timur 65145 INDONESIA (telp: 0341-573090; e-mail: heru\_wh@um.ac.id, triyannaw.ft@um.ac.id)

<sup>3</sup> Jurusan Teknik Elektro Universitas Negeri Malang, Jl. Semarang No.5 Sumbersari, Kec. Lowokwaru, Kota Malang, Jawa Timur 65145 INDONESIA (telp: 085859455293; e-mail: ppyindma23@gmail.com)



Gbr. 1 Tahapan penelitian.

memengaruhi kecocokan dengan pupuk, sehingga dapat menghasilkan hasil panen yang optimal.

Prediksi merupakan proses memperkirakan sesuatu yang akan terjadi di masa depan secara sistematis berdasarkan informasi yang dimiliki di masa lalu dan sekarang [6]. Algoritme *linear regression* adalah metode statistik yang digunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel. Variabel merupakan besaran yang berubah-ubah nilainya. Variabel yang memengaruhi disebut dengan variabel bebas, variabel independen, atau variabel penjelas. Variabel yang dipengaruhi disebut dengan variabel terikat atau variabel dependen [7].

Penelitian terkait yang telah dilakukan tentang estimasi luas panen padi menggunakan algoritme *linear regression* menghasilkan nilai kinerja sebesar 185,52 [8]. Sedangkan pada penelitian pengenalan pola curah hujan menggunakan algoritme yang sama dihasilkan tingkat kecocokan model sebesar 57,78% [9]. Oleh karena itu, algoritme *linear regression* merupakan model prediksi atau peramalan yang terdapat lebih dari satu prediktor yang dapat digunakan untuk memprediksi hasil panen padi berdasarkan luas lahan, bibit, dan pupuk yang dilakukan pada makalah ini.

## II. METODE

Tahap-tahap yang dilakukan meliputi pengumpulan data, *pre-processing* data, penerapan algoritme *linear regression*, dan validasi hasil, seperti ditunjukkan pada Gbr. 1.

### A. Pengumpulan Data

Dalam makalah ini digunakan data yang dikumpulkan melalui survei kepada petani yang ada di Lamongan dengan memberikan angket kepada responden. Angket tersebut berisi informasi mengenai nama petani, luas lahan yang digunakan, jumlah bibit padi yang ditanam, hasil panen tanaman padi dalam sekali tanam, jenis pupuk yang digunakan, serta takaran pupuk yang digunakan dalam sekali tanam. Informasi mengenai curah hujan, kondisi tanah, dan sebagainya tidak ada dalam makalah ini karena penelitian hanya fokus pada luas lahan yang digunakan, jumlah bibit padi yang ditanam, hasil panen tanaman padi dalam sekali tanam, dan jenis pupuk yang digunakan serta takaran pupuk yang digunakan dalam sekali

tanam. Terdapat enam atribut yang digunakan, yaitu Nama Petani, Luas Lahan, Jumlah Bibit, Pupuk Urea, Pupuk NPK Phonska, dan Hasil Panen. Jumlah data yang digunakan adalah 300 data.

### B. Pre-processing Data

*Pre-processing* merupakan langkah awal yang dilakukan dalam pengolahan data untuk membantu metode yang digunakan agar dapat berjalan dengan baik dan nantinya menghasilkan faktor kesalahan *Root Mean Squared Error* (RMSE) yang rendah. Pada tahap *pre-processing*, dilakukan proses *cleaning* data yang digunakan untuk menghilangkan informasi yang tidak diperlukan dalam proses *pre-processing*, yaitu menghilangkan atribut Nama Petani.

### C. Penerapan Algoritme Linear Regression

Regresi merupakan teknik membangun model yang digunakan untuk prediksi nilai dari data masukan yang diberikan. Regresi adalah ukuran statistik yang digunakan untuk menentukan kekuatan hubungan antara variabel dependen (tak bebas) dengan variabel independen (bebas).

Metode utama untuk melakukan prediksi yakni membangun model regresi dengan mencari hubungan antara satu atau lebih variabel independen atau prediktor ( $X$ ) dengan variabel dependen atau respons ( $Y$ ). *Linear regression* memodelkan hubungan antara variabel skalar dan satu atau lebih variabel penjelas.

Metode *linear regression* tersusun atas dasar pola hubungan data yang relevan di masa lalu [10]. Secara umum, algoritme *linear regression* dibagi menjadi dua jenis, yaitu *simple linear regression* dan *multiple linear regression*. *Simple linear regression* merupakan hubungan antara satu variabel dependen dengan satu variabel independen, sedangkan *multiple linear regression* merupakan hubungan antara satu variabel dependen dengan dua atau lebih variabel independen.

Dalam makalah ini, digunakan *multiple linear regression* karena terdapat lebih dari satu buah variabel independen. *Dataset* penelitian terdiri atas satu variabel dependen ( $Y$ ) dan empat variabel independen ( $X$ ). Variabel dependen tersebut adalah atribut Hasil Panen, sedangkan variabel independen adalah atribut Luas Lahan, Jumlah Bibit, Pupuk Urea, dan Pupuk NPK Phonska.

Perhitungan yang digunakan untuk *multiple linear regression* dinyatakan dalam (1).

$$Y = a + a_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

dengan

$Y$  = variabel dependen atau variabel tak bebas  
(nilai yang diprediksikan)

$X_1, X_2, \dots, X_n$  = variabel independen atau variabel bebas

$a$  = konstanta

$b_1, b_2, \dots, b_n$  = koefisien regresi.

Dengan mengacu pada (1), untuk menghitung hasil panen tanaman padi dengan algoritme *linear regression*, digunakan (2).

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 \quad (2)$$

dengan

- $Y$  = hasil panen
- $a$  = konstanta
- $b_1$  = koefisien variabel Luas Lahan
- $X_1$  = Luas Lahan
- $b_2$  = koefisien variabel Jumlah Bibit
- $X_2$  = Jumlah Bibit
- $b_3$  = koefisien variabel Pupuk Urea
- $X_3$  = Pupuk Urea
- $b_4$  = koefisien variabel Pupuk NPK Phonska
- $X_4$  = Pupuk NPK Phonska.

Langkah awal yang dilakukan adalah pembentukan model, dengan mencari nilai  $a, b_1, b_2, \dots, b_n$  menggunakan kuadrat terkecil dengan persamaan umum metode kuadrat terkecil, mengacu pada (3).

$$\begin{aligned}
 a_n + b_1 \sum X_1 + b_2 \sum X_2 + b_3 \sum X_3 + b_4 \sum X_4 &= \sum Y \\
 a \sum X_1 + b_1 \sum (X_1)^2 + b_2 \sum (X_1 X_2) + b_3 \sum (X_1 X_3) \\
 + b_4 \sum (X_1 X_4) &= \sum X_1 Y \\
 a \sum X_2 + b_1 \sum (X_1 X_2) + b_2 \sum (X_2)^2 + b_3 \sum (X_2 X_3) \\
 + b_4 \sum (X_2 X_4) &= \sum X_2 Y \quad (3) \\
 a \sum X_3 + b_1 \sum (X_1 X_3) + b_2 \sum (X_2 X_3) + b_3 \sum (X_3)^2 \\
 + b_4 \sum (X_3 X_4) &= \sum X_3 Y \\
 a \sum X_4 + b_1 \sum (X_1 X_4) + b_2 \sum (X_2 X_4) + b_3 \sum (X_3 X_4) \\
 + b_4 \sum (X_4)^2 &= \sum X_4 Y.
 \end{aligned}$$

Setelah hasil inversi diketahui, selanjutnya dilakukan perkalian matriks determinan dengan  $\sum Y, \sum X_1 Y, \sum X_2 Y, \sum X_3 Y, \sum X_4 Y$ . Untuk menghitung determinan matriks  $A, A_0, A_1, A_2, A_3$ , dan  $A_4$ , digunakan perhitungan berdasarkan (4) berikut.

$$\begin{bmatrix} N & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 X_2 & \sum X_1 X_3 & \sum X_1 X_4 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 X_2 & \sum X_2 X_3 & \sum X_2 X_4 \\ \sum X_3 & \sum X_3 X_1 & \sum X_3 X_2 & \sum X_3 X_3 & \sum X_3 X_4 \\ \sum X_4 & \sum X_4 X_1 & \sum X_4 X_2 & \sum X_4 X_3 & \sum X_4 X_4 \end{bmatrix} \times \begin{bmatrix} a \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \\ \sum X_3 Y \\ \sum X_4 Y \end{bmatrix} \quad (4)$$

Selanjutnya, hasil perhitungan determinan matriks digunakan untuk mencari nilai  $a, b_1, b_2, b_3$ , dan  $b_4$  dengan mengacu pada (5).

$$\begin{aligned}
 a &= \frac{\text{Det}(A_0)}{\text{Det}(A)} \\
 b_1 &= \frac{\text{Det}(A_1)}{\text{Det}(A)} \\
 b_2 &= \frac{\text{Det}(A_2)}{\text{Det}(A)} \\
 b_3 &= \frac{\text{Det}(A_3)}{\text{Det}(A)} \\
 b_4 &= \frac{\text{Det}(A_4)}{\text{Det}(A)}. \quad (5)
 \end{aligned}$$

Selanjutnya, dilakukan perhitungan uji korelasi parsial untuk mengetahui tingkat keterkaitan masing-masing variabel independen terhadap variabel dependen. Keterkaitan nilai yang dicari adalah nilai  $rX_1Y, rX_2Y, rX_3Y, rX_4Y, rX_1X_2, rX_1X_3, rX_1X_4,$

$rX_2X_3, rX_2X_4,$  dan  $rX_3X_4$ . Perhitungan nilai korelasi dilakukan menggunakan (6).

$$\begin{aligned}
 \sum X_1 Y &= \sum X_1 Y - \frac{(\sum X_1) \cdot (\sum Y)}{n} \\
 \sum X_2 Y &= \sum X_2 Y - \frac{(\sum X_2) \cdot (\sum Y)}{n} \\
 \sum X_3 Y &= \sum X_3 Y - \frac{(\sum X_3) \cdot (\sum Y)}{n} \\
 \sum X_4 Y &= \sum X_4 Y - \frac{(\sum X_4) \cdot (\sum Y)}{n} \\
 \sum X_1 X_2 &= \sum X_1 X_2 - \frac{(\sum X_1) \cdot (\sum X_2)}{n} \\
 \sum X_1 X_3 &= \sum X_1 X_3 - \frac{(\sum X_1) \cdot (\sum X_3)}{n} \\
 \sum X_1 X_4 &= \sum X_1 X_4 - \frac{(\sum X_1) \cdot (\sum X_4)}{n} \\
 \sum X_2 X_3 &= \sum X_2 X_3 - \frac{(\sum X_2) \cdot (\sum X_3)}{n} \\
 \sum X_2 X_4 &= \sum X_2 X_4 - \frac{(\sum X_2) \cdot (\sum X_4)}{n}. \quad (6)
 \end{aligned}$$

Selanjutnya, dilakukan perhitungan uji koefisien determinasi untuk mengetahui besar pengaruh antara variabel independen terhadap variabel dependen, sehingga diketahui kecocokan model *linear regression*. Koefisien determinasi berkisar antara 0 sampai 1. Jika  $R^2 = 0$ , maka tidak ada pengaruh antara variabel independen dengan variabel dependen, dan jika  $R^2$  semakin mendekati nilai 1, maka semakin kuat pengaruh variabel independen terhadap variabel dependen. Perhitungan untuk mencari  $R^2$  mengacu pada (7).

$$R^2 = 1 - \frac{SS \text{ error}}{SS \text{ total}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (7)$$

dengan

- $y_i$  = observasi respons ke- $i$
- $\bar{y}$  = rata-rata
- $\hat{y}_i$  = ramalan respons ke- $i$ .

Sedangkan koefisien determinasi dapat dihitung dengan mengacu pada (8).

$$Kd = R^2 \times 100\% \quad (8)$$

dengan

- $Kd$  = besar atau jumlah koefisien determinasi
- $R^2$  = nilai koefisien korelasi.

#### D. Validasi Hasil

Evaluasi dilakukan menggunakan bantuan perangkat lunak RapidMiner Studio yang digunakan untuk menguji dan melihat tingkat akurasi algoritme *linear regression* terhadap pemrosesan data. Dalam pengujian metode digunakan *cross validation* untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan *dataset* tertentu. *Cross validation* adalah cara untuk menemukan parameter terbaik dari suatu model dengan menguji besarnya *error* pada data uji [11]. *K-Fold cross validation* adalah metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan melakukan

TABEL I  
SKEMA 10-FOLD CROSS VALIDATION

Fold ke-	Fold1	Fold2	Fold3	Fold4	Fold5	...	Fold1 0
Fold1	Data ke-1-30	Data ke-31-60	Data ke-61-90	Data ke-91-120	Data ke-121-150	...	Data ke-271-300
Fold2	Data ke-1-30	Data ke-31-60	Data ke-61-90	Data ke-91-120	Data ke-121-150	...	Data ke-271-300
Fold3	Data ke-1-30	Data ke-31-60	Data ke-61-90	Data ke-91-120	Data ke-121-150	...	Data ke-271-300
Fold4	Data ke-1-30	Data ke-31-60	Data ke-61-90	Data ke-91-120	Data ke-121-150	...	Data ke-271-300
Fold5	Data ke-1-30	Data ke-31-60	Data ke-61-90	Data ke-91-120	Data ke-121-150	...	Data ke-271-300
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Fold1 0	Data ke-1-30	Data ke-31-60	Data ke-61-90	Data ke-91-120	Data ke-121-150	...	Data ke-271-300

perulangan dengan mengacak atribut masukan untuk menguji atribut *input* yang dimasukkan. *K-Fold cross validation* dilakukan dengan membagi data sejumlah *n-fold*. Pada makalah ini digunakan perhitungan *K-Fold* hingga *10-fold* dikarenakan banyak hasil eksperimen yang menunjukkan bahwa *10-fold* merupakan pilihan terbaik untuk mendapatkan estimasi yang akurat. Pada Tabel I berikut diperlihatkan *10-fold cross validation* dengan data uji ditampilkan berwarna biru, sedangkan data latih ditampilkan dengan warna putih.

Pengujian akurasi dilakukan dengan melihat perkembangan nilai RMSE dengan mengacu pada (9). RMSE merupakan metode untuk mengevaluasi hasil teknik peramalan yang digunakan dengan mengukur tingkat akurasi dari hasil prakiraan suatu model [12].

$$RMSE = \frac{\sum_{i=1}^n \sqrt{(y_i - \hat{y}_i)^2}}{n} \tag{9}$$

dengan

- $Y_i$  = data awal
- $\hat{Y}_i$  = data akhir
- $n$  = jumlah data.

RMSE merupakan nilai rata-rata dari jumlah kuadrat kesalahan yang dapat menyatakan ukuran besarnya kesalahan yang dihasilkan oleh suatu model prakiraan. Nilai RMSE yang rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai aslinya [13]. Sedangkan apabila nilai RMSE semakin besar, maka keakuratan suatu nilai yang dihasilkan semakin sedikit atau tidak akurat

### III. HASIL DAN PEMBAHASAN

Terdapat *dataset* petani sebanyak 300 data yang diperoleh melalui survei. Tabel II menunjukkan sampel *dataset* petani yang berisi atribut Nama Petani, Luas Lahan, Jumlah Bibit, Pupuk Urea, Pupuk NPK Phonska, dan Hasil Panen.

Pada tahap *pre-processing* dilakukan proses *cleaning* data. Proses *cleaning* data digunakan untuk menghilangkan atribut data yang tidak dibutuhkan dalam proses *pre-processing*, yaitu dengan menghilangkan atau membuang informasi atribut Nama Petani yang tidak diperlukan dalam pengolahan data.

TABEL II  
SAMPEL DATASET PETANI

Nama Petani	Luas Lahan	Jumlah Bibit	Pupuk Urea	Pupuk NPK Phonska	Hasil Panen
Abdul Aziz	1,7	80	340	340	8,0
Abdul K	0,3	25	60	60	1,5
Abdul R	0,4	25	80	80	2,0
Abdul W	0,2	15	40	40	1,4
Abdul Yati	0,6	40	120	120	3,7
Achmad	0,4	25	80	80	2,0
Akhsanul F	0,7	40	140	140	4,0
Alimin	0,3	25	60	60	1,5
Arjo	0,9	50	180	180	6,0
Asim	1,3	60	260	250	8,0
Asim A	0,7	40	140	140	3,8

TABEL III  
HASIL PRE-PROCESSING DATA

Luas Lahan	Jumlah Bibit	Pupuk Urea	Pupuk NPK Phonska	Hasil Panen
1,7	80	340	340	8,0
0,3	25	60	60	1,5
0,4	25	80	80	2,0
0,2	15	40	40	1,4
0,6	40	120	120	3,7
0,4	25	80	80	2,0
0,7	40	140	140	4,0
0,3	25	60	60	1,5
0,9	50	180	180	6,0
1,3	60	260	250	8,0
0,7	40	140	140	3,8

Tabel III menampilkan hasil dari *pre-processing* data, meliputi atribut Luas Lahan, Jumlah Bibit, Pupuk Urea, Pupuk NPK Phonska, dan Hasil Panen.

Hasil perhitungan algoritme *linear regression* dengan bentuk model *multiple linear regression* menyatakan hubungan antara Luas Lahan, Jumlah Bibit, Pupuk Urea, Pupuk NPK Phonska, dan Hasil Panen dengan mengacu pada (10).

$$Y = a + b_1 (\text{Luas Lahan}) + b_2 (\text{Jumlah Bibit}) + b_3 (\text{Pupuk Urea}) + b_4 (\text{Pupuk NPK Phonska}) \tag{10}$$

Langkah pertama untuk mendapatkan persamaan *linear regression* berganda adalah melakukan pembentukan model *multiple linear regression*.

#### A. Multiple Linear Regression

Nilai  $a$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , dan  $b_4$  merupakan nilai konstanta dan koefisien regresi yang dapat diperoleh dengan menggunakan perhitungan matriks determinan. Pada penelitian ini, terdapat persamaan variabel yang tidak diketahui nilainya, yaitu  $a$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , dan  $b_4$ . Persamaan tersebut dapat dinyatakan dalam persamaan matriks seperti ditunjukkan pada (11).

$$A = \begin{bmatrix} N & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 X_2 & \sum X_1 X_3 & \sum X_1 X_4 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 X_2 & \sum X_2 X_3 & \sum X_2 X_4 \\ \sum X_3 & \sum X_3 X_1 & \sum X_3 X_2 & \sum X_3 X_3 & \sum X_3 X_4 \\ \sum X_4 & \sum X_4 X_1 & \sum X_4 X_2 & \sum X_4 X_3 & \sum X_4 X_4 \end{bmatrix} \tag{11}$$

A merupakan persamaan matriks yang digunakan untuk mencari nilai  $Det(A)$ , dengan  $N$  adalah banyaknya data,  $\sum X_1$  adalah jumlah dari variabel Luas Lahan,  $\sum X_2$  adalah jumlah dari variabel Jumlah Bibit,  $\sum X_3$  adalah jumlah dari variabel Pupuk Urea, dan  $\sum X_4$  adalah jumlah dari variabel Pupuk NPK Phonska.

$$A_0 = \begin{bmatrix} \sum Y & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 \\ \sum X_1 Y & \sum X_1 X_1 & \sum X_1 X_2 & \sum X_1 X_3 & \sum X_1 X_4 \\ \sum X_2 Y & \sum X_2 X_1 & \sum X_2 X_2 & \sum X_2 X_3 & \sum X_2 X_4 \\ \sum X_3 Y & \sum X_3 X_1 & \sum X_3 X_2 & \sum X_3 X_3 & \sum X_3 X_4 \\ \sum X_4 Y & \sum X_4 X_1 & \sum X_4 X_2 & \sum X_4 X_3 & \sum X_4 X_4 \end{bmatrix} \quad (12)$$

$A_0$  merupakan persamaan matriks yang digunakan untuk mencari nilai  $Det(A_0)$  yang dihitung dengan mengacu pada (12), dengan  $\sum Y$  adalah jumlah dari variabel Hasil Panen,  $\sum X_1$  adalah jumlah dari variabel Luas Lahan,  $\sum X_2$  adalah jumlah dari variabel Jumlah Bibit,  $\sum X_3$  adalah jumlah dari variabel Pupuk Urea, dan  $\sum X_4$  adalah jumlah dari variabel Pupuk NPK Phonska.

$$A_1 = \begin{bmatrix} N & \sum Y & \sum X_2 & \sum X_3 & \sum X_4 \\ \sum X_1 & \sum X_1 Y & \sum X_1 X_2 & \sum X_1 X_3 & \sum X_1 X_4 \\ \sum X_2 & \sum X_2 Y & \sum X_2 X_2 & \sum X_2 X_3 & \sum X_2 X_4 \\ \sum X_3 & \sum X_3 Y & \sum X_3 X_2 & \sum X_3 X_3 & \sum X_3 X_4 \\ \sum X_4 & \sum X_4 Y & \sum X_4 X_2 & \sum X_4 X_3 & \sum X_4 X_4 \end{bmatrix} \quad (13)$$

$A_1$  merupakan persamaan matriks yang digunakan untuk mencari nilai  $Det(A_1)$  yang dihitung dengan mengacu pada (13), dengan  $N$  adalah banyaknya data,  $\sum Y$  adalah jumlah dari variabel Hasil Panen,  $\sum X_2$  adalah jumlah dari variabel Jumlah Bibit,  $\sum X_3$  adalah jumlah dari variabel Pupuk Urea, dan  $\sum X_4$  adalah jumlah dari variabel Pupuk NPK Phonska.

$$A_2 = \begin{bmatrix} N & \sum X_1 & \sum Y & \sum X_3 & \sum X_4 \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 Y & \sum X_1 X_3 & \sum X_1 X_4 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 Y & \sum X_2 X_3 & \sum X_2 X_4 \\ \sum X_3 & \sum X_3 X_1 & \sum X_3 Y & \sum X_3 X_3 & \sum X_3 X_4 \\ \sum X_4 & \sum X_4 X_1 & \sum X_4 Y & \sum X_4 X_3 & \sum X_4 X_4 \end{bmatrix} \quad (14)$$

$A_2$  merupakan persamaan matriks yang digunakan untuk mencari nilai  $Det(A_2)$  yang dihitung dengan mengacu pada (14), dengan  $N$  adalah banyaknya data,  $\sum X_1$  adalah jumlah dari variabel Luas Lahan,  $\sum Y$  adalah jumlah dari variabel Hasil Panen,  $\sum X_3$  adalah jumlah dari variabel Pupuk Urea, dan  $\sum X_4$  adalah jumlah dari variabel Pupuk NPK Phonska.

$$A_3 = \begin{bmatrix} N & \sum X_1 & \sum X_2 & \sum Y & \sum X_4 \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 X_2 & \sum X_1 Y & \sum X_1 X_4 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 X_2 & \sum X_2 Y & \sum X_2 X_4 \\ \sum X_3 & \sum X_3 X_1 & \sum X_3 X_2 & \sum X_3 Y & \sum X_3 X_4 \\ \sum X_4 & \sum X_4 X_1 & \sum X_4 X_2 & \sum X_4 Y & \sum X_4 X_4 \end{bmatrix} \quad (15)$$

$$A_4 = \begin{bmatrix} N & \sum X_1 & \sum X_2 & \sum X_3 & \sum Y \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 X_2 & \sum X_1 X_3 & \sum X_1 Y \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 X_2 & \sum X_2 X_3 & \sum X_2 Y \\ \sum X_3 & \sum X_3 X_1 & \sum X_3 X_2 & \sum X_3 X_3 & \sum X_3 Y \\ \sum X_4 & \sum X_4 X_1 & \sum X_4 X_2 & \sum X_4 X_3 & \sum X_4 Y \end{bmatrix} \quad (16)$$

$A_3$  merupakan persamaan matriks yang digunakan untuk mencari nilai  $Det(A_3)$  yang dihitung dengan mengacu pada (15), dengan  $N$  adalah banyaknya data,  $\sum X_1$  adalah jumlah dari variabel Luas Lahan,  $\sum X_2$  adalah jumlah dari variabel Jumlah

Bibit,  $\sum Y$  adalah jumlah variabel Hasil Panen,  $\sum X_4$  adalah jumlah dari variabel Pupuk NPK Phonska.

$A_4$  merupakan persamaan matriks yang digunakan untuk mencari nilai  $Det(A_4)$  yang dihitung dengan mengacu pada (16), dengan  $N$  adalah banyaknya data,  $\sum X_1$  adalah jumlah variabel Luas Lahan,  $\sum X_2$  adalah jumlah variabel Jumlah Bibit,  $\sum X_3$  adalah jumlah variabel dari Pupuk Urea, dan  $\sum Y$  adalah jumlah dari variabel Hasil Panen. Langkah selanjutnya adalah perhitungan untuk menentukan determinasi matriks  $A, A_0, A_1, A_2, A_3,$  dan  $A_4$  sebagai berikut.

$$Det(A) = \{N \cdot \sum(X_1 X_1) \cdot \sum(X_2 X_2) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_4)\} + \{\sum X_1 \cdot \sum(X_1 X_2) \cdot \sum(X_2 X_3) \cdot \sum(X_3 X_4) \cdot \sum X_4\} + \{\sum X_2 \cdot \sum(X_1 X_3) \cdot \sum(X_2 X_4) \cdot \sum X_3 \cdot \sum(X_4 X_1)\} + \{\sum X_3 \cdot \sum(X_1 X_4) \cdot \sum X_2 \cdot \sum(X_3 X_1) \cdot \sum(X_4 X_2)\} + \{\sum X_4 \cdot \sum X_1 \cdot \sum(X_2 X_1) \cdot \sum(X_3 X_2) \cdot \sum(X_4 X_3)\} - \{\sum X_4 \cdot \sum(X_1 X_3) \cdot \sum(X_2 X_2) \cdot \sum(X_3 X_1) \cdot \sum X_4\} - \{\sum X_3 \cdot \sum(X_1 X_2) \cdot \sum(X_2 X_1) \cdot \sum X_3 \cdot \sum(X_4 X_4)\} - \{\sum X_2 \cdot \sum(X_1 X_1) \cdot \sum X_2 \cdot \sum(X_3 X_4) \cdot \sum(X_4 X_3)\} - \{\sum X_1 \cdot \sum X_1 \cdot \sum(X_2 X_4) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_2)\} - \{N \cdot \sum(X_1 X_4) \cdot \sum(X_2 X_3) \cdot \sum(X_3 X_2) \cdot \sum(X_4 X_1)\}$$

$$Det(A_0) = \{\sum Y \cdot \sum(X_1 X_1) \cdot \sum(X_2 X_2) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_4)\} + \{\sum X_1 \cdot \sum(X_1 X_2) \cdot \sum(X_2 X_3) \cdot \sum(X_3 X_4) \cdot \sum X_4 Y\} + \{\sum X_2 \cdot \sum(X_1 X_3) \cdot \sum(X_2 X_4) \cdot \sum(X_3 \cdot Y) \cdot \sum(X_4 X_1)\} + \{\sum X_3 \cdot \sum(X_1 X_4) \cdot \sum X_2 Y \cdot \sum(X_3 X_1) \cdot \sum(X_4 X_2)\} + \{\sum X_4 \cdot \sum X_1 Y \cdot \sum(X_2 X_1) \cdot \sum(X_3 X_2) \cdot \sum(X_4 X_3)\} - \{\sum X_4 \cdot \sum(X_1 X_3) \cdot \sum(X_2 X_2) \cdot \sum(X_3 X_1) \cdot \sum X_4\} - \{\sum X_3 \cdot \sum(X_1 X_2) \cdot \sum(X_2 X_1) \cdot \sum X_3 \cdot \sum(X_4 Y)\} - \{\sum X_2 \cdot \sum(X_1 X_1) \cdot \sum(X_2 Y) \cdot \sum(X_3 X_4) \cdot \sum(X_4 X_3)\} - \{\sum X_1 \cdot \sum(X_1 Y) \cdot \sum(X_2 X_4) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_2)\} - \{\sum Y \cdot \sum(X_1 X_4) \cdot \sum(X_2 X_3) \cdot \sum(X_3 X_2) \cdot \sum(X_4 X_1)\}$$

$$Det(A_1) = \{\sum N \cdot \sum(X_1 Y) \cdot \sum(X_2 X_2) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_4)\} + \{\sum Y \cdot \sum(X_1 X_2) \cdot \sum(X_2 X_3) \cdot \sum(X_3 X_4) \cdot \sum X_4\} + \{\sum X_2 \cdot \sum(X_1 X_3) \cdot \sum(X_2 X_4) \cdot \sum X_3 \cdot \sum(X_4 Y)\} + \{\sum X_3 \cdot \sum(X_1 X_4) \cdot \sum X_2 \cdot \sum(X_3 Y) \cdot \sum(X_4 X_2)\} + \{\sum X_4 \cdot \sum X_1 \cdot \sum(X_2 Y) \cdot \sum(X_3 X_2) \cdot \sum(X_4 X_3)\} - \{\sum X_4 \cdot \sum(X_1 X_3) \cdot \sum(X_2 X_2) \cdot \sum(X_3 Y) \cdot \sum X_4\} - \{\sum X_3 \cdot \sum(X_1 X_2) \cdot \sum(X_2 Y) \cdot \sum X_3 \cdot \sum(X_4 X_4)\} - \{\sum X_2 \cdot \sum(X_1 Y) \cdot \sum X_2 \cdot \sum(X_3 X_4) \cdot \sum(X_4 X_3)\} - \{\sum Y \cdot \sum X_1 \cdot \sum(X_2 X_4) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_2)\} - \{\sum N \cdot \sum(X_1 X_4) \cdot \sum(X_2 X_3) \cdot \sum(X_3 X_2) \cdot \sum(X_4 X_1)\}$$

$$Det(A_2) = \{\sum N \cdot \sum(X_1 X_1) \cdot \sum(X_2 Y) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_4)\} + \{\sum X_1 \cdot \sum(X_1 Y) \cdot \sum(X_2 X_3) \cdot \sum(X_3 X_4) \cdot \sum X_4\} + \{\sum Y \cdot \sum(X_1 X_3) \cdot \sum(X_2 X_4) \cdot \sum X_3 \cdot \sum(X_4 X_1)\} + \{\sum X_3 \cdot \sum(X_1 X_4) \cdot \sum X_2 \cdot \sum(X_3 X_1) \cdot \sum(X_4 Y)\} + \{\sum X_4 \cdot \sum X_1 \cdot \sum(X_2 X_1) \cdot \sum(X_3 Y) \cdot \sum(X_4 X_3)\} - \{\sum X_4 \cdot \sum(X_1 X_3) \cdot \sum(X_2 Y) \cdot \sum(X_3 X_1) \cdot \sum X_4\} - \{\sum X_3 \cdot \sum(X_1 Y) \cdot \sum(X_2 X_1) \cdot \sum X_3 \cdot \sum(X_4 X_4)\} - \{\sum Y \cdot \sum(X_1 X_1) \cdot \sum X_2 \cdot \sum(X_3 X_4) \cdot \sum(X_4 X_3)\} - \{\sum X_1 \cdot \sum X_1 \cdot \sum(X_2 X_4) \cdot \sum(X_3 X_3) \cdot \sum(X_4 X_2)\} - \{\sum N \cdot \sum(X_1 X_4) \cdot \sum(X_2 X_3) \cdot \sum(X_3 Y) \cdot \sum(X_4 X_1)\}$$

$$Det(A_3) = \{\sum N \cdot \sum(X_1 X_1) \cdot \sum(X_2 X_2) \cdot \sum(X_3 Y) \cdot \sum(X_4 X_4)\} + \{\sum X_1 \cdot \sum(X_1 X_2) \cdot \sum(X_2 Y) \cdot \sum(X_3 X_4) \cdot \sum X_4\} + \{\sum X_2 \cdot \sum(X_1 Y) \cdot \sum(X_2 X_4) \cdot \sum X_3 \cdot \sum(X_4 X_1)\} + \{\sum Y \cdot \sum(X_1 X_4) \cdot \sum X_2 \cdot \sum(X_3 X_1) \cdot \sum(X_4 X_2)\} + \{\sum X_4 \cdot \sum X_1 \cdot \sum(X_2 X_1) \cdot \sum(X_3 X_2) \cdot \sum(X_4 Y)\} - \{\sum X_4 \cdot \sum(X_1 Y) \cdot \sum(X_2 X_2) \cdot \sum(X_3 X_1) \cdot \sum X_4\} - \{\sum Y \cdot \sum(X_1 X_2) \cdot \sum(X_2 X_1) \cdot \sum X_3 \cdot \sum(X_4 X_4)\} - \{\sum X_2 \cdot \sum(X_1 X_1) \cdot \sum X_2 \cdot \sum(X_3 X_4) \cdot \sum(X_4 Y)\} - \{\sum X_1 \cdot \sum X_1 \cdot \sum(X_2 X_4) \cdot \sum(X_3 Y) \cdot \sum(X_4 X_2)\} - \{\sum N \cdot \sum(X_1 X_4) \cdot \sum(X_2 Y) \cdot \sum(X_3 X_2) \cdot \sum(X_4 X_1)\}$$

TABEL IV  
HASIL PERHITUNGAN

$a$	$b_1$	$b_2$	$b_3$	$b_4$
-0, 0,72368	5,684211	0,914474	0,776316	2,980263

$$\begin{aligned} \text{Det (A}_4) = & \{ \sum N \cdot \sum (X_1 X_1) \cdot \sum (X_2 X_2) \cdot \sum (X_3 X_3) \cdot \sum (X_4 X_4) \} + \\ & \{ \sum X_1 \cdot \sum (X_1 X_2) \cdot \sum (X_2 X_3) \cdot \sum (X_3 Y) \cdot \sum X_4 \} + \\ & \{ \sum X_2 \cdot \sum (X_1 X_3) \cdot \sum (X_2 Y) \cdot \sum X_3 \cdot \sum (X_4 X_1) \} + \\ & \{ \sum X_3 \cdot \sum (X_1 Y) \cdot \sum X_2 \cdot \sum (X_3 X_1) \cdot \sum (X_4 X_2) \} + \\ & \{ \sum Y \cdot \sum X_1 \cdot \sum (X_2 X_1) \cdot \sum (X_3 X_2) \cdot \sum (X_4 X_3) \} - \\ & \{ \sum Y \cdot \sum (X_1 X_3) \cdot \sum (X_2 X_2) \cdot \sum (X_3 X_1) \cdot \sum X_4 \} - \\ & \{ \sum X_3 \cdot \sum (X_1 X_2) \cdot \sum (X_2 X_1) \cdot \sum X_3 \cdot \sum (X_4 Y) \} - \\ & \{ \sum X_2 \cdot \sum (X_1 X_1) \cdot \sum X_2 \cdot \sum (X_3 Y) \cdot \sum (X_4 X_3) \} - \\ & \{ \sum X_1 \cdot \sum X_1 \cdot \sum (X_2 Y) \cdot \sum (X_3 X_3) \cdot \sum (X_4 X_2) \} - \\ & \{ \sum N \cdot \sum (X_1 Y) \cdot \sum (X_2 X_3) \cdot \sum (X_3 X_2) \cdot \sum (X_4 X_1) \} \end{aligned}$$

Berdasarkan perhitungan tersebut, didapatkan hasil nilai  $a$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , dan  $b_4$ , seperti ditunjukkan pada Tabel IV. Berdasarkan hasil yang diperoleh untuk koefisien  $a$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , dan  $b_4$  tersebut, hasil model *multiple linear regression* menjadi sebagai berikut.

$$\begin{aligned} \Sigma Y = & -0, 0,72368 + 5,684211 \Sigma X_1 + 0,914474 \Sigma X_2 + \\ & 0,776316 \Sigma X_3 + 2,980263 \Sigma X_4 \end{aligned}$$

### B. Uji Korelasi Parsial

Perhitungan uji korelasi parsial digunakan untuk mengetahui besar keterkaitan masing-masing variabel independen terhadap variabel dependen. Untuk menghitung nilai korelasi, diperlukan perhitungan untuk mengetahui nilai  $rX_1Y$ ,  $rX_2Y$ ,  $rX_3Y$ ,  $rX_4Y$ ,  $rX_1X_2$ ,  $rX_1X_3$ ,  $rX_1X_4$ ,  $rX_2X_3$ ,  $rX_2X_4$ , dan  $rX_3X_4$  sebagai berikut.

Perhitungan  $rX_1Y$  :

$$\begin{aligned} \Sigma X_1 Y = & \Sigma X_1 Y - \frac{(\Sigma X_1) \cdot (\Sigma Y)}{n} \\ = & 800,64 - \frac{(185,4) \cdot (1.058,3)}{300} \\ = & 146,6106 \end{aligned}$$

Perhitungan  $rX_2Y$  :

$$\begin{aligned} \Sigma X_2 Y = & \Sigma X_2 Y - \frac{(\Sigma X_2) \cdot (\Sigma Y)}{n} \\ = & 43.388 - \frac{(10.425) \cdot (1.058,3)}{300} \\ = & 6.612,075 \end{aligned}$$

Perhitungan  $rX_3Y$  :

$$\begin{aligned} \Sigma X_3 Y = & \Sigma X_3 Y - \frac{(\Sigma X_3) \cdot (\Sigma Y)}{n} \\ = & 160280 - \frac{(37.120) \cdot (1.058,3)}{300} \\ = & 29.333,013 \end{aligned}$$

Perhitungan  $rX_4Y$  :

$$\Sigma X_4 Y = \Sigma X_4 Y - \frac{(\Sigma X_4) \cdot (\Sigma Y)}{n}$$

$$\begin{aligned} = & 160.302 - \frac{(37.140) \cdot (1.058,3)}{300} \\ = & 29.284,46 \end{aligned}$$

Perhitungan  $rX_1X_2$  :

$$\begin{aligned} \Sigma X_1 X_2 = & \Sigma X_1 X_2 - \frac{(\Sigma X_1) \cdot (\Sigma X_2)}{n} \\ = & 74.95,5 - \frac{(185,4) \cdot (10.425)}{300} \\ = & 1.052,85 \end{aligned}$$

Perhitungan  $rX_1X_3$  :

$$\begin{aligned} \Sigma X_1 X_3 = & \Sigma X_1 X_3 - \frac{(\Sigma X_1) \cdot (\Sigma X_3)}{n} \\ = & 27.996 - \frac{(185,4) \cdot (37.120)}{300} \\ = & 5.055,84 \end{aligned}$$

Perhitungan  $rX_1X_4$  :

$$\begin{aligned} \Sigma X_1 X_4 = & \Sigma X_1 X_4 - \frac{(\Sigma X_1) \cdot (\Sigma X_4)}{n} \\ = & 28.004 - \frac{(185,4) \cdot (37.140)}{300} \\ = & 5.051,48 \end{aligned}$$

Perhitungan  $rX_2X_3$  :

$$\begin{aligned} \Sigma X_2 X_3 = & \Sigma X_2 X_3 - \frac{(\Sigma X_2) \cdot (\Sigma X_3)}{n} \\ = & 1.500.700 - \frac{(10.425) \cdot (37.120)}{300} \\ = & 210.780 \end{aligned}$$

Perhitungan  $rX_2X_4$  :

$$\begin{aligned} \Sigma X_2 X_4 = & \Sigma X_2 X_4 - \frac{(\Sigma X_2) \cdot (\Sigma X_4)}{n} \\ = & 1.501.300 - \frac{(10.425) \cdot (37.140)}{300} \\ = & 210.685 \end{aligned}$$

Perhitungan  $rX_3X_4$  :

$$\begin{aligned} \Sigma X_3 X_4 = & \Sigma X_3 X_4 - \frac{(\Sigma X_3) \cdot (\Sigma X_4)}{n} \\ = & 5.606.400 - \frac{(37.120) \cdot (37.140)}{300} \\ = & 1.010.944 \end{aligned}$$

Hasil perhitungan uji korelasi parsial untuk model *multiple linear regression* dari Luas Lahan, Jumlah Bibit, Pupuk Urea, Pupuk NPK Phonska, dan Hasil Panen disajikan pada Tabel V.

### C. Uji Koefisien Determinasi

Perhitungan uji koefisien determinasi digunakan untuk mengetahui besar pengaruh antara variabel bebas terhadap variabel tak bebas, sehingga dapat diketahui kecocokan model

TABEL V  
HASIL UJI KORELASI PARSIAL

Keterkaitan Variabel	Nilai Korelasi
$r_{X_1Y}$	146,6106
$r_{X_2Y}$	6.612,0750
$r_{X_3Y}$	29.333,0130
$r_{X_4Y}$	29.284,4600
$r_{X_1X_2}$	1.052,8500
$r_{X_1X_3}$	5.055,8400
$r_{X_1X_4}$	5.051,4800
$r_{X_2X_3}$	210.780,0000
$r_{X_2X_4}$	210.685,0000
$r_{X_3X_4}$	1.010.944,0000

TABEL VI  
HASIL UJI KOEFISIEN DETERMINASI

R	R <sup>2</sup>
0,945873857	0,945139943

TABEL VII  
NILAI RMSE

K-Fold Validasi	Nilai RMSE
2	0,460
3	0,461
4	0,443
5	0,448
6	0,445
7	0,435
8	0,434
9	0,435
10	0,432

*linear regression*. Koefisien determinasi berkisar antara 0 sampai 1. Nilai  $R^2 = 0$  menunjukkan tidak adanya pengaruh antara variabel independen dengan variabel dependen dan jika  $R^2$  semakin mendekati 1, maka semakin kuat pengaruh variabel independen terhadap variabel dependen.

Dari hasil model pada Tabel VI terlihat hasil koefisien determinasi sebesar 0,945139943. Artinya tingkat kecocokan model *multiple linear regression* memiliki tingkat keandalan sebesar 94,51%. Sebanyak 94,51% variasi nilai Hasil Panen bergantung pada variabel independen yang diukur, yaitu meliputi Luas Lahan, Jumlah Bibit, Pupuk Urea, dan Pupuk NPK Phonska. Sedangkan sisanya, yaitu sebesar 5,49%, dipengaruhi oleh variabel lain yang tidak diukur pada makalah ini.

Validasi hasil dilakukan menggunakan perangkat lunak RapidMiner Studio dengan melihat perkembangan nilai RMSE. RMSE adalah metode untuk mengevaluasi hasil teknik peramalan yang digunakan untuk mengukur tingkat akurasi dari hasil prakiraan suatu model [12].

Tabel VII menunjukkan hasil RMSE dengan nilai *K-Fold* dari 2 sampai dengan 10. Dapat disimpulkan bahwa hasil nilai rata-rata akurasi RMSE adalah sebesar 0,432. Hal tersebut menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan yang dihasilkan mendekati akurat.

#### IV. KESIMPULAN

Berdasarkan analisis data dan pembahasan yang telah dipaparkan, dapat ditarik kesimpulan bahwa algoritme *linear regression* dapat diterapkan untuk prediksi hasil panen tanaman padi dengan beberapa variabel yang memengaruhi, yaitu Luas Lahan, Jumlah Bibit, Pupuk Urea dan Pupuk NPK Phonska.

Dari jumlah data sebanyak 300 *instance*, dihasilkan tingkat kecocokan model *multiple linear regression* sebesar 94,51%. Artinya sebanyak 94,51% variasi nilai hasil panen bergantung pada variabel bebas yang diukur, yaitu Luas Lahan, Varietas Bibit, Jumlah Bibit, Pupuk Urea, dan Pupuk NPK Phonska. Sedangkan sisanya, yaitu sebesar 5,49%, dipengaruhi oleh variabel lain yang tidak diukur pada makalah ini. Hasil nilai rata-rata akurasi RMSE adalah sebesar 0,432. Hal tersebut menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan yang dihasilkan mendekati akurat

#### REFERENSI

- [1] M. Muslihudin dan T.F. Abdillah, "Sistem Pendukung Keputusan untuk Menentukan Kualitas Bibit Padi (Kasus Petani Podosari)," *Jurnal TAM (Technology Accept. Model)*, Vol. 2, hal.26-32, Jul. 2014.
- [2] E.E. Pratiwi, A.W. Widodo, dan W.F. Mahmudy, "Penerapan Algoritme Genetika pada Kasus Optimasi Penentuan Bibit dan Pemerataan Subsidi pupuk (Studi Kasus: Desa Pandansari, Kabupaten Kediri)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 2, No. 5, hal. 1803–1812, Sep. 2017.
- [3] C.V. Donggulo, I.M. Lapanjang, dan U. Made, "Pertumbuhan dan Hasil Tanaman Padi (*Oryza sativa* L) pada Berbagai Pola Jajar Legowo dan Jarak Tanam," *J. Agrol.*, Vol. 24, hal. 27–35, April 2017.
- [4] G. Ramadhona, B.D. Setiawan, dan F.A. Bachtiar, "Prediksi Produktivitas Padi Menggunakan Jaringan Syaraf Tiruan Backpropagation," *J. Pengemb. Teknol. Inf. dan Ilmu Komputer*, Vol. 2, No. 12, hal. 6048–6057, 2018.
- [5] W.N. Tenggara, "Kajian Pengembangan Varietas Unggul Baru Padi Sawah dengan Pendekatan Pengelolaan Tanaman Terpadu di Dompu Nusa Tenggara Barat," *Penelit. Pertan. Tanam. Pangan*, Vol. 2, No. 2, hal. 95–99, 2018.
- [6] Z.A. Matondang, "Sistem Pendukung Keputusan Forecasting Harga Emas Lelang pada Pegadaian dengan Metode Single Moving Average," *J. Tek. Inform. Unika St. Thomas*, Vol. 3, No. 1, hal. 72–77, Jun. 2018.
- [7] I.M. Kamal, T. Hendro P., dan R. Ilyas, "Prediksi Penjualan Buku Menggunakan Data Mining di PT. Niaga Swadaya," *Semin. Nas. Teknol. Inf. dan Multimedia*, 2017, hal. 49–54.
- [8] F. Anis dan Suprayogi, "Estimasi Luas Panen Padi di Kabupaten Rembang Menggunakan Algoritma Linear Regression," Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia, 2015.
- [9] B. Irwan. dan N.A. Artesya, "Aplikasi Data Mining Menggunakan Multiple Linear Regression untuk Pengenalan Pola Curah Hujan," *Jurnal Ilmiah KLIK*, Vol. 2, No. 1, hal. 34–44, 2015.
- [10] M.F. Saputri dan S. Slamet, "Analisa Data Penjualan Menggunakan Metode Regresi Linier untuk Prediksi Persediaan Barang pada TB.Kawankita," Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia, 2016.
- [11] A. Fikri, "Penerapan Data Mining untuk Mengetahui Tingkat Kekuatan Beton yang Dihasilkan dengan Metode Estimasi Menggunakan Linear Regression," Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia, 2009.
- [12] T. Chai dan R.R. Draxler, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? -Arguments Against Avoiding RMSE in the Literature," *Geosci. Model Dev.*, Vol. 7, No. 3, hal. 1247–1250, 2014.
- [13] P. Choirunisa dan Kariyam, "Perbandingan Metode Triple Exponential Smoothing dan Metode Seasonal ARIMA untuk Peramalan Inflasi di Kota Tamjung Pandan," *Prosiding Sendika*, Vol. 5, No. 2, hal. 76–83, 2019.