

PERAMALAN HASIL PANEN PADI KABUPATEN CIREBON MENGGUNAKAN ALGORITMA REGRESI LINEAR BERGANDA

Rita Andia, Kaslani, Sandy Eka Permana, Tineka Handayani

Program Studi Teknik Informatika S1, STMIK IKMI Cirebon

Jalan Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon, Jawa Barat 45131

ritaandia5200@gmail.com

ABSTRAK

Pertanian telah lama menjadi sektor ekonomi yang signifikan di Indonesia. Pertanian berkontribusi dalam aspek perekonomian, kesejahteraan masyarakat dan pelestarian alam. Kabupaten Cirebon, sebagai bagian dari negara agraris, memiliki mayoritas penduduk yang menggeluti sektor pertanian, terutama dalam kegiatan bertani padi. Hasil produksi panen padi Kabupaten Cirebon mengalami fluktuasi setiap tahunnya, perubahan tersebut menjadi fokus pada penelitian ini. Pada tahun 2020 produksi padi Kabupaten Cirebon mencapai 709.105 Ton/GKP sedangkan pada tahun 2021 produksi padi mengalami penurunan menjadi 703.044 Ton/GKP. Kemudian pada tahun 2022 produksi padi mengalami penurunan kembali menjadi 701.059 Ton/GKP. Penurunan produksi padi selama dua tahun secara berurutan menimbulkan kekhawatiran akan produksi padi pada tahun yang akan datang. Tujuan dalam penelitian ini yaitu melakukan peramalan sebagai upaya untuk memberikan gambaran produksi padi pada tahun 2023 menggunakan algoritma regresi linear berganda dengan menerapkan tahapan *Knowledge Discovery in Database* (KDD). Regresi linear berganda merupakan teknik statistik yang digunakan untuk meramalkan nilai dari suatu indikator dependen (Y) berdasarkan lebih dari satu indikator independen (X). Hasil peramalan produksi padi tahun 2023 sebesar 685.403 Ton/GKP, dengan hasil evaluasi peramalan menggunakan matriks evaluasi yaitu MSE sebesar 1691845,33, RMSE sebesar 1300,70, MAE sebesar 844,70 dan R2-Score sebesar 0,99. Berdasarkan hasil peramalan, produksi padi tahun 2023 mengalami penurunan.

Kata kunci : Data Mining, Regresi Linear Ganda, Forecasting, Rice Harvest, Teknologi Informasi dan Komunikasi

1. PENDAHULUAN

Kabupaten Cirebon merupakan salah satu lumbung padi di Jawa Barat. Berdasarkan data yang dirilis oleh Badan Pusat Statistik (BPS) Kabupaten Cirebon, penduduk di Kabupaten Cirebon pada tahun 2022 tercatat sebanyak 2.315.417 jiwa. Sebagian besar penduduk Kabupaten Cirebon berprofesi sebagai petani. Padi menjadi tanaman pangan utama karena mayoritas penduduknya cenderung memilih beras sebagai konsumsi utama dalam kehidupan sehari-hari, yang kemudian diolah menjadi nasi. Kabupaten Cirebon, dengan perannya sebagai lumbung padi, memiliki kontribusi signifikan dalam menjaga ketahanan pangan di tingkat Kabupaten. Rata-rata konsumsi beras (beras lokal, kualitas unggul, impor) per kapita pada tahun 2022 tercatat sebanyak 1,418 Kg (Statistik, 2022)

Pada setiap tahunnya produksi padi Kabupaten Cirebon memperoleh angka produksi gabah kering panen yang fluktuasi. Berdasarkan data yang dirilis oleh Badan Pusat Statistik (BPS) Kabupaten Cirebon, pada tahun 2018, Kabupaten Cirebon dapat memproduksi padi sebanyak 672.882 Ton/GKP. Kemudian pada tahun 2019 produksi padi mengalami penurunan menjadi 672.630 Ton/GKP. Sedangkan pada tahun 2020 produksi padi mengalami peningkatan menjadi 709.105 Ton/GKP. Pada tahun 2021 produksi padi mengalami penurunan menjadi 703.044 Ton/GKP. Namun, pada tahun 2022 produksi padi mengalami penurunan kembali menjadi 701.059 Ton/GKP. Data ini menunjukkan bahwa telah terjadi

penurunan produksi padi selama 2 (dua) tahun terakhir secara beruntun yang menimbulkan kekhawatiran terhadap produksi padi pada tahun yang akan datang. Pada tahun 2018 Kabupaten Cirebon memiliki lahan tanam seluas 92.196 Ha. Kemudian pada tahun 2019 adanya pengurangan lahan tanam mencapai 87.169 Ha. Pada tahun 2020 adanya penambahan lahan tanam menjadi 90.778 Ha. Kemudian pada tahun 2021 adanya penambahan lahan tanam kembali mencapai 92.539 Ha. Namun, pada tahun 2022 lahan tanam berkurang kembali menjadi 91.382 Ha (Publikasi, 2018-2022).

Kenaikan dan penurunan produksi padi dapat disebabkan oleh produksi padi tahun sebelumnya, adanya pengurangan dan penambahan lahan pertanian, serta faktor lain yaitu fenomena alam yang sedang terjadi pada saat ini di tahun 2023 yaitu *El Nino* (Pemanasan suhu laut di wilayah Pasifik Tengah dan Timur) yang membawa dampak kemarau panjang sehingga dapat menyebabkan gagal panen. Sebagai upaya untuk mengatasi kekhawatiran akan produksi padi tahun 2023 yang akan datang, maka penelitian ini bertujuan untuk melakukan peramalan menggunakan algoritma regresi linear berganda dengan menerapkan tahapan dari proses *Knowledge Discovery in Database* sebagai metode peramalan untuk memberikan gambaran mengenai produksi padi pada tahun 2023 dengan menggunakan data pada tahun-tahun sebelumnya. Adapun data yang di gunakan meliputi kecamatan, tahun, luas tanam, luas panen dan produksi GKP. Peramalan merupakan *estimasi* mengenai

kejadian yang belum terjadi atau segala sesuatu cenderung tidak pasti [1]. Algoritma regresi linear berganda merupakan teknik statistika yang digunakan untuk meramalkan nilai indikator Y berdasarkan nilai indikator X1, X2, dan seterusnya. Indikator X1, X2 dan seterusnya disebut indikator independen (indikator bebas) sedangkan indikator dependen (indikator terikat) di lambangkan sebagai Y [2]. Faktor-faktor yang dapat mempengaruhi di tentukan berdasarkan nilai korelasi antar indikator. Indikator independen yang dapat mempengaruhi adalah luas tanam dan luas panen.

2. TINJAUAN PUSTAKA

Berdasarkan penelitian sebelumnya yaitu *paper satu*, penelitian yang dilakukan oleh Heru Wahyu Herwanto dengan judul “Penerapan Algoritma Linear Regression untuk Prediksi Hasil Panen Tanaman Padi”, menjelaskan bahwa permasalahan pada penelitian ini yaitu adanya ketidakseimbangan antara produksi dan konsumsi. Berdasarkan data dari Badan Pusat Statistik (BPS) pada tahun 2018, Indonesia memproduksi beras sebesar 32,4 juta Ton, dengan target produksi yang lebih tinggi hingga akhir 48 juta Ton menurut Kementerian Pertanian. Namun, perkiraan konsumsi beras hingga akhir tahun mencapai 29,6 juta Ton, menyisakan selisih sekitar 2,85 Ton. Regresi merupakan suatu metode statistik yang digunakan untuk mengukur kekuatan hubungan antara indikator dependen dan indikator independen. Secara umum regresi terbagi menjadi 2 (dua) yaitu *simple linear regression* dan *multiple linear regression*. *Multiple linear regression* di mana indikator dependen adalah atribut hasil panen (Y) dan indikator independen terdiri dari dua atau lebih yang terdiri dari atribut luas lahan, jumlah bibit, pupuk urea, dan pupuk NPK Phonska. Hasil penelitian dari 300 data yang di analisis, model regresi linear berganda menghasilkan tingkat kesesuaian sebesar 94,51%. Ini mengidentifikasi bahwa sekitar 94,51% dalam penelitian ini dapat dijelaskan oleh indikator-indikator independen yang diukur, termasuk luas lahan, jumlah bibit, pupuk urea dan pupuk NPK Phonska. Sedangkan sisanya, yaitu 5,49% dipengaruhi oleh indikator lain yang tidak di ukur pada penelitian ini. Hasil tersebut menunjukkan nilai rata-rata akurasi RMSE sebesar 0,432 [3].

Paper dua berdasarkan penelitian sebelumnya yang dilakukan oleh Yaya Asohi dengan judul “Implementasi Algoritma Regresi Linier Berganda Untuk Prediksi Penjualan”. Permasalahan pada penelitian ini yaitu adanya ketidakpastian dalam jumlah penjualan barang yang menyebabkan kesulitan dalam merencanakan persediaan, manajemen keuangan, dan strategi pemasaran. Penelitian ini menggunakan algoritma regresi linier berganda dengan menerapkan tahapan *Knowledge Discovery in Database*. KDD adalah pemaparan tentang proses pengungkapan informasi yang tersembunyi dalam suatu himpunan data yang besar. Selain itu, konsep

KDD juga dipergunakan untuk menjelaskan langkah-langkah dalam mencari informasi yang belum diketahui dalam basis data. Adapun tahapan KDD meliputi *data selection* (menyeleksi data), *preprocessing* (*cleaning data*), *transformation* (mengubah data sesuai kebutuhan), *Data Mining* (pencarian pola atau pengetahuan), dan *evaluation* (mengartikan pola yang diperoleh dari tahapan *Data Mining*). Algoritma regresi linier berganda adalah metode regresi linier yang melibatkan lebih dari satu indikator independen (X1, X2, hingga Xn). Algoritma ini memungkinkan adanya hubungan dan pengaruh antara indikator-indikator ini terhadap indikator dependen (Y). Hasil peramalan penjualan barang untuk tahun 2020 mencapai 169.715 unit, dan ditemukan bahwa barang dengan jenis *FORTUNE PP 1L/24* memiliki peramalan penjualan tertinggi sebesar 419,494 unit, sementara barang berjenis *TEH BOTOL KOTAK SOSRO 200 ML* memiliki peramalan penjualan terendah dengan jumlah 23,872 unit dengan nilai *Root Mean Square Error* (RMSE) sebesar 40,476. Proses ini memberikan wawasan mendalam tentang performa penjualan berbagai jenis barang serta memberikan gambaran tentang sejauh mana ketepatan peramalan tersebut [2].

Paper tiga berdasarkan penelitian sebelumnya yang dilakukan oleh Ghebyla Najla Ayuni dengan judul “Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ” menjelaskan bahwa permasalahan pada penelitian ini yaitu fluktuasi harga pasar dan minat pembeli yang berubah-ubah sehingga diperlukan pengetahuan terkait nilai penjualan di masa depan. Indikator independen (x) adalah atribut tanggal penjualan dan indikator dependen (y) adalah tipe properti. Metode yang digunakan adalah regresi linear. Regresi adalah pendekatan statistik yang memproyeksikan hasil peramalan dengan mengembangkan hubungan matematis antara indikator dependen (Y) dan indikator Independen (X). Pengujian akurasi peramalan pada penelitian ini menggunakan matriks evaluasi yaitu *Mean Squared Error* (MSE) yang merupakan nilai rata-rata dari selisih kuadrat antara nilai yang di ramalkan dan nilai yang di amati. *Root Mean Squared Error* (RMSE) sebagai akar kuadrat dari MSE, mengukur besaran sebenarnya dari kesalahan peramalan. Sedangkan *Mean Absolute Percentage Error* (MAPE) adalah nilai rata-rata dari persentase absolut perbedaan antara nilai yang diramalkan dan nilai aktual. Hasil dari penelitian ini yaitu properti rumah 59 berdasarkan peramalan terjual 4 unit, properti rumah sudut berdasarkan peramalan terjual 2 unit, enam tipe lainnya (rumah 28, 39, 62, 73, 84, dan 115) akan terjual masing-masing 1 unit, lima tipe properti lainnya (*kavling*, ruko, dan rumah 58, 67 serta 72) berdasarkan peramalan tidak akan mengalami penjualan [4].

Paper empat, berdasarkan penelitian terdahulu yang dilakukan oleh Diyanti dengan judul “Prediksi Hasil Panen Padi Tahun 2023 Menggunakan Regresi

Linier Di Kabupaten Indramayu”, menjelaskan bahwa permasalahan pada penelitian ini adalah produktivitas padi yang berubah-ubah setiap tahunnya dan belum diketahui faktor apa saja yang mempengaruhi produksi padi pada penelitian ini. Faktor-faktor penting dalam meningkatkan produksi padi tersebut meliputi atribut luas tanam, luas panen, curah hujan, dan puso. Penelitian ini menggunakan regresi linier sebagai algoritma untuk meramalkan hasil panen padi. Sebagai evaluasi hasil peramalan, penelitian ini menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE) dan R2-Score. Hasil peramalan panen padi di Kabupaten Indramayu sebanyak 1510403 Ton/GKP dengan nilai MAE sebesar 5449.45, MSE sebesar 72325540.80, RMSE sebesar 8504.44, dan R2 Score sebesar 0.93 dengan peramalan produksi padi tahun 2023 mengalami penurunan [5].

Paper lima, berdasarkan penelitian terdahulu yang dilakukan oleh Andik Adi Suryanto dengan judul “Penerapan Metode *Mean Absolute Error* (MEA) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi” menjelaskan bahwa penelitian ini dilatarbelakangi oleh kebutuhan untuk mengevaluasi tingkat keakuratan model peramalan produksi padi menggunakan regresi linear. Atribut yang digunakan yaitu jumlah penduduk dan jumlah produksi padi per tahun. Metode pada penelitian ini menggunakan *mean absolute error* (MAE) yang merupakan salah satu metode yang dipakai untuk menilai seberapa akurat suatu model peramalan. Hasil dari MAE mencerminkan rata-rata kesalahan mutlak atau absolut antara nilai sebenarnya dan nilai yang diramalkan pada

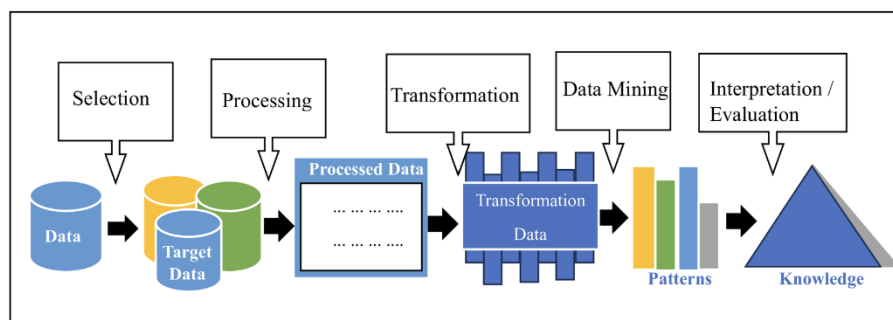
penelitian ini. Rumus MAE adalah $MAE = \frac{1}{n} \sum_{i=1}^n |F_i - Y_i|$. Di mana F_i adalah nilai peramalan, Y_i adalah nilai sebenarnya, dan n adalah jumlah data. Peramalan produksi padi tahun 2017, dengan hasil estimasi sekitar 5.347.763 kwintal menggunakan algoritma regresi linear mendapatkan nilai pengujian akurasi menggunakan Mean Absolute Error (MAE) sebesar 1,48950 [6].

Paper enam, berdasarkan penelitian terdahulu yang dilakukan oleh Hendra Di Kusuma dengan judul “Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda” menjelaskan bahwa permasalahan pada penelitian ini belum adanya sistem yang dapat melakukan peramalan jumlah mahasiswa baru untuk masa yang akan datang di STMIK Bina Nusantara Jaya. Penelitian ini menggunakan *algoritma regresi linear* berganda dengan 3 indikator independen. Atribut yang digunakan yaitu mahasiswa baru sebagai indikator dependen (Y) dan indikator independen (X) terdiri dari biaya, dan pendaftar. Algoritma regresi linear ganda digunakan untuk menilai apakah terdapat pengaruh yang signifikan antara dua atau lebih indikator independen. Rumus regresi linear berganda 2 (dua) indikator independen adalah $Y = \alpha + b_1.X_1 + b_2.X_2$. Jika data X_1 (biaya) adalah 499, dan data X_2 (pendaftar) adalah 56 tahun 2022, maka Y untuk tahun 2023 adalah $Y = 34,396 + (-0,070666667).499 + 0,7.56$ maka $Y = 38,608$. Hasil peramalan mahasiswa tahun 2023 pada penelitian ini didapatkan sebanyak 38 mahasiswa baru [7].

3. METODE PENELITIAN

3.1. Metode Penelitian

Penelitian ini mengimplementasikan metode *Knowledge Discovery in Database* (KDD) dalam melakukan analisis *Data Mining* dengan maksud untuk menemukan informasi yang tersembunyi dalam data yang besar [2][8]. Proses ini melibatkan serangkaian langkah-langkah tertentu yang dinyatakan pada gambar 1 berikut ini.



Gambar 1. Metode Penelitian

Untuk memberikan pemahaman yang lebih mendalam dalam setiap langkah dari proses KDD pada gambar 1, maka tahapan tersebut dijabarkan secara rinci sebagai berikut:

- Data selection* merupakan proses yang melibatkan pengambilan atau pemilihan data dari jumlah data yang ada [2] [8]. Pada tahap ini, melakukan aktivitas menyeleksi data sesuai

kebutuhan, adapun data yang esensial dalam konteks penelitian ini meliputi atribut kecamatan, luas tanam, luas panen, tahun, dan produksi gabah kering panen (GKP). Data yang digunakan merupakan data hasil panen padi Kabupaten Cirebon dalam kurun waktu 5 tahun (2018-2022).

- Preprocessing* data merupakan proses untuk memeriksa konsistensi data, menghilangkan

duplikasi, dan memperbaiki kesalahan data [8]. Pada tahap ini melakukan aktivitas memeriksa data untuk memastikan tidak ada *missing value* pada data serta mendeskripsikan data untuk mengetahui informasi pada data.

- c. *Transformation* merupakan proses untuk mengubah data ke dalam format tertentu [5]. Proses ini bertujuan agar data dapat di proses melalui proses *Data Mining* [9]. Pada tahap ini, melakukan aktivitas mengubah format data yang semula data itu diperoleh melalui publikasi Badan Pusat Statistik (BPS) dalam file PDF kemudian dilakukan transformasi menjadi file excel.
- d. *Data Mining*, pada tahap ini melakukan aktivitas mencari informasi pada data dengan mengimplementasikan *algoritma regresi linear* sebagai teknik analisis data. Adapun proses pada tahap ini yaitu mencari korelasi antar indikator, membagi data menjadi data *training* dan *testing*, mengimplementasikan *algoritma regresi linear*, mencari nilai koefisien dan konstanta (*intercept*) pada indikator Y dan indikator X, menguji data training menggunakan data testing (melakukan peramalan).
- e. *Evaluation/Interpretation* merupakan proses untuk mengevaluasi atau menilai hasil peramalan beserta kinerja model peramalan [2] [5]. Evaluasi pada tahap ini menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE) dan *R2-Score*.

3.2. Teknik Analisis Data

Algoritma regresi linear cukup efektif untuk melakukan peramalan menggunakan *Data Mining*. *Regresi* merupakan suatu metode statistik yang digunakan untuk mengukur kekuatan hubungan antara *indikator dependen* dan *indikator independen*. Secara umum regresi terbagi menjadi 2 (dua) yaitu *simple linear regression* dan *multiple linear regression* [3]. Pada jurnal ini menggunakan algoritma regresi linear berganda karena lebih dari satu *indikator independen*. Adapun faktor tersebut adalah luas tanam dan luas panen.

Algoritma regresi linier berganda adalah metode regresi linier yang melibatkan lebih dari satu *indikator independen* (X1, X2, hingga Xn). Untuk menentukan *indikator independen* dapat menggunakan matriks determinan, dengan mencari nilai korelasi dan konstanta [2]. Adapun rumusnya dinyatakan pada pernyataan (1) sampai (8) berikut ini.

$$A = \begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_2 \cdot X_1 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{bmatrix} \quad (1)$$

$$H = \begin{bmatrix} \sum Y \\ \sum Y \cdot X_1 \\ \sum Y \cdot X_2 \end{bmatrix} \quad (2)$$

Untuk menentukan matriks A, A1, A2 dan A3 sebagai berikut.

- a. Menentukan matriks A

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_2 \cdot X_1 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{bmatrix} \quad (3)$$

- b. Menentukan matriks A1

$$\begin{bmatrix} \sum Y & \sum X_1 & \sum X_2 \\ \sum Y \cdot X_1 & \sum X_1^2 & \sum X_2 \cdot X_1 \\ \sum Y \cdot X_2 & \sum X_1 X_2 & \sum X_2^2 \end{bmatrix} \quad (4)$$

- c. Menentukan matriks A2

$$\begin{bmatrix} n & \sum Y & \sum X_2 \\ \sum X_1 & \sum Y \cdot X_1 & \sum X_2 \cdot X_1 \\ \sum X_2 & \sum Y \cdot X_2 & \sum X_2^2 \end{bmatrix} \quad (5)$$

- d. Menentukan matriks A3

$$\begin{bmatrix} n & \sum X_1 & \sum Y \\ \sum X_1 & \sum X_1^2 & \sum Y \cdot X_1 \\ \sum X_2 & \sum X_1 X_2 & \sum Y \cdot X_2 \end{bmatrix} \quad (6)$$

Adapun penjelasan dari rumus pada pernyataan (1) sampai (6) adalah :

n : jumlah data pada indikator dependen
 $\sum Y$: jumlah total data dari indikator dependen
 $\sum X_1$: jumlah total data dari indikator X1
 $\sum X_2$: jumlah total data dari indikator X2
 $\sum X_1^2$: jumlah total dari indikator x_1^2
 $\sum X_2^2$: jumlah total dari indikator x_2^2

Kemudian setelah mendapatkan determinan dari matriks A, A1, A2, dan A3 maka tahap selanjutnya menentukan koefisien dan konstanta, adapun rumusnya dijabarkan pada pernyataan (7) berikut ini.

$$a = \frac{Det(A1)}{Det(A)} \quad B1 = \frac{Det(A2)}{Det(A)} \quad B2 = \frac{Det(A3)}{Det(A)} \quad (7)$$

Adapun penjelasan dari rumus pada pernyataan (7) adalah :

a adalah konstanta
 b1 adalah koefisien indikator X1
 b2 adalah koefisien indikator X2

Tahap selanjutnya, menerapkan nilai-nilai yang didapatkan ke dalam rumus regresi linear berganda 2 indikator independen. Rumus regresi linear berganda tersebut di paparkan pada pernyataan (8) berikut ini [2].

$$Y = a + b_1.X_1 + b_2.X_2 \quad (8)$$

Adapun penjelasan dari rumus regresi linear pada pernyataan (5) yaitu:

Y merupakan nilai indikator dependen

a merupakan konstanta

b merupakan koefisien, dan

X1 merupakan nilai indikator independen ke-1

X2 merupakan nilai indikator independen ke-2

Untuk melakukan penilaian hasil dari model peramalan dapat menggunakan MAE, MSE, RMSE, MAPE dan R2-Score. MSE merupakan singkatan dari *Mean Squared Error*, yang mencerminkan nilai rata-rata perbedaan kuadrat antara nilai yang diperbaiki dan nilai yang diamati sedangkan RMSE merupakan singkatan dari *Root Mean Squared Error* yang bertujuan mengukur akar kuadrat rata-rata dari selisih kuadrat antara nilai yang diramalkan dan nilai aktual [4]. Adapun rumus MSE dan RMSE dinyatakan pada pernyataan (6) dan (7) berikut ini.

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

MAE merupakan singkatan dari *Mean Absolute Error*, yang mengukur rata-rata dari nilai absolut dari selisih antara nilai yang diramalkan dan nilai aktual [6]. Adapun rumus MAE dinyatakan pada pernyataan (8) berikut ini.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (8)$$

R2 Score merupakan tahap untuk menilai sejauh mana pengaruh antara dua indikator dapat dipahami melalui pengujian koefisien determinasi, yang memberikan tingkat akurasi dari penerapan regresi. Koefisien ini memiliki rentang nilai antara 0 hingga 1. Jika nilai R2 sama dengan 0, dapat disimpulkan bahwa tidak terdapat pengaruh atau hubungan antara kedua indikator. Sebaliknya jika nilai R2 sama dengan 1 atau mendekati 1, ini menandakan adanya hubungan yang semakin kuat antara indikator dependen dan independen [6]. Adapun rumus untuk mencari R Squared dinyatakan pada pernyataan (9) berikut:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (9)$$

Di mana :

n adalah jumlah pengamatan

y_i adalah nilai yang diamati untuk pengamatan ke-i

$\hat{y}_i$ adalah nilai yang diramalkan untuk pengamatan ke-i

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Data Selection

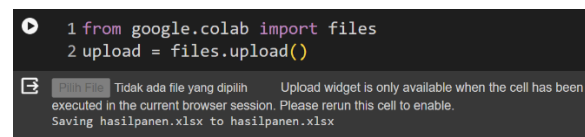
Pada tahap ini melakukan aktivitas menyeleksi data yang esensial untuk proses peramalan. Data tersebut dipresentasikan pada tabel 1 yang terlampir.

Tabel 1. Data yang sudah di seleksi

No	Kecamatan	Tahun	Luas Tanam	Luas Panen	Produksi GKP
1	Waled	2018	2123	2123	16507
2	Pasaleman	2018	93	93	670
3	Ciledug	2018	1035	1033	7726
4	Pabuaran	2018	723	723	5485
5	Losari	2018	1796	1779	13240
.....					
200	Kaliwedi	2022	4467	4467	31441

Data pada tabel 1, merupakan data hasil panen padi yang di rilis pada publikasi resmi Badan Pusat Statistik (BPS) Kabupaten Cirebon dalam rentang waktu 5 (lima) tahun. Data tersebut sebanyak 200 data, dengan 5 (lima) atribut. Adapun atribut tersebut meliputi kecamatan, tahun, luas tanam, luas panen, dan produksi gabah kering panen.

Data tersebut kemudian di unggah ke dalam proses *Data Mining*. Peramalan ini menggunakan bahasa pemrograman *Python* pada *google colabs*. Adapun proses tersebut di nyatakan pada gambar 2 berikut ini.



Gambar 2. Proses mengunggah data

Setelah data berhasil di unggah maka tahap selanjutnya menampilkan data. Adapun proses tersebut dinyatakan pada gambar 3 berikut ini.

```
1 data = pd.read_excel("hasilpanen.xlsx")
2 print(data)
```

Gambar 3. Proses membaca data

Hasil dari proses pada gambar 3 di tampilkan seperti pada tabel 1 *data selection*. Selanjutnya melakukan tahap kedua dari tahapan *Knowledge Discovery in Database* yaitu preprocessing.

4.1.1.1. Preprocessing

Pada tahap ini melakukan aktivitas memeriksa data serta menampilkan *type data*. Adapun tahap tersebut di nyatakan pada gambar 4 terlampir.

```
1 print(data.info())
2 print(data.isnull().sum())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  -
 0   kecamatan    200 non-null   object  
 1   tahun         200 non-null   int64   
 2   luas_tanam    200 non-null   int64   
 3   luas_panen    200 non-null   int64   
 4   produksi_gkp  200 non-null   int64   
dtypes: int64(4), object(1)
memory usage: 7.9+ KB
None
kecamatan    0
tahun        0
luas_tanam   0
luas_panen   0
produksi_gkp 0
dtype: int64
```

Gambar 4. Mengecek missing value

Proses pemeriksaan pada gambar 4 bertujuan untuk memastikan tidak terdapat *missing value* pada data dan menampilkan *type atribut* pada data. Hasil

dari proses tersebut menunjukkan bahwa tidak ada *missing value* atau nilai yang hilang pada data

```
1 total_produksi_per_kecamatan = data.groupby('kecamatan')['produksi_gkp'].sum().reset_index()
2 plt.figure(figsize=(10, 6))
3 bars = plt.bar(total_produksi_per_kecamatan['kecamatan'], total_produksi_per_kecamatan['produksi_gkp'])
4 plt.xlabel('kecamatan', fontsize=12)
5 plt.ylabel('Total Produksi GKP (Ton)', fontsize=12)
6 plt.title('Total Produksi GKP Per Kecamatan Selama 5 Tahun', fontsize=14)
7 plt.xticks(rotation=45, ha='right', fontsize=10)
8 plt.tight_layout()
9 for bar in bars:
10     plt.text(bar.get_x() + bar.get_width() / 2, bar.get_height() + 0.1, f'{bar.get_height():.2f}', ha='center', va='bottom', fontsize=8)
11 plt.show()
```

Gambar 6 Proses visualisasi total produksi per kecamatan

Pada gambar 6, menjelaskan bahwa total produksi per kecamatan di ambil dari *database* bernama "data" yang diambil dari atribut kecamatan dan produksi GKP kemudian di jumlahkan menggunakan fungsi *sum()*, setelah itu di urutan kembali berdasarkan abjad. Pada sumbu *x label*

tersebut. Adapun data tersebut terdiri dari 1 atribut dengan *type data* objek, yaitu kecamatan dan 4 atribut lainnya dengan *type data* integer yang meliputi tahun, luas tanam, luas panen, dan produksi gabah kering panen (GKP). Proses selanjutnya yaitu mendeskripsikan data. Adapun proses tersebut di paparkan pada gambar 5 berikut.

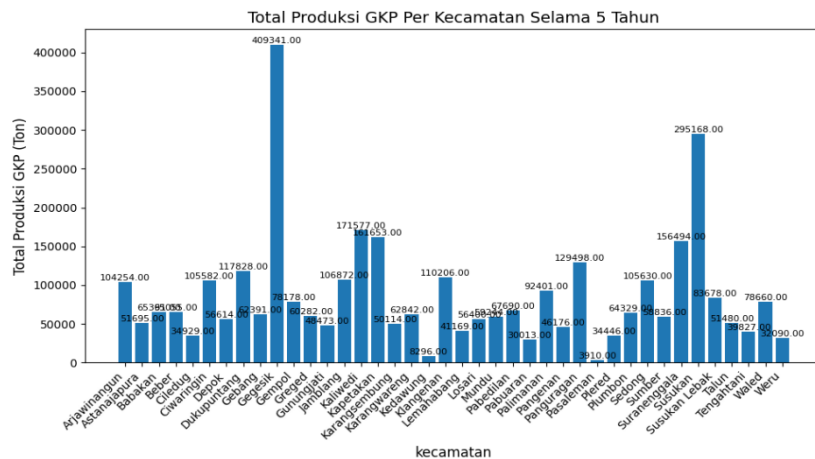
```
1 print(data.describe())
```

	tahun	luas_tanam	luas_panen	produksi_gkp
count	200.000000	200.000000	200.000000	200.000000
mean	2020.000000	2270.325000	2256.825000	17293.500000
std	1.417762	1901.38146	1905.141661	15168.48371
min	2018.000000	34.000000	34.000000	267.000000
25%	2019.000000	1379.000000	1317.250000	9746.000000
50%	2020.000000	1721.500000	1721.500000	12887.000000
75%	2021.000000	2635.000000	2526.750000	19545.500000
max	2022.000000	10497.000000	10497.000000	85876.000000

Gambar 5. Mendeskripsikan data

Hasil dari proses pada gambar 5, menjelaskan bahwa data yang digunakan sebanyak 200 data, dalam kurun waktu 5 tahun. Rata-rata luas tanam berada pada angka seluas 2270,325 Ha, rata-rata luas panen berada pada angka 2256,825 Ha, dan rata-rata produksi gabah kering panen (GKP) sebanyak 17293,560 Ton/GKP. Sedangkan luas tanam dan luas panen berdasarkan tingkat minimum berada pada angka 34 Ha dan produksi padi minimum sebanyak 267 Ton/GKP. Kemudian luas tanam dan luas panen maksimum pada data mencapai 10497 Ha dan produksi padi maksimum sebanyak 85876 Ton/GKP. Selanjutnya proses memvisualisasikan data berdasarkan total data aktual produksi GKP per kecamatan, hingga memudahkan dalam melihat informasi mengenai kecamatan mana saja yang memiliki kontribusi signifikan dalam produksi gabah kering panen di Kabupaten Cirebon menggunakan diagram batang. Adapun proses tersebut di nyatakan pada gambar 6 berikut.

(*horizontal*) digunakan untuk mengidentifikasi kecamatan, sedangkan sumbu *y label* digunakan untuk mengidentifikasi total produksi GKP per Ton. Adapun hasil dari perintah di atas dijabarkan pada gambar 7 berikut ini.



Gambar 7. Hasil visualisasi produksi per kecamatan

Pada gambar 7 menjelaskan bahwa kecamatan yang memiliki kontribusi terbesar terhadap produksi GKP di Kabupaten Cirebon adalah Kecamatan gegesik dengan total produksi selama 5 tahun sebanyak 409.341 Ton/GKP, kemudian di susul oleh Kecamatan Susukan sebesar 295.168 Ton/GKP sedangkan kecamatan yang memiliki tingkat produksi padi minimum yaitu Pasaleman (3910 Ton/GKP) dan Kedawung (8296 Ton/GKP).

4.1.2. Transformation

Pada tahap ini melakukan aktivitas mengubah data yang semula di dapatkan pada publikasi resmi Badan Pusat Statistik Kabupaten Cirebon berupa PDF di rubah (di sajikan kembali) menjadi excel.

4.1.3. Data Mining

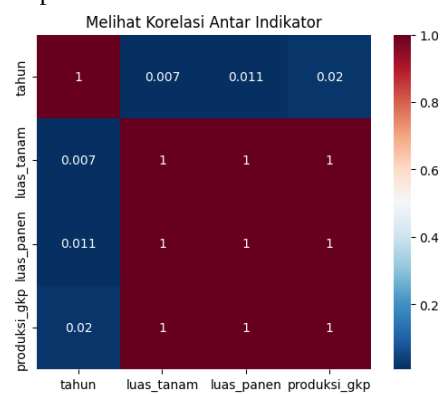
Pada tahap ini aktivitas yang dilakukan yaitu mengimplementasikan algoritma regresi linear sebagai teknik analisis *Data Mining*. *Data Mining* merupakan sebuah terminologi untuk menggali informasi dari sekumpulan data yang besar menggunakan metode tertentu [8][10]. Informasi ini memiliki nilai yang bermanfaat [11].

Tahap pertama yaitu mencari korelasi antar indikator dan memvisualisasikan menggunakan *heatmap*. Adapun proses tersebut di paparkan pada gambar 8 berikut ini.

```
1 sns.heatmap(data.corr(), cmap = "RdBu_r", square = True, annot=True, cbar=True)
2 plt.title("Melihat Korelasi Antar Indikator")
3 plt.show()
```

Gambar 8. Mencari korelasi antar indikator

Gambar 8 menjelaskan proses perhitungan matriks korelasi menggunakan *heatmap* pada *data frame* yang diberi nama "data" menggunakan fungsi "corr()", dengan visualisasi hasil korelasi antar indikator terlampir pada gambar 9 berikut ini.



Gambar 9. Hasil korelasi antar indikator

Pada gambar 9 menjelaskan bahwa terdapat hubungan korelasi sempurna dan bernilai positif. Dapat dikatakan sempurna karena korelasi bernilai 1. Untuk mempermudah melihat informasi, maka hasil tersebut di tampilkan ke dalam bentuk tabel 2 terlampir.

Tabel 2. Nilai koefisien antar indikator

No	Atribut pertama	Atribut Kedua	Korelasi
1	Tahun	Luas tanam	0.007
2	Tahun	Luas panen	0.011
3	Tahun	Produksi GKP	0.02
4	Luas tanam	Luas panen	1
5	Luas tanam	Produksi GKP	1
6	Luas panen	Produksi GKP	1

Berdasarkan hasil korelasi maka atribut yang terpilih untuk proses peramalan menggunakan *Data Mining* menggunakan *algoritma regresi linear berganda* adalah produksi GKP, luas tanam, dan luas panen. Produksi GKP sebagai indikator dependen (Y), dan luas tanam sebagai indikator independen (X1) serta luas panen sebagai indikator Independen (X2). Atribut ini di pilih karena memiliki nilai korelasi sebesar 1 (satu).

Tahap dua, data set penelitian dibagi menjadi 2 (dua) yaitu *data training* dan *data testing*. Adapun

proses tersebut dapat dilihat pada gambar 10 berikut ini.

```
1 x = data[['luas_tanam', 'luas_panen']]
2 y = data['produksi_gkp']
3 x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=21)
4 print(x_train.shape)
5 print(y_train.shape)
6 print(x_test.shape)
7 print(y_test.shape)
```

(160, 2)
(160,)
(40, 2)
(40,)

Gambar 10. Menyiapkan data training dan data testing

Proses pada gambar 10 menjelaskan bahwa *data training* yang digunakan sebanyak 80% sedangkan *data testing* sebanyak 20% dengan pembagian data secara acak. Sehingga *data training* yang diperoleh sebanyak 160 data sedangkan *data testing* sebanyak 40 data dari total 200 data. Data ini terdiri dari indikator dependen yaitu produksi GKP dan indikator independen yaitu atribut luas tanam dan luas panen. *Data training* merupakan kumpulan informasi yang digunakan untuk menghitung probabilitas berdasarkan pembelajaran yang dilakukan sementara *data testing* merupakan data yang digunakan sebagai elemen evaluasi yang sebelumnya telah diperoleh dari *data training* [8].

Tahap tiga, peneliti mengimplementasikan *algoritma regresi linear berganda* dan membuat model regresi dengan nama “model” untuk memanggil kelas *LinearRegression()*. Adapun proses tersebut dipaparkan pada gambar 11 berikut.

```
1 model = LinearRegression()
2 model.fit(x_train, y_train)
```

LinearRegression
LinearRegression()

Gambar 11. Implementasi algoritma regresi linear

Pada gambar 11 menjelaskan bahwa model di uji menggunakan data yang telah disiapkan pada proses sebelumnya yaitu pada gambar 10.

Tahap empat, melakukan aktivitas mencari koefisien dan konstanta (*intercept*). Koefisien merupakan cara untuk mengukur seberapa besar perubahan rata-rata dalam indikator dependen (Y) yang terjadi ketika indikator X sama dengan nol [6]. Proses untuk mencari nilai *koefisien* dan *intercept* dinyatakan pada gambar 12 berikut ini.

```
1 print("Koefisien Luas Tanam:", model.coef_[0])
2 print("Koefisien Luas Panen:", model.coef_[1])
3 print("Intercept:", model.intercept_)
```

Koefisien Luas Tanam: 4.1053889503612595
Koefisien Luas Panen: 3.89952456954694
Intercept: -788.0981161010968

Gambar 12. Mencari nilai koefisien dan intercept

Pada gambar 12 menjelaskan bahwa luas tanam mempengaruhi produksi padi sebesar 4,105 sedangkan luas panen dapat mempengaruhi produksi padi sebesar 3,899.

Tahap lima, meramalkan produksi GKP menggunakan *algoritma regresi linear berganda* dengan cara menguji *data training* menggunakan *data testing*, kemudian menampilkan hasil peramalan. Adapun proses tersebut dinyatakan pada gambar 13 berikut ini.

```
1 y_pred = model.predict(x_test)
2 print(y_pred)
```

[2921.66463942 12259.91092135 13524.68725749 28509.88536676
12876.49512676 12339.96005655 20608.85365556 16391.88734829
13244.5152843 11146.81621332 17470.70979271 6480.36335998
2774.08840026 10716.19779829 12283.92566191 20886.53145886
10913.66819667 21682.10586304 18545.20908514 11153.39197484
6592.63801364 26036.36708911 15702.02373491 83239.47910238
22983.1774106 20043.36309965 28687.22865049 58121.4896289
7489.18832786 17871.3552988 14205.10490669 11519.35348857
1024.91186397 11131.21811504 15106.36630434 24547.45317441
13606.38330774 10250.67762785 12144.54830198 8369.52295067]

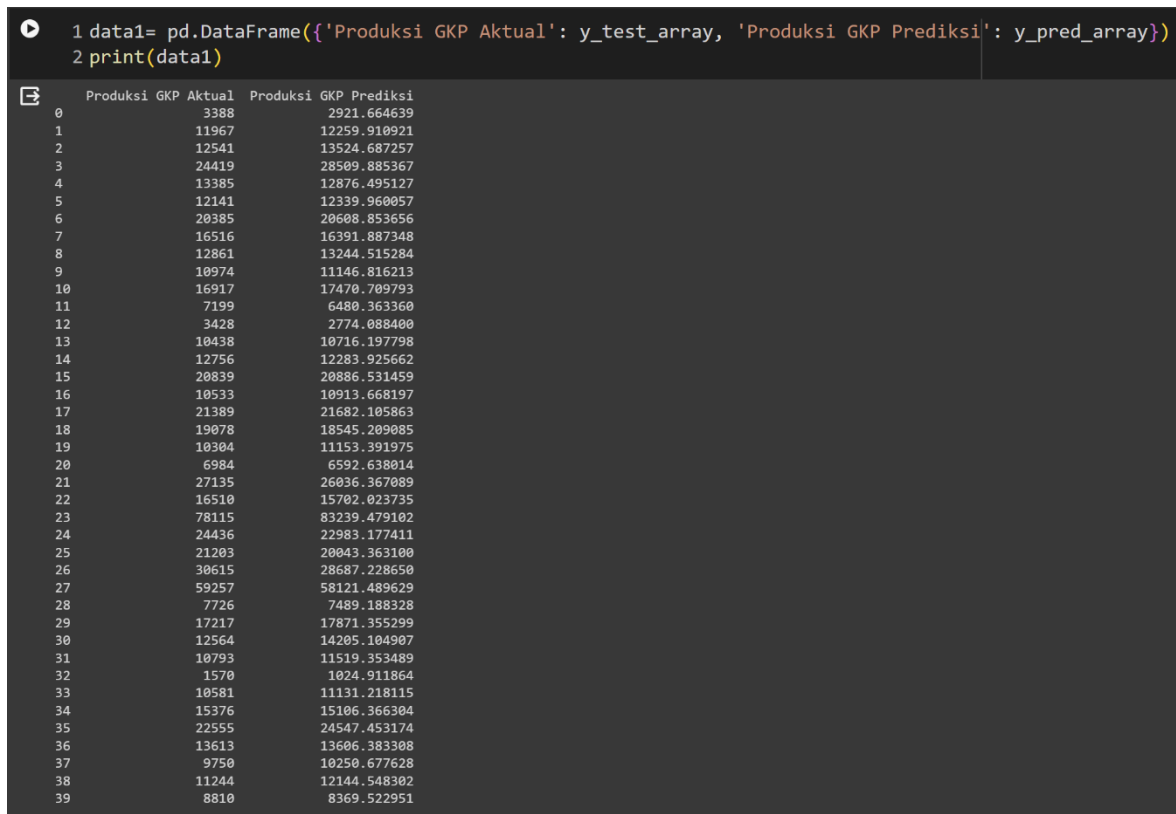
Gambar 13. Menguji data training dengan data testing

Sebelumnya pada tahap lima, hasil peramalan yang dilakukan telah didapatkan dalam bentuk *array*. Kemudian pada tahap enam dilakukan perubahan hasil peramalan yang semula berbentuk *array* menjadi *list*. Kemudian menampilkan untuk memudahkan pemanggilan data pada tahap berikutnya. Adapun proses tersebut dinyatakan pada gambar 14 berikut ini.

```
1 y_test_array = y_test.values.flatten() if hasattr(y_test, 'values') else y_test.flatten()
2 y_pred_array = y_pred.flatten()
```

Gambar 14. Merubah hasil peramalan array to list

Tahap tujuh, pada tahap ini menampilkan data peramalan yang semula berbentuk *array* menjadi *list* dan menampilkan data aktual dari produksi padi. Tujuan dari menampilkan data aktual dan peramalan dalam bentuk *list* agar dapat memudahkan dalam melihat perbandingan pada data tersebut. Adapun proses tersebut dinyatakan pada gambar 15 berikut ini.



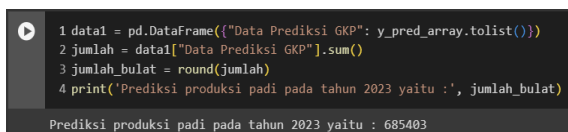
```
1 data1= pd.DataFrame({'Produksi GKP Aktual': y_test_array, 'Produksi GKP Prediksi': y_pred_array})
2 print(data1)
```

	Produksi GKP Aktual	Produksi GKP Prediksi
0	3388	2921.664639
1	11967	12259.910921
2	12541	13524.687257
3	24419	28509.885367
4	13385	12876.495127
5	12141	12339.960057
6	20385	20608.853656
7	16516	16391.887348
8	12861	13244.515284
9	10974	11146.816213
10	16917	17470.709793
11	7199	6480.363360
12	3428	2774.088400
13	10438	10716.197798
14	12756	12283.925662
15	20839	20886.531459
16	10533	10913.668197
17	21389	21682.105863
18	19078	18545.209085
19	10304	11153.391975
20	6984	6592.638014
21	27135	26036.367089
22	16510	15702.023735
23	78115	83239.479102
24	24436	22983.177411
25	21203	20043.363100
26	30615	28687.228650
27	59257	58121.489629
28	7726	7489.188328
29	17217	17871.355299
30	12564	14205.104907
31	10793	11519.353489
32	1570	1024.911864
33	10581	11131.218115
34	15376	15106.366304
35	22555	24547.453174
36	13613	13606.383308
37	9750	10250.677628
38	11244	12144.548302
39	8810	8369.522951

Gambar 15 Menguji data training menggunakan data testing

Hasil dari proses pada gambar 15, menjelaskan bahwa data hasil peramalan sebanyak 40 data, di mana *data testing* (aktual) di beri nama “Produksi GKP Aktual” yang di ambil dari “*y_test_array*” di sandingkan dengan data hasil peramalan yang diberi nama “Produksi GKP Prediksi” yang di ambil dari data hasil peramalan yaitu “*y_pred_array*”. Data ini di tampilkan dengan memanggil nama *dataframe* baru yang diberi nama “*data1*”.

Tahap delapan, meramalkan hasil panen padi pada tahun 2023, dengan menjumlahkan semua hasil peramalan yang di dapatkan. Adapun proses tersebut dinyatakan pada gambar 11 berikut ini.



```
1 data1 = pd.DataFrame({"Data Prediksi GKP": y_pred_array.tolist()})
2 jumlah = data1["Data Prediksi GKP"].sum()
3 jumlah_bulat = round(jumlah)
4 print('Prediksi produksi padi pada tahun 2023 yaitu :', jumlah_bulat)
```

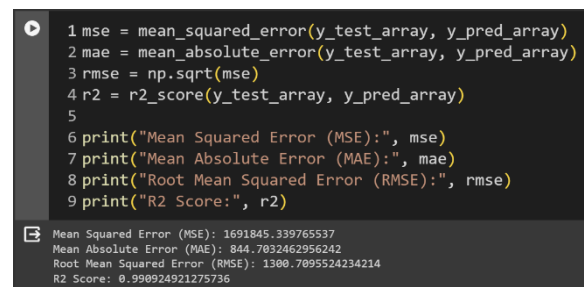
Prediksi produksi padi pada tahun 2023 yaitu : 685403

Gambar 16 Meramalkan hasil panen padi tahun 2023

Proses pada gambar 16 menjelaskan bahwa hasil peramalan produksi gabah kering panen (GKP) pada tahun 2023 menggunakan algoritma regresi linear berganda 2 indikator independen (X) yang terdiri dari atribut luas tanam dan luas panen memperoleh nilai peramalan produksi GKP sebesar 685.403 Ton/GKP.

4.1.4. Interpretation/Evaluation

Pada tahap ini, aktivitas yang dilakukan adalah mengukur hasil peramalan dan kinerja model dalam melakukan peramalan. Alur yang dihasilkan harus dalam bentuk yang mudah dimengerti [12]. Evaluasi pada peramalan ini menggunakan matriks evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE) dan *R-2 Score*. Adapun proses tersebut di nyatakan pada gambar 17 berikut ini.



```
1 mse = mean_squared_error(y_test_array, y_pred_array)
2 mae = mean_absolute_error(y_test_array, y_pred_array)
3 rmse = np.sqrt(mse)
4 r2 = r2_score(y_test_array, y_pred_array)
5
6 print("Mean Squared Error (MSE):", mse)
7 print("Mean Absolute Error (MAE):", mae)
8 print("Root Mean Squared Error (RMSE):", rmse)
9 print("R2 Score:", r2)
```

Mean Squared Error (MSE): 1691845.339765537
Mean Absolute Error (MAE): 844.7032462956242
Root Mean Squared Error (RMSE): 1300.7095524234214
R2 Score: 0.990924921275736

Gambar 17. Mengevaluasi hasil peramalan

4.2. Pembahasan

Peramalan ini menggunakan algoritma regresi linear berganda dengan indikator dependen adalah atribut produksi GKP yang dilambangkan dengan indikator Y dan 2 indikator independen yaitu luas tanam dan luas panen yang dilambangkan dengan indikator X. Peramalan merupakan estimasi mengenai kejadian yang belum terjadi atas segala sesuatu cenderung tidak pasti [1]. Pada gambar 16 dan 17

menjelaskan bahwa dari hasil peramalan produksi gabah kering panen yang didapatkan melalui proses peramalan *Data Mining* menggunakan *regresi linear berganda* mendapatkan nilai peramalan sebesar 685403 Ton/GKP dengan *Mean Squared Error* sebesar 1691845,3397, *Mean Absolute Error* sebesar 844,7032, *Root Mean Squared Error* 1300,7095 dan *R2-Score* 0,9909. Menurut beberapa penelitian terdahulu mengenai peramalan hasil panen padi menjelaskan bahwa nilai *R-Squared* berkisar antara 0 dan 1[5]. Jika koefisien determinan *R2* sama dengan 0, itu menunjukkan bahwa tidak ada hubungan atau pengaruh antara indikator independen dan indikator dependen. Sebaliknya, jika nilai *R2* mendekati 1, itu mengidentifikasi bahwa hubungan antara indikator independen dan indikator dependen semakin erat [3].

5. KESIMPULAN

Kesimpulan dari peramalan produksi padi Kabupaten Cirebon tahun 2023 menggunakan algoritma regresi linear berganda adalah sekitar 685403 Ton/GKP sementara tahun 2022 produksi padi mencapai 701.059 Ton/GKP. Meskipun terjadi penurunan hasil panen pada tahun 2023, perlu diingat bahwa peramalan hanya memberikan gambaran dan tidak memberikan jawaban pasti. Nilai korelasi antara luas tanam dengan produksi GKP, serta luas panen dengan produksi GKP, menunjukkan korelasi positif. Model peramalan menggunakan algoritma regresi linear berganda memiliki nilai *Squared Error* mendekati 1 (0,9909), menunjukkan kemampuan model mengikuti variasi data dengan nilai *MSE* sebesar 1691845,339, *MAE* sebesar 844,703 dan *RMSE* sebesar 1300,709. Produksi padi tertinggi yaitu Kecamatan Gegesik (409341 Ton/GKP), sementara produksi minimum yaitu Kecamatan Pasaleman (3910 Ton/GKP).

DAFTAR PUSTAKA

- [1] M. N. Fawaiq, A. Jazuli, and M. M. Hakim, "Prediksi Hasil Pertanian Padi Di Kabupaten Kudus Dengan Metode Brown'S Double Exponential Smoothing," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 4, no. 2, p. 78, 2019, doi: 10.29100/jupi.v4i2.1421.
- [2] Y. Asohi and A. Andri, "Impelementasi Algoritma Regresi Linier Berganda Untuk Prediksi Penjualan," *J. Nas. Ilmu Komput.*, vol. 1, no. 3, pp. 149–158, 2020, doi: 10.47747/jurnalnik.v1i3.161.
- [3] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, "Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi," *JNTETI*, vol. 8, no. 4, pp. 364–370, 2019, [Online]. Available: <https://journal.ugm.ac.id/v3/JNTETI/article/view/2563/694>
- [4] G. N. Ayuni and D. Fitriana, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2019, [Online]. Available: <https://journal.ithb.ac.id/telematika/article/view/321>
- [5] A. B. Diyanti, Martanto, "PREDIKSI HASIL PANEN PADI TAHUN 2023 MENGGUNAKAN METODE REGRESI LINIER DI KABUPATEN INDRAMAYU," *J. Inform. Terpadu HIVE-HADOOP*, vol. 6, no. 1, pp. 20–28, 2020, [Online]. Available: <https://journal.nurulfikri.sc.id/index.php/JIT>
- [6] A. A. Suryanto, "Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi," *Saintekbu*, vol. 11, no. 1, pp. 78–83, 2019, doi: 10.32764/saintekbu.v11i1.298.
- [7] Hendra Di Kesuma, D. Apriadi, H. Juliansa, and E. Etriyanti, "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda," *J. Ilm. Bin. STMIK Bina Nusantara. Jaya Lubuklinggau*, vol. 4, no. 2, pp. 62–66, 2022, doi: 10.52303/jb.v4i2.74.
- [8] M. Walid and F. Halimiyah, "Klasifikasi Kemandirian Siswa SMA/MA Double Track Menggunakan Metode Naive Bayes," *J. ICT Inf. Commun. Technol.*, vol. 22, pp. 190–197, 2022.
- [9] Syafi'i, O. Nurdiawan, and G. Dwilestari, "Penerapan Machine Learning Untuk Menentukan Kelayakan Kredit Menggunakan Metode Support Vektor Machine," *J. Sist. Inf. dan Manaj.*, vol. 10, no. 2, pp. 1–6, 2022.
- [10] R. R. Khaerullah, N. Suarna, and O. Nurdiawan, "Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means," *J. Inform. dan Teknol. Inf.*, vol. 1, no. 2, pp. 125–132, 2023, doi: 10.56854/jt.v1i2.135.
- [11] F. Firmansyah and O. Nurdiawan, "Penerapan Data Mining Menggunakan Algoritma Frequent Pattern - Growth Untuk Menentukan Pola Pembelian Produk Chemicals," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 547–551, 2023, doi: 10.36040/jati.v7i1.6371.
- [12] M. R. Sulistio, N. Suarna, and O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.