

Prediksi Hasil Panen Padi Di Kabupaten Jembrana Dengan Metode Linear Regression

I. B. M. Swarbawa^{a1}, I. G. Arta Wibawa^{a2}, I. K. G. Suhartana^{a3}

^aInformatics Engineering, Faculty of Math and Science, University of Udayana
South Kuta, Badung, Bali, Indonesia

¹ibmswarbawa@gmail.com

²gede.arta@unud.ac.id

³ikg.suhartana@unud.ac.id

Abstract

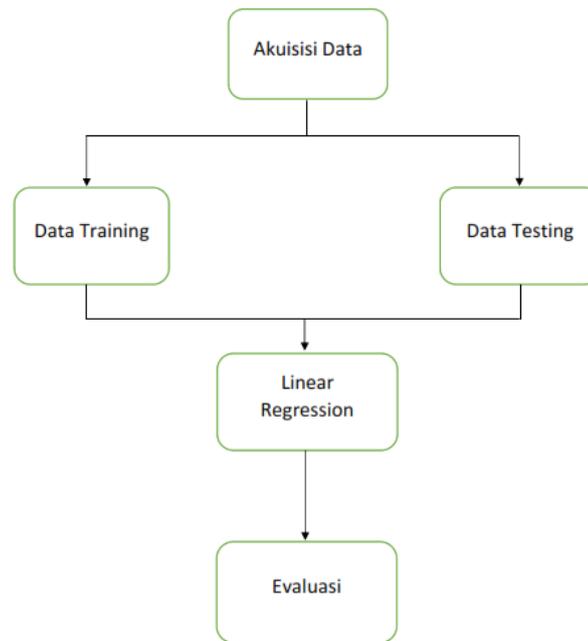
Prediction of agricultural yields is very much needed in terms of planning and decision making as well as in national food security policies. One of the strategic commodities that requires special attention is rice. This study aims to predict rice yields, especially in Jembrana Regency by using Linear Regression. Prediction is an important tool in effective and efficient planning. These results indicate that the rice yield prediction model can be used for forecasting purposes in the future. It is hoped that with this rice yield prediction system, it can help predict rice yields in the coming year and overcome the soaring increase in rice prices.

Keywords : *Yields, Prediction, Linear Regression*

1. Pendahuluan

Jembrana merupakan salah satu kabupaten penyangga pangan di provinsi Bali. Ada berbagai macam tanaman pangan yang dihasilkan di Kabupaten Jembrana setiap tahunnya salah satunya adalah padi. Padi merupakan salah satu tanaman pangan terpenting, karena padi menjadi sumber karbohidrat yang utama [1]. Kebutuhan bahan pangan terutama beras akan terus meningkat sejalan dengan pertambahan jumlah penduduk dan peningkatan konsumsi perkapita akibat peningkatan pendapatan. Hasil panen padi di kabupaten Jembrana mengalami naik turun. Sebagai upaya untuk mengatasi kenaikan harga beras yang melambung tinggi perlu adanya ramalan yang tepat tentang hasil panen padi di Kabupaten Jembrana yaitu sistem prediksi.

Prediksi merupakan proses memperkirakan sesuatu yang akan terjadi di masa depan secara sistematis berdasarkan informasi yang dimiliki di masa lalu dan sekarang [2]. Dalam penelitian ini akan digunakan metode Linear Regression untuk mendapatkan prediksi dari hasil panen padi di Kabupaten Jembrana. Algoritma 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 [3].



Gambar 1. Tahapan penelitian

Tahapan-tahapan yang dilakukan dalam penelitian adalah pengumpulan data, kemudian membagi data training dan data testing, kemudian mencari prediksi dengan regresi linier dan terakhir melakukan pengujian atau evaluasi.

2. Metode Penelitian

Dalam penelitian ini, tahap-tahap yang akan dilakukan adalah pengumpulan data, membagi data training dan data testing, penerapan algoritma linear regression, dan validasi hasil.

2.1. Pengumpulan Data

Dalam penelitian ini data yang digunakan adalah data yang berasal website Jembrana Satu Data Dari Desa atau disingkat JSDDD yang mana websitenya bernama <https://survey.jembranakab.go.id/>. Terdapat lima atribut yang digunakan dalam data yaitu nama kecamatan, luas tanam, luas panen, hasil produksi padi dan rata-rata hasil produksi padi.

2.2. Pre-Processing

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 Kecamatan

2.3. Linear Regression

Metode Prediksi yang akan digunakan pada penelitian ini adalah Linear Regression. Linear Regression digunakan untuk membangun model untuk memprediksi hubungan antara dua variabel dengan menerapkan persamaan linier pada data yang diamati.

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 [4]. Algoritma 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 penelitian ini digunakan multiple linear regression karena terdapat lebih dari satu variabel independen. Dataset penelitian terdiri dari satu variabel dependen (Y) dan tiga variabel independen (X). Variabel dependen tersebut adalah atribut Hasil Produksi Padi, sedangkan variabel independen adalah atribut Luas Tanam, Luas Panen, dan Rata-Rata Produksi Padi.

Perhitungan yang digunakan untuk multiple linear regression dinyatakan dalam persamaan 1

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

dimana

Y = variabel dependen (nilai yang akan diprediksi)

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 persamaan 1, untuk menghitung hasil panen tanaman padi dengan algoritme *linear regression*, digunakan persamaan 2

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

dimana

Y = hasil produksi

a = konstanta

b_1 = koefisien variabel Luas Tanam

X_1 = Luas Tanam

b_2 = koefisien variabel Luas Panen

X_2 = Luas Panen

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

$$\begin{aligned} a_n + b_1\sum X_1 + b_2\sum X_2 &= \sum Y \\ a\sum X_1 + b_1\sum (X_1)^2 + b_2\sum (X_1X_2) &= \sum X_1Y \\ a\sum X_2 + b_1\sum (X_1X_2) + b_2\sum (X_2)^2 &= \sum X_2Y \end{aligned} \quad (3)$$

Setelah hasil inversi diketahui, selanjutnya dilakukan perkalian matriks determinan dengan $\sum Y, \sum X_1Y, \sum X_2Y$. Untuk menghitung determinan matriks A, $A_0, A_1,$ dan A_2 , digunakan persamaan 4

$$\begin{bmatrix} N & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1X_1 & \sum X_1X_2 \\ \sum X_2 & \sum X_2X_1 & \sum X_2X_2 \end{bmatrix} \times \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix} \quad (4)$$

Selanjutnya, mencari nilai a, $b_1,$ dan b_2 dengan persamaan 5

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

Selanjutnya, dilakukan perhitungan uji korelasi parsial untuk mengetahui tingkat keterkaitan masing-masing variabel independen terhadap variabel dependen. Perhitungan korelasi didapatkan dengan persamaan 6

$$\sum X_1Y = \sum X_1Y - \frac{(\sum X_1) \cdot (\sum Y)}{n} \quad (6)$$

$$\sum X_2 Y = \sum X_2 Y - \frac{(\sum X_2) \cdot (\sum Y)}{n}$$

$$\sum X_1 X_2 = \sum X_1 X_2 - \frac{(\sum X_1) \cdot (\sum X_2)}{n}$$

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 R² = 0, maka tidak ada pengaruh antara variabel independen dengan variabel dependen, dan jika R² semakin mendekati nilai 1, maka semakin kuat pengaruh variabel independen terhadap variabel dependen. Perhitungan untuk mencari R² dapat dilihat pada persamaan 7

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

dimana

y_i = observasi respons ke-i

\bar{y} = rata-rata

\hat{y}_i = ramalan respons ke-i

Koefisien determinasi dapat dihitung dengan persamaan 8

$$R = \frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y}{\sum Y^2} \quad (8)$$

dimana

R = jumlah koefisien determinasi

2.4. Evaluasi

Evaluasi dilakukan 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 [5].

K-Fold Cross Validation adalah metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan melakukan perulangan dengan mengacak atribut masukan untuk menguji atribut input yang dimasukkan. K-Fold cross validation dilakukan dengan membagi data sejumlah n-fold.

Percobaan 1	Test	Train	Train	Train	Train
Percobaan 2	Train	Test	Train	Train	Train
Percobaan 3	Train	Train	Test	Train	Train
Percobaan 4	Train	Train	Train	Test	Train
Percobaan 5	Train	Train	Train	Train	Test

Gambar 2. K-Fold Cross Validation

Pada penelitian ini akan dilakukan perhitungan K-Fold hingga 5-fold. Pada gambar 2 diperlihatkan 5-fold cross validation dengan data testing ditampilkan berwarna kuning dan data latih ditampilkan berwarna putih.

Pengujian akurasi dilakukan dengan melihat perkembangan nilai RMSE. RMSE merupakan metode untuk mengevaluasi hasil teknik peramalan yang digunakan dengan mengukur tingkat akurasi dari hasil prakiraan suatu model [6].

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}} \quad (9)$$

dimana

y_i = data awal

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 [7]. Sedangkan apabila nilai RMSE semakin besar, maka keakuratan suatu nilai yang dihasilkan semakin sedikit atau tidak akurat.

3. Hasil dan Pembahasan

Terdapat 5 kecamatan di kabupaten jembrana. Tabel II menunjukkan data hasil panen padi yang berisi atribut Kecamatan, Luas Tanam, Luas Panen, Hasil Produksi Padi dan Rata-Rata Hasil Produksi Padi. Pada tahap Pre-Processing dilakukan proses data cleaning. Proses data cleaning digunakan untuk menghilangkan atribut data yang tidak dibutuhkan dalam proses pre-processing, yaitu dengan menghilangkan atau membuang informasi atribut Kecamatan yang tidak diperlukan dalam pengolahan data

Tabel I
Dataset Padi

Kecamatan	Luas Tanam	Luas Panen	Hasil Produksi Padi
Pekutatan	7	6	3
Mendoyo	40	40	19
Jembrana	11	12	7
Negara	40	32	17
Melaya	20	18	9

Tabel II
Data Cleaning

Luas Tanam	Luas Panen	Hasil Produksi Padi
7	6	3
40	40	19
11	12	7
40	32	17
20	18	9

Hasil perhitungan algoritma linear regression me nyatakan hubungan antara Luas Tanam, Luas Panen dan Hasil Produksi Padi.

$$Y = a + b_1(\text{Luas Tanam}) + b_2(\text{Luas Panen}) \quad (10)$$

Untuk mendapatkan persamaan linear regression berganda dilakukan model *multiple linear regression*.

3.1. Multiple Linear Regression

Nilai a, b1, dan b2 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, b1, dan b2. Persamaannya adalah

$$A = \begin{bmatrix} N & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1 X_1 & \sum X_1 X_2 \\ \sum X_2 & \sum X_2 X_1 & \sum X_2 X_2 \end{bmatrix} \quad (11)$$

A merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A), dengan N adalah banyaknya data, $\sum X_1$ adalah jumlah dari variabel Luas Tanam dan $\sum X_2$ adalah jumlah dari variabel Luas Panen.

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

A_0 merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A_0), dengan $\sum Y$ adalah jumlah dari variabel Hasil Produksi, $\sum X_1$ adalah jumlah dari variabel Luas Tanam, $\sum X_2$ adalah jumlah dari variabel Luas Panen

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

A_1 merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A_1), dengan N adalah banyaknya data, $\sum Y$ adalah jumlah dari variabel Hasil Produksi, $\sum X_2$ adalah jumlah dari variabel Luas Panen.

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

A_2 merupakan persamaan matriks yang digunakan untuk mencari nilai Det(A_2), dengan N adalah banyaknya data, $\sum X_1$ adalah jumlah dari variabel Luas Tanam, $\sum Y$ adalah jumlah dari variabel Hasil Produksi. Langkah selanjutnya adalah menentukan determinasi matriks A, A_0 , A_1 , A_2

$$\begin{aligned} \text{Det}(A) &= \{N \cdot \sum X_1 X_1 \cdot \sum X_2 X_2\} + \{\sum X_1 \cdot \sum X_1 X_2 \cdot \sum X_2\} + \\ &\quad \{\sum X_2 \cdot \sum X_1 \cdot \sum X_2 X_1\} - \{\sum X_2 \cdot \sum X_1 X_1 \cdot \sum X_2\} - \\ &\quad \{N \cdot \sum X_1 X_2 \cdot \sum X_2 X_1\} - \{\sum X_1 \cdot \sum X_1 \cdot \sum X_2 X_2\} \\ \text{Det}(A_0) &= \{\sum Y \cdot \sum X_1 X_1 \cdot \sum X_2 X_2\} + \{\sum X_1 \cdot \sum X_1 X_2 \cdot \sum X_2 Y\} + \\ &\quad \{\sum X_2 \cdot \sum X_1 Y \cdot \sum X_2 X_1\} - \{\sum X_2 \cdot \sum X_1 X_1 \cdot \sum X_2 Y\} - \\ &\quad \{\sum Y \cdot \sum X_1 X_2 \cdot \sum X_2 X_1\} - \{\sum X_1 \cdot \sum X_1 Y \cdot \sum X_2 X_2\} \\ \text{Det}(A_1) &= \{N \cdot \sum X_1 Y \cdot \sum X_2 X_2\} + \{\sum Y \cdot \sum X_1 X_2 \cdot \sum X_2\} + \\ &\quad \{\sum X_2 \cdot \sum X_1 \cdot \sum X_2 Y\} - \{\sum X_2 \cdot \sum X_1 Y \cdot \sum X_2\} - \\ &\quad \{N \cdot \sum X_1 X_2 \cdot \sum X_2 Y\} - \{\sum Y \cdot \sum X_1 \cdot \sum X_2 X_2\} \\ \text{Det}(A_2) &= \{N \cdot \sum X_1 X_1 \cdot \sum X_2 Y\} + \{\sum X_1 \cdot \sum X_1 Y \cdot \sum X_2\} + \\ &\quad \{\sum Y \cdot \sum X_1 \cdot \sum X_2 X_1\} - \{\sum Y \cdot \sum X_1 X_1 \cdot \sum X_2\} - \\ &\quad \{N \cdot \sum X_1 Y \cdot \sum X_2 X_1\} - \{\sum X_1 \cdot \sum X_1 \cdot \sum X_2 Y\} \end{aligned}$$

Berdasarkan perhitungan determinasi matriks, didapatlah hasil nilai a, b1, dan b2, seperti ditunjukkan pada Tabel III.

Tabel III.
Hasil Perhitungan Konstanta dan Koefisien

a	b ₁	b ₂
-0,79822	0,14939	0,31532

Berdasarkan hasil yang diperoleh untuk koefisien a, b1, dan b2 tersebut, hasil model multiple linear regression menjadi sebagai berikut.

$$\sum Y = -0,79822 + 0,14939 \sum X_1 + 0,31532 \sum X_2$$

Tabel IV
Data aktual dan hasil prediksi

Data Aktual	Hasil Prediksi
3	2,1
19	18
7	5
17	16
9	8

$$\begin{aligned}
 RMSE &= \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}} \\
 &= \sqrt{\frac{0,9^2 + 1^2 + 2^2 + 1^2 + 1^2}{5}} \\
 &= 1,249
 \end{aligned}$$

3.2. Uji Korelasi

Perhitungan uji korelasi parsial digunakan untuk mengetahui besar keterkaitan masing-masing variabel independen terhadap variabel dependen.

Perhitungan r_{X_1Y} :

$$\begin{aligned}
 \sum X_1Y &= \sum X_1Y - \frac{(\sum X_1) \cdot (\sum Y)}{n} \\
 &= 1718 - \frac{(118) \cdot (55)}{5} \\
 &= 420
 \end{aligned}$$

Perhitungan r_{X_2Y} :

$$\begin{aligned}
 \sum X_2Y &= \sum X_2Y - \frac{(\sum X_2) \cdot (\sum Y)}{n} \\
 &= 1568 - \frac{(108) \cdot (55)}{5} \\
 &= 380
 \end{aligned}$$

Perhitungan $r_{X_1X_2}$:

$$\begin{aligned}
 \sum X_1X_2 &= \sum X_1X_2 - \frac{(\sum X_1) \cdot (\sum X_2)}{n} \\
 &= 3414 - \frac{(118) \cdot (108)}{5} \\
 &= 865,2
 \end{aligned}$$

Tabel V
Hasil Uji Korelasi

Keterkaitan Variabel	Nilai Korelasi
r_{X_1Y}	420
r_{X_2Y}	380
$r_{X_1X_2}$	865,2

3.3. Uji Koefisien Determinasi

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

$$\begin{aligned}
 R &= \frac{b_1 \sum X_1Y + b_2 \sum X_2Y}{\sum Y^2} \\
 &= \frac{0,149 \cdot 1718 + 0,315 \cdot 1568}{789} \\
 &= 0,950446
 \end{aligned}$$

Hasil koefisien determinasi sebesar 0,950446. Artinya tingkat kecocokan model multiple linear regression memiliki tingkat keandalan sebesar 95,04%. Sebanyak 95,04% variasi nilai Hasil Produksi

bergantung pada variabel independen yang diukur, yaitu Luas Tanam dan Luas Panen Sedangkan sisanya, yaitu sebesar 4,96%, dipengaruhi oleh variabel lain yang tidak diukur pada penelitian ini.

4. Kesimpulan

Berdasarkan analisis data dan pembahasan yang telah dipaparkan, dapat ditarik kesimpulan bahwa algoritma linear regression dapat diterapkan untuk prediksi hasil panen padi dengan beberapa variabel yang memengaruhi yaitu Luas Tanam, Luas Panen, dan Hasil Produksi Padi

Dari jumlah data sebanyak 5 Kecamatan, dihasilkan tingkat kecocokan model multiple linear regression sebesar 95,04%. Artinya sebanyak 95,04% variasi nilai hasil panen bergantung pada variabel independen yang diukur, yaitu Luas Tanam dan Luas Panen. Sedangkan sisanya sebesar 4,96%, dipengaruhi oleh variabel lain yang tidak diukur pada penelitian ini.

Referensi

- [1] 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. Vol. 24, p. hal. 27–35, 2017.
- [2] Z.A. Matondang, "Sistem Pendukung Keputusan Forecasting Harga Emas Lelang pada Pegadaian dengan Metode Single Moving Average," *J. Tek. Inform. Unika St. Thomas*, Vols. Vol. 3, No. 1, p. hal. 72–77, 2018.
- [3] F. Anis dan Suprayogi, "Estimasi Luas Panen Padi di Kabupaten Rembang Menggunakan Algoritma Linear Regression," *Skripsi, Universitas Dian Nuswantoro, Semarang, Indonesia*, 2015.
- [4] 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.
- [5] 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.
- [6] 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*, Vols. Vol. 7, No. 3, p. hal. 1247–1250, 2014.
- [7] P. Choirunisa dan Kariyam, "Perbandingan Metode Triple Exponential Smoothing dan Metode Seasonal ARIMA untuk Peramalan Inflasi di Kota Tamjung Pandan," *Prosiding Sendika*, Vols. Vol. 5, No. 2, p. hal. 76–83, 2019.