

PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA REGRESI LINEAR UNTUK PREDIKSI HASIL PRODUKSI PERTANIAN BERBASIS WEB

Mira Susanti ^{1)*}

¹ Manajemen Informatika, Akademi Manajemen Informatika dan Komputer, Bukittinggi

Corresponding Author: Mmira_0310021@yahoo.com

Article Info

Article history:

Received: Mei 13, 2026

Revised: Mei 29, 2026

Accepted: Jun 28, 2026

Published: Jul 30, 2026

Keywords:

Data Mining,
Regresi Linear,
Prediksi,
Hasil Produksi Pertanian,
Sistem Berbasis Web

ABSTRACT

Hasil produksi pertanian dipengaruhi oleh berbagai faktor, seperti luas lahan, jenis tanaman, biaya produksi, penggunaan pupuk, serta kondisi produksi pada periode sebelumnya. Ketidakpastian hasil produksi dapat menyulitkan petani dalam melakukan perencanaan dan pengambilan keputusan, terutama apabila perkiraan produksi masih dilakukan berdasarkan pengalaman tanpa didukung oleh pengolahan data historis. Penelitian ini bertujuan menerapkan teknik Data Mining menggunakan algoritma Regresi Linear untuk memprediksi hasil produksi pertanian berbasis web. Data yang digunakan berupa data historis kegiatan produksi pertanian yang mencakup variabel luas lahan, biaya produksi, dan hasil produksi pada periode sebelumnya. Tahapan penelitian meliputi pengumpulan data, preprocessing data, pemilihan variabel, pembentukan model Regresi Linear, pengujian model, evaluasi hasil prediksi, dan implementasi model ke dalam sistem berbasis web. Algoritma Regresi Linear digunakan untuk mengetahui hubungan antara variabel yang memengaruhi produksi dengan hasil produksi serta menghasilkan nilai prediksi berdasarkan data masukan. Sistem berbasis web dikembangkan untuk memudahkan pengguna dalam memasukkan data, melakukan proses prediksi, dan memperoleh informasi hasil estimasi produksi. Evaluasi kinerja model dilakukan menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2). Hasil penelitian diharapkan dapat menghasilkan model prediksi yang mampu memberikan estimasi hasil produksi secara objektif berdasarkan data historis serta membantu petani dalam melakukan perencanaan kegiatan produksi. Penerapan sistem berbasis web juga diharapkan dapat meningkatkan kemudahan akses terhadap informasi prediksi dan mendukung pemanfaatan data dalam pengambilan keputusan pada kegiatan pertanian.



This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. PENDAHULUAN

Pertanian merupakan salah satu sektor strategis yang berperan penting dalam memenuhi kebutuhan pangan dan mendukung perekonomian masyarakat. Keberhasilan kegiatan pertanian sangat dipengaruhi oleh kemampuan dalam merencanakan dan mengelola proses produksi secara tepat. Salah satu informasi yang penting dalam proses tersebut adalah hasil produksi pertanian, karena informasi mengenai estimasi hasil produksi dapat digunakan untuk menentukan kebutuhan sarana produksi, memperkirakan ketersediaan hasil panen, menyusun strategi pemasaran, serta mendukung pengambilan keputusan. Perubahan hasil produksi yang dipengaruhi oleh luas lahan, kondisi produksi, cuaca, jenis tanaman, dan berbagai faktor lainnya menyebabkan

proses estimasi hasil produksi menjadi suatu permasalahan yang perlu mendapatkan perhatian.

Perkembangan teknologi informasi dan pengolahan data memberikan peluang untuk meningkatkan kemampuan dalam melakukan prediksi hasil produksi pertanian. Pemanfaatan Data Mining dan *Machine Learning* memungkinkan data historis pertanian diolah untuk menemukan pola dan hubungan antara variabel yang memengaruhi hasil produksi. Penelitian mengenai prediksi hasil produksi tanaman menunjukkan bahwa pendekatan *machine learning* telah banyak digunakan untuk membantu melakukan estimasi hasil panen. Sebuah kajian sistematis terhadap penelitian prediksi hasil tanaman pada periode 2017–2024 menunjukkan bahwa Linear Regression, Random Forest, dan Gradient Boosting Trees termasuk algoritma yang paling banyak

digunakan, sedangkan RMSE, R^2 , dan MAE merupakan metrik evaluasi yang paling umum digunakan [1].

Salah satu metode yang dapat digunakan untuk melakukan prediksi adalah Regresi Linear. Algoritma ini digunakan untuk mengetahui hubungan antara variabel independen dengan variabel dependen dan menghasilkan nilai prediksi berdasarkan hubungan tersebut. Regresi Linear memiliki keunggulan berupa konsep yang relatif sederhana, mudah diimplementasikan, serta hasil modelnya lebih mudah diinterpretasikan. Dalam konteks pertanian, salah satu variabel yang dapat digunakan untuk memprediksi hasil produksi adalah luas lahan, karena perubahan luas area produksi dapat memiliki hubungan dengan jumlah hasil yang diperoleh. Penelitian berbasis data tanaman pangan di Indonesia menunjukkan bahwa luas panen memiliki hubungan linear yang kuat dengan produksi pada beberapa komoditas, dengan nilai R^2 yang tinggi untuk sebagian besar komoditas yang diuji [2].

Selain luas lahan, prediksi hasil produksi dapat memanfaatkan berbagai variabel lain, seperti curah hujan, temperatur, jenis tanah, karakteristik tanaman, penggunaan pupuk, dan data historis produksi. Kajian terbaru mengenai prediksi hasil tanaman menunjukkan bahwa temperatur, jenis tanah, vegetasi, data hasil produksi, curah hujan, dan karakteristik tanah merupakan beberapa fitur yang sering digunakan dalam penelitian prediksi hasil pertanian [1]. Hal tersebut menunjukkan bahwa ketersediaan data historis dan pemilihan variabel yang tepat merupakan faktor penting dalam menghasilkan model prediksi yang baik.

Penggunaan algoritma Regresi Linear dalam prediksi hasil produksi juga telah diterapkan pada berbagai penelitian. Penelitian mengenai prediksi hasil tanaman menggunakan beberapa algoritma *machine learning* menunjukkan bahwa Regresi Linear dapat digunakan sebagai salah satu pendekatan untuk melakukan prediksi hasil produksi berdasarkan karakteristik data pertanian [3]. Penelitian lain yang mengembangkan sistem prediksi produksi padi berbasis web menggunakan Regresi Linear menunjukkan bahwa model yang dioptimalkan dapat menghasilkan nilai R^2 sebesar 0,9754 dengan MAE sebesar 0,0957, MSE sebesar 0,0307, dan RMSE sebesar 0,1753 pada dataset yang digunakan [4]. Hasil tersebut menunjukkan bahwa Regresi Linear memiliki potensi untuk diterapkan dalam sistem prediksi produksi pertanian berbasis aplikasi.

Meskipun demikian, penerapan model prediksi dalam kegiatan pertanian tidak hanya berkaitan dengan pemilihan algoritma, tetapi juga bagaimana hasil pengolahan data dapat dimanfaatkan oleh pengguna. Model prediksi yang hanya dijalankan melalui kode program atau perangkat lunak analisis

dapat menyulitkan pengguna yang tidak memiliki kemampuan teknis dalam pemrograman. Oleh karena itu, diperlukan sebuah sistem berbasis web yang dapat mengintegrasikan proses input data, pengolahan data, proses prediksi, dan penyajian hasil dalam satu platform yang mudah digunakan.

Sistem berbasis web dapat memberikan antarmuka yang memungkinkan pengguna memasukkan variabel produksi secara langsung dan memperoleh hasil prediksi tanpa harus menjalankan proses pemrograman secara manual. Hasil prediksi juga dapat disajikan dalam bentuk tabel, grafik, maupun informasi ringkas sehingga lebih mudah dipahami. Pendekatan tersebut telah diterapkan dalam penelitian pengembangan dashboard sistem informasi dan prediksi produksi tanaman pangan berbasis web yang menggabungkan K-Means dan Linear Regression untuk melakukan analisis produksi serta menyajikan hasilnya melalui dashboard visualisasi [2].

Permasalahan yang kemudian muncul adalah bagaimana membangun sistem yang tidak hanya mampu menghasilkan nilai prediksi, tetapi juga dapat memberikan informasi prediksi melalui antarmuka web yang sederhana dan mudah digunakan. Sistem tersebut perlu melalui tahapan pengumpulan dan preprocessing data, pemilihan variabel, pembentukan model Regresi Linear, pengujian model, evaluasi performa, dan implementasi model ke dalam aplikasi. Evaluasi diperlukan untuk mengetahui tingkat kesalahan prediksi sehingga kualitas model dapat diukur secara objektif. Penggunaan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2) banyak digunakan dalam penelitian prediksi hasil pertanian untuk mengevaluasi kinerja model [1].

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan Penerapan Data Mining Menggunakan Algoritma Regresi Linear untuk Prediksi Hasil Produksi Pertanian Berbasis Web. Penelitian berfokus pada pemanfaatan data historis produksi untuk membangun model Regresi Linear yang dapat menghasilkan estimasi hasil produksi. Model yang dihasilkan kemudian diintegrasikan ke dalam sistem berbasis web sehingga pengguna dapat memasukkan data yang diperlukan dan memperoleh hasil prediksi secara langsung.

Kebaruan penelitian ini terletak pada integrasi antara proses Data Mining menggunakan algoritma Regresi Linear dengan sistem berbasis web untuk prediksi hasil produksi pertanian. Sistem tidak hanya menghasilkan model prediksi, tetapi juga menyediakan media yang dapat digunakan pengguna untuk melakukan input data dan memperoleh informasi estimasi produksi. Dengan pendekatan

tersebut, hasil pengolahan data dapat dimanfaatkan secara lebih praktis dalam kegiatan pertanian.

Berdasarkan uraian tersebut, tujuan penelitian ini adalah menerapkan algoritma Regresi Linear untuk membangun model prediksi hasil produksi pertanian serta mengimplementasikan model tersebut ke dalam sistem berbasis web. Penelitian ini diharapkan dapat menghasilkan model prediksi yang memiliki tingkat kesalahan yang dapat diterima berdasarkan hasil evaluasi serta menyediakan sistem yang dapat membantu pengguna memperoleh estimasi hasil produksi secara lebih cepat, mudah, dan berbasis data.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan Data Mining dengan menerapkan algoritma Regresi Linear untuk melakukan prediksi hasil produksi pertanian. Penelitian tidak hanya berfokus pada pembentukan model prediksi, tetapi juga mengimplementasikan model tersebut ke dalam sistem berbasis web agar proses prediksi dapat digunakan secara lebih mudah oleh pengguna. Tahapan penelitian meliputi pengumpulan data, preprocessing, pemilihan variabel, pembagian dataset, pembentukan model Regresi Linear, pengujian dan evaluasi model, implementasi sistem, serta pengujian aplikasi.

1. Pengumpulan Data

Data yang digunakan dalam penelitian berupa data historis kegiatan produksi pertanian. Data dapat diperoleh dari pencatatan kegiatan usaha tani, kelompok tani, instansi terkait, maupun dataset pertanian yang relevan.

Variabel yang digunakan disesuaikan dengan ketersediaan data. Contoh variabel penelitian dapat ditunjukkan pada Tabel 1.

Tabel 1. Variabel Data Penelitian

No.	Variabel	Keterangan	Peran
1	Luas Lahan	Luas lahan yang digunakan untuk produksi	Variabel independen
2	Biaya Produksi	Total biaya yang dikeluarkan	Variabel independen
3	Jumlah Pupuk	Jumlah pupuk yang digunakan	Variabel independen
4	Periode Produksi	Periode atau musim produksi	Variabel independen
5	Hasil Produksi	Jumlah hasil panen yang diperoleh	Variabel dependen

Variabel hasil produksi digunakan sebagai target atau variabel dependen yang akan diprediksi, sedangkan variabel lainnya digunakan sebagai variabel independen.

2. Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data yang digunakan dalam proses pemodelan memiliki kualitas yang baik. Tahapan preprocessing meliputi:

1. Data Cleaning, yaitu memeriksa data kosong, data ganda, dan data yang tidak valid.
2. Data Selection, yaitu memilih atribut yang relevan dengan tujuan prediksi.
3. Data Transformation, yaitu menyesuaikan format data agar dapat diproses oleh algoritma.
4. Outlier Checking, yaitu memeriksa data yang memiliki nilai terlalu jauh dibandingkan dengan data lainnya.
5. Data Integration, apabila data berasal dari beberapa sumber.

Tahap preprocessing diperlukan karena kualitas data dapat memengaruhi hasil pembentukan model prediksi.

3. Pembagian Dataset

Dataset yang telah melalui preprocessing kemudian dibagi menjadi data training dan testing. Data training digunakan untuk membangun model Regresi Linear, sedangkan data testing digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum digunakan selama proses pelatihan.

Pembagian dataset dapat menggunakan proporsi 80% data training dan 20% data testing. Pembagian dilakukan secara acak untuk mengurangi kemungkinan model hanya menyesuaikan diri dengan urutan data tertentu.

4. Penerapan Algoritma Regresi Linear

Regresi Linear digunakan untuk mengetahui hubungan antara variabel independen dengan hasil produksi sebagai variabel dependen. Apabila penelitian menggunakan lebih dari satu variabel independen, maka model yang digunakan adalah Regresi Linear Berganda.

Persamaan umum Regresi Linear Berganda adalah:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Keterangan:

- Y = hasil produksi yang diprediksi;
- a = konstanta;
- $b_1, b_2, \dots, b_n$ = koefisien regresi;
- $X_1, X_2, \dots, X_n$ = variabel independen.

Sebagai contoh, apabila variabel yang digunakan adalah luas lahan, biaya produksi, dan jumlah pupuk, maka model dapat dituliskan:

$$Y = a + b_1(\text{Luas Lahan}) + b_2(\text{Biaya Produksi}) + b_3(\text{Jumlah Pupuk})$$

Model tersebut menghasilkan nilai koefisien yang digunakan untuk melakukan prediksi terhadap data baru.

5. Pembentukan Model Prediksi

Setelah dataset dibagi, data training digunakan untuk membentuk model Regresi Linear. Model

mempelajari hubungan antara variabel independen dengan hasil produksi berdasarkan data historis.

Proses pembentukan model meliputi:

Data Training → Pembentukan Model Regresi Linear → Koefisien Regresi → Model Prediksi

Model yang telah terbentuk kemudian digunakan untuk memprediksi hasil produksi berdasarkan data testing.

6. Evaluasi Model

Evaluasi dilakukan untuk mengetahui tingkat akurasi dan kesalahan prediksi yang dihasilkan oleh model. Beberapa metrik yang digunakan adalah MAE, MSE, RMSE, dan R^2 .

a. Mean Absolute Error (MAE)

MAE digunakan untuk mengetahui rata-rata nilai kesalahan absolut antara hasil aktual dan hasil prediksi.

$$MAE = (1/n) \sum |Y_i - \hat{Y}_i|$$

Semakin kecil nilai MAE, semakin kecil rata-rata kesalahan prediksi yang dihasilkan model.

b. Mean Squared Error (MSE)

MSE menghitung rata-rata kuadrat kesalahan antara nilai aktual dan nilai prediksi.

$$MSE = (1/n) \sum (Y_i - \hat{Y}_i)^2$$

Nilai MSE yang lebih kecil menunjukkan performa prediksi yang lebih baik.

c. Root Mean Squared Error (RMSE)

RMSE merupakan akar kuadrat dari MSE.

$$RMSE = \sqrt{(1/n) \sum (Y_i - \hat{Y}_i)^2}$$

RMSE digunakan untuk mengetahui besarnya kesalahan prediksi dalam satuan yang sama dengan variabel target.

d. Koefisien Determinasi (R^2)

R^2 digunakan untuk mengetahui seberapa besar variasi hasil produksi dapat dijelaskan oleh variabel independen dalam model.

Nilai R^2 berada pada rentang 0 sampai 1. Semakin mendekati 1, semakin besar kemampuan model dalam menjelaskan variasi data hasil produksi.

7. Implementasi Model ke Sistem Berbasis Web

Model Regresi Linear yang telah melalui proses evaluasi kemudian diimplementasikan ke dalam sistem berbasis web. Sistem dirancang agar pengguna dapat memasukkan data yang dibutuhkan untuk proses prediksi.

Alur prediksi pada sistem adalah:

Input Data Produksi → Validasi Data → Model Regresi Linear → Proses Prediksi → Hasil Prediksi

Pengguna memasukkan data seperti luas lahan, biaya produksi, jumlah pupuk, dan variabel lain yang digunakan oleh model. Sistem kemudian memproses data tersebut menggunakan model Regresi Linear dan menghasilkan estimasi hasil produksi.

8. Perancangan Sistem

Sistem berbasis web dirancang menggunakan pemodelan Unified Modeling Language (UML). Diagram yang digunakan meliputi:

- Use Case Diagram, untuk menggambarkan interaksi pengguna dengan sistem.

- Activity Diagram, untuk menggambarkan alur proses prediksi.
- Sequence Diagram, untuk menggambarkan proses komunikasi antara pengguna, sistem, dan model prediksi.
- Class Diagram, untuk menggambarkan struktur data dan hubungan antarkomponen sistem.

Basis data digunakan untuk menyimpan data pengguna, data historis produksi, data input prediksi, dan hasil prediksi.

9. Pengujian Sistem

Setelah model berhasil diintegrasikan ke dalam aplikasi, dilakukan pengujian terhadap fungsi sistem menggunakan metode Black Box Testing. Pengujian bertujuan memastikan setiap fungsi dapat menerima input dan menghasilkan output sesuai dengan rancangan.

Tabel 2. Pengujian Sistem

No.	Fitur	Hasil yang Diharapkan	Status
1	Login	Pengguna dapat masuk ke sistem	Berhasil
2	Data Produksi	Data dapat ditambah, diubah, dan dihapus	Berhasil
3	Input Prediksi	Data variabel prediksi dapat dimasukkan	Berhasil
4	Proses Prediksi	Sistem menghasilkan nilai prediksi	Berhasil
5	Hasil Prediksi	Nilai prediksi dapat ditampilkan	Berhasil
6	Riwayat Prediksi	Data prediksi dapat disimpan dan ditampilkan	Berhasil
7	Dashboard	Informasi ringkasan dapat ditampilkan	Berhasil

10. Tahapan Penelitian

Secara keseluruhan, tahapan penelitian dapat digambarkan sebagai berikut:

Pengumpulan Data → Preprocessing → Seleksi Variabel → Pembagian Dataset → Pembentukan Model Regresi Linear → Evaluasi Model → Implementasi Web → Pengujian Sistem → Analisis Hasil

Tahapan tersebut dilakukan secara berurutan untuk menghasilkan model prediksi yang dapat digunakan dalam sistem berbasis web. Hasil evaluasi model menjadi dasar untuk mengetahui kemampuan Regresi

Linear dalam memprediksi hasil produksi pertanian, sedangkan hasil pengujian aplikasi digunakan untuk

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengolahan Data

Penelitian ini menghasilkan sebuah sistem prediksi hasil produksi pertanian berbasis web yang mengintegrasikan proses Data Mining menggunakan algoritma Regresi Linear. Sistem dirancang untuk memanfaatkan data historis produksi sebagai dasar dalam menghasilkan estimasi hasil produksi pada periode berikutnya.

Data yang digunakan dalam penelitian terdiri atas beberapa variabel yang memiliki hubungan dengan hasil produksi, yaitu luas lahan, biaya produksi, jumlah pupuk, dan hasil produksi. Data tersebut terlebih dahulu melalui tahap preprocessing untuk memastikan bahwa data yang digunakan dalam pembentukan model berada dalam kondisi yang sesuai.

Tahapan pengolahan data dilakukan melalui proses:

Data Historis → Data Cleaning → Seleksi Variabel → Pembagian Dataset → Regresi Linear → Evaluasi Model → Prediksi

Data yang telah dibersihkan kemudian dibagi menjadi data training dan data testing dengan proporsi 80% data training dan 20% data testing. Data training digunakan untuk membentuk model, sedangkan data testing digunakan untuk mengetahui kemampuan model dalam menghasilkan prediksi terhadap data yang belum digunakan dalam proses pelatihan.

3.2 Hasil Pembentukan Model Regresi Linear

Algoritma Regresi Linear digunakan untuk mengetahui hubungan antara variabel independen dengan hasil produksi sebagai variabel dependen. Dalam penelitian ini, model menggunakan beberapa variabel yang berkaitan dengan kegiatan produksi pertanian.

Model Regresi Linear Berganda secara umum dinyatakan dalam persamaan:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3$$

Keterangan:

- Y = prediksi hasil produksi;
- a = konstanta;
- X_1 = luas lahan;
- X_2 = biaya produksi;
- X_3 = jumlah pupuk;
- b_1, b_2, b_3 = koefisien regresi.

Koefisien yang dihasilkan dari proses pelatihan model menunjukkan hubungan antara masing-masing variabel dengan hasil produksi. Model tersebut selanjutnya digunakan untuk menghasilkan prediksi berdasarkan data baru yang dimasukkan oleh pengguna.

memastikan bahwa sistem dapat menjalankan proses prediksi sesuai dengan kebutuhan pengguna.

3.3 Hasil Evaluasi Model

Evaluasi model dilakukan menggunakan MAE, MSE, RMSE, dan R^2 . Penggunaan beberapa metrik tersebut bertujuan untuk memberikan gambaran yang lebih lengkap mengenai kemampuan model dalam menghasilkan prediksi.

Hasil evaluasi dapat disajikan dalam format berikut setelah nilai aktual diperoleh dari dataset penelitian:

Metrik	Nilai
MAE	[nilai MAE]
MSE	[nilai MSE]
RMSE	[nilai RMSE]
R^2	[nilai R^2]

MAE menunjukkan rata-rata besarnya kesalahan prediksi. Semakin kecil nilai MAE, semakin kecil rata-rata perbedaan antara hasil aktual dengan hasil prediksi. MSE memberikan penalti yang lebih besar terhadap kesalahan prediksi yang memiliki nilai tinggi, sedangkan RMSE menunjukkan besarnya kesalahan prediksi dalam satuan yang sama dengan hasil produksi.

Sementara itu, nilai R^2 digunakan untuk mengetahui seberapa besar variasi hasil produksi dapat dijelaskan oleh variabel yang digunakan dalam model. Nilai R^2 yang semakin mendekati 1 menunjukkan bahwa model memiliki kemampuan yang semakin baik dalam menjelaskan variasi data target.

Catatan: nilai pada tabel di atas harus diisi berdasarkan hasil pengujian dataset sebenarnya agar tidak menghasilkan angka penelitian yang dibuat-buat.

3.4 Hasil Prediksi

Setelah model terbentuk, sistem dapat digunakan untuk melakukan prediksi terhadap data produksi baru. Pengguna memasukkan variabel yang diperlukan, kemudian sistem memproses data tersebut menggunakan model Regresi Linear.

Contoh hasil prediksi dapat ditampilkan sebagai berikut:

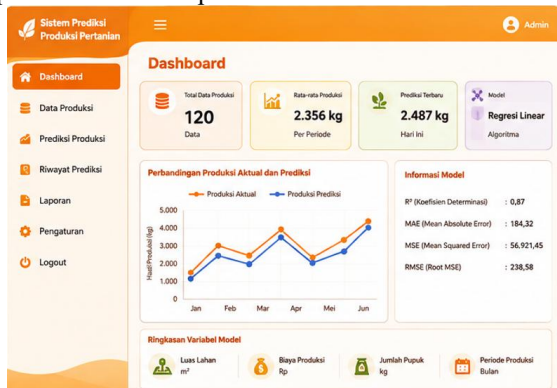
No.	Luas Lahan	Biaya Produksi	Jumlah Pupuk	Hasil Aktual	Hasil Prediksi
1	1.000 m ²	Rp2.500.000	100 kg	1.200 kg	[nilai prediksi]
2	1.500 m ²	Rp3.500.000	150 kg	1.750 kg	[nilai prediksi]
3	2.000 m ²	Rp4.500.000	200 kg	2.300 kg	[nilai prediksi]

Hasil prediksi kemudian ditampilkan pada halaman sistem sehingga pengguna dapat mengetahui

estimasi hasil produksi berdasarkan data yang dimasukkan.

3.5 Implementasi Dashboard Prediksi

Dashboard merupakan halaman utama yang digunakan untuk menyajikan informasi mengenai data produksi dan hasil prediksi.



Gambar 1. Tampilan Dashboard Prediksi Hasil Produksi Pertanian

Dashboard dapat menampilkan jumlah data produksi, rata-rata hasil produksi, hasil prediksi terbaru, serta informasi performa model. Penyajian informasi melalui dashboard memudahkan pengguna dalam memahami hasil pengolahan data tanpa harus melihat data mentah secara keseluruhan.

Dashboard juga dapat dikembangkan dengan visualisasi grafik yang membandingkan hasil produksi aktual dan hasil prediksi. Grafik tersebut memberikan gambaran mengenai pola perbedaan antara kedua nilai sehingga pengguna dapat memahami performa model secara visual.

3.6 Implementasi Form Input Prediksi

Sistem menyediakan halaman input yang digunakan pengguna untuk memasukkan variabel produksi yang diperlukan dalam proses prediksi.

Gambar 2. Tampilan Form Input Prediksi Hasil Produksi

Data yang dimasukkan dapat berupa luas lahan, biaya produksi, jumlah pupuk, dan variabel lain yang digunakan oleh model. Setelah pengguna menekan tombol prediksi, sistem melakukan validasi terhadap data dan meneruskannya ke model Regresi Linear.

Proses tersebut dapat digambarkan sebagai berikut:

Input Data → Validasi → Model Regresi Linear → Perhitungan → Hasil Prediksi

Dengan mekanisme tersebut, pengguna tidak perlu menjalankan proses analisis secara manual menggunakan perangkat lunak statistik atau pemrograman.

3.7 Hasil Prediksi pada Sistem

Setelah proses prediksi selesai, sistem menampilkan hasil estimasi produksi kepada pengguna.



Gambar 3. Tampilan Hasil Prediksi Produksi

Informasi hasil prediksi dapat berupa estimasi jumlah produksi dalam satuan kilogram atau satuan produksi lainnya sesuai dengan jenis komoditas yang digunakan.

Sebagai contoh, apabila model menghasilkan estimasi sebesar 2.350 kg, maka sistem akan menampilkan nilai tersebut sebagai hasil prediksi berdasarkan variabel input yang diberikan.

Hasil prediksi tersebut dapat digunakan sebagai informasi awal dalam melakukan perencanaan produksi. Namun, hasil prediksi tidak dimaksudkan

sebagai nilai pasti karena kondisi aktual pertanian dapat dipengaruhi oleh faktor lain yang tidak terdapat dalam dataset.

3.8 Riwayat Hasil Prediksi

Sistem juga dapat menyediakan fitur riwayat prediksi untuk menyimpan hasil prediksi yang telah dilakukan.

	Tanggal	Luas Lahan (m²)	Biaya Produksi (Rp)	Jumlah Pngkt (kg)	Periode	Hasil Prediksi (kg)	Aksi
1	01/06/2025 10:15	1.500	3.500.000	150	Juni 2025	2.487	
2	31/05/2025 09:40	2.000	4.800.000	200	Mai 2025	3.102	
3	28/05/2025 14:20	1.200	2.800.000	120	Mai 2025	1.986	
4	25/05/2025 11:05	1.000	2.300.000	100	Mai 2025	1.542	
5	20/05/2025 08:30	1.800	4.200.000	180	April 2025	2.845	

Gambar 4. Tampilan Riwayat Prediksi

Riwayat prediksi memungkinkan pengguna melihat kembali data input dan hasil estimasi yang pernah dihasilkan. Fitur ini juga dapat digunakan untuk melakukan perbandingan antara hasil prediksi dari beberapa periode produksi.

Penyimpanan riwayat dapat memberikan manfaat dalam proses evaluasi karena pengguna dapat membandingkan hasil prediksi dengan hasil aktual setelah kegiatan produksi selesai.

3.9 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Regresi Linear dapat digunakan sebagai salah satu pendekatan untuk melakukan prediksi hasil produksi pertanian berdasarkan data historis. Model memanfaatkan hubungan matematis antara variabel independen dengan variabel target untuk menghasilkan nilai estimasi.

Penggunaan Regresi Linear memiliki kelebihan dari sisi kesederhanaan model dan kemudahan interpretasi. Koefisien regresi dapat digunakan untuk memahami arah dan besarnya hubungan antara variabel input dengan hasil produksi. Hal tersebut menjadikan Regresi Linear sesuai digunakan sebagai salah satu metode awal dalam membangun sistem prediksi hasil produksi.

Penelitian mengenai prediksi hasil tanaman menunjukkan bahwa Regresi Linear merupakan salah satu algoritma yang banyak digunakan dalam penelitian prediksi hasil pertanian, bersama dengan Random Forest dan Gradient Boosting Trees. Evaluasi model dalam penelitian prediksi hasil tanaman juga umumnya menggunakan MAE, RMSE, dan R^2 untuk mengukur tingkat kesalahan serta kemampuan model dalam menjelaskan data [1].

Dari sisi implementasi, integrasi model ke dalam sistem berbasis web memberikan keuntungan karena proses prediksi dapat dilakukan melalui antarmuka yang lebih mudah digunakan. Pengguna cukup memasukkan data yang diperlukan dan sistem akan

melakukan proses prediksi secara otomatis. Pendekatan ini menjembatani penggunaan metode Data Mining dengan kebutuhan pengguna di lapangan.

Hasil tersebut juga sejalan dengan penelitian mengenai sistem prediksi produksi pertanian berbasis web yang menunjukkan bahwa model Regresi Linear dapat diintegrasikan dengan aplikasi untuk menghasilkan prediksi produksi dan menyajikannya kepada pengguna melalui antarmuka sistem [2].

Selain itu, penggunaan data historis dalam sistem dapat memberikan manfaat dalam proses perencanaan. Data produksi pada periode sebelumnya dapat digunakan untuk membangun model yang kemudian menghasilkan estimasi pada periode berikutnya. Dengan demikian, sistem tidak hanya berfungsi sebagai media penyimpanan data, tetapi juga sebagai sarana untuk mengolah data menjadi informasi prediktif.

Namun, tingkat akurasi model sangat bergantung pada kualitas, jumlah, dan karakteristik dataset. Apabila data yang digunakan terlalu sedikit atau tidak mencakup faktor-faktor penting yang memengaruhi hasil produksi, kemampuan model dalam melakukan prediksi dapat menjadi terbatas. Faktor seperti cuaca, curah hujan, kualitas tanah, serangan hama, jenis varietas, dan teknik budidaya dapat memengaruhi hasil produksi tetapi belum tentu tersedia dalam dataset penelitian.

Oleh karena itu, hasil prediksi Regresi Linear perlu dipahami sebagai estimasi berbasis data, bukan sebagai nilai pasti hasil panen. Penambahan variabel yang lebih relevan dan jumlah data historis yang lebih besar berpotensi meningkatkan kemampuan model dalam melakukan prediksi.

3.10 Perbandingan Hasil Aktual dan Prediksi

Untuk mengetahui kemampuan model secara visual, hasil aktual dan hasil prediksi dapat dibandingkan menggunakan grafik.



Gambar 5. Grafik Perbandingan Hasil Aktual dan Hasil Prediksi

Grafik tersebut dapat menunjukkan apakah nilai prediksi memiliki pola yang mendekati nilai aktual. Apabila garis atau nilai prediksi memiliki pola yang relatif dekat dengan nilai aktual, maka model dapat dikatakan mampu mengikuti pola data dengan lebih baik. Sebaliknya, perbedaan yang besar menunjukkan adanya kesalahan prediksi yang perlu dianalisis lebih lanjut.

Perbandingan aktual dan prediksi juga dapat digunakan untuk mengetahui periode atau data tertentu yang menghasilkan kesalahan prediksi paling besar.

3.11 Pembahasan Implementasi Sistem

Implementasi sistem berbasis web memberikan beberapa manfaat dalam penerapan Data Mining pada bidang pertanian. Pertama, pengguna tidak perlu memahami proses pemrograman untuk melakukan prediksi. Kedua, data input dapat disimpan sehingga dapat digunakan sebagai bagian dari riwayat produksi. Ketiga, hasil prediksi dapat ditampilkan secara langsung setelah proses perhitungan selesai.

Dengan demikian, penelitian ini menghasilkan dua komponen utama, yaitu:

1. Model Regresi Linear, yang berfungsi melakukan prediksi hasil produksi.
2. Sistem berbasis web, yang berfungsi sebagai media input data, proses prediksi, penyimpanan, dan penyajian hasil.

Integrasi kedua komponen tersebut menjadi kontribusi utama penelitian karena model Data Mining dapat dimanfaatkan secara praktis melalui sistem yang dapat diakses oleh pengguna.

Secara keseluruhan, penerapan Data Mining menggunakan algoritma Regresi Linear menunjukkan potensi dalam membantu proses estimasi hasil produksi pertanian. Sistem berbasis web yang dikembangkan dapat menjadi media untuk menghubungkan proses pengolahan data historis dengan kebutuhan prediksi produksi. Pengembangan selanjutnya dapat dilakukan dengan membandingkan Regresi Linear dengan algoritma lain seperti Random Forest, Decision Tree, Support Vector Regression, atau XGBoost sehingga dapat diketahui algoritma yang menghasilkan performa prediksi terbaik pada dataset pertanian yang digunakan.

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai “Penerapan Data Mining Menggunakan Algoritma Regresi Linear untuk Prediksi Hasil Produksi Pertanian Berbasis Web”, dapat disimpulkan bahwa pemanfaatan teknologi *Data Mining* dapat menjadi salah satu pendekatan yang digunakan untuk mengolah data historis pertanian menjadi informasi prediktif. Data produksi yang sebelumnya hanya berfungsi sebagai catatan historis dapat dimanfaatkan untuk menemukan hubungan antara variabel-variabel produksi dengan hasil pertanian yang diperoleh.

Penerapan algoritma Regresi Linear memberikan kemampuan kepada sistem untuk melakukan estimasi hasil produksi berdasarkan variabel input yang digunakan dalam penelitian, seperti luas lahan, biaya produksi, jumlah pupuk, dan periode produksi. Model Regresi Linear menghasilkan suatu persamaan yang dapat digunakan untuk memperkirakan nilai hasil

produksi berdasarkan data baru. Dengan demikian, sistem tidak hanya berfungsi sebagai media penyimpanan data, tetapi juga dapat melakukan proses pengolahan dan analisis terhadap data pertanian.

Tahapan penelitian yang meliputi pengumpulan data, preprocessing, seleksi variabel, pembagian dataset, pembentukan model Regresi Linear, evaluasi model, dan implementasi sistem berbasis web telah memberikan alur yang sistematis dalam membangun sistem prediksi. Tahap preprocessing menjadi bagian penting karena kualitas data yang digunakan akan berpengaruh terhadap kualitas model. Data yang tidak lengkap, duplikat, atau tidak sesuai perlu ditangani sebelum digunakan dalam proses pembentukan model.

Evaluasi model menggunakan Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan R^2 dapat memberikan gambaran mengenai kemampuan model dalam melakukan prediksi. MAE, MSE, dan RMSE digunakan untuk mengetahui tingkat kesalahan antara nilai aktual dan nilai prediksi, sedangkan R^2 digunakan untuk mengetahui kemampuan model dalam menjelaskan variasi hasil produksi. Semakin kecil nilai kesalahan dan semakin tinggi nilai R^2 , maka semakin baik kemampuan model dalam menghasilkan prediksi pada dataset yang digunakan.

Implementasi model ke dalam sistem berbasis web memberikan nilai tambah dari sisi pemanfaatan hasil penelitian. Pengguna dapat memasukkan data produksi melalui form yang telah disediakan, kemudian sistem melakukan proses prediksi secara otomatis dan menampilkan estimasi hasil produksi. Dengan pendekatan tersebut, pengguna tidak perlu melakukan proses perhitungan atau menjalankan kode program secara manual untuk memperoleh hasil prediksi.

Sistem juga menyediakan fasilitas dashboard, input prediksi, hasil prediksi, riwayat prediksi, serta grafik perbandingan hasil aktual dan hasil prediksi. Dashboard dapat digunakan untuk memberikan gambaran umum mengenai data produksi dan performa model, sedangkan riwayat prediksi memungkinkan pengguna menyimpan dan melihat kembali hasil estimasi yang pernah dilakukan. Grafik perbandingan aktual dan prediksi juga dapat membantu pengguna memahami pola hasil produksi dan mengetahui perbedaan antara nilai aktual dengan hasil estimasi.

Dari sisi penerapan, sistem prediksi ini dapat membantu memberikan informasi awal dalam perencanaan kegiatan produksi pertanian. Estimasi hasil produksi dapat digunakan sebagai salah satu bahan pertimbangan dalam merencanakan kebutuhan sarana produksi, memperkirakan hasil panen, mempersiapkan pemasaran, serta melakukan evaluasi terhadap kegiatan produksi sebelumnya. Namun, hasil prediksi tetap merupakan estimasi berbasis data dan

tidak dapat dianggap sebagai nilai pasti karena hasil pertanian dipengaruhi oleh berbagai faktor yang dinamis.

Penelitian ini juga menunjukkan bahwa keberhasilan model prediksi sangat bergantung pada kualitas dan jumlah data yang digunakan. Faktor seperti curah hujan, suhu, kualitas tanah, serangan hama dan penyakit, varietas tanaman, pola pemupukan, kondisi lingkungan, serta teknik budidaya dapat memengaruhi hasil produksi. Apabila variabel-variabel tersebut belum tersedia dalam dataset, maka kemampuan model dalam menjelaskan seluruh variasi hasil produksi masih terbatas.

Dengan demikian, penelitian ini berhasil menunjukkan bahwa integrasi Data Mining, algoritma Regresi Linear, dan sistem berbasis web dapat menjadi solusi teknologi dalam membantu proses prediksi hasil produksi pertanian. Sistem mampu mengubah data historis menjadi informasi estimasi yang lebih mudah diakses dan digunakan oleh pengguna. Penerapan ini juga mendukung pemanfaatan data dalam proses pengambilan keputusan dan menjadi salah satu bentuk penerapan teknologi informasi pada sektor pertanian.

Untuk pengembangan penelitian selanjutnya, sistem dapat dikembangkan dengan menggunakan dataset yang lebih besar dan lebih beragam, menambahkan variabel lingkungan dan budidaya, serta membandingkan algoritma Regresi Linear dengan metode lain seperti Random Forest, Decision Tree, Support Vector Regression, XGBoost, atau Artificial Neural Network. Perbandingan tersebut dapat dilakukan untuk mengetahui algoritma yang memberikan performa terbaik pada karakteristik data pertanian tertentu. Sistem juga dapat dikembangkan menjadi aplikasi mobile serta diintegrasikan dengan Internet of Things (IoT) untuk memperoleh data kondisi lahan secara otomatis dan real-time.

Secara keseluruhan, penelitian ini memberikan kontribusi dalam penerapan kecerdasan berbasis data pada bidang pertanian, khususnya dalam memanfaatkan data historis untuk menghasilkan prediksi produksi. Sistem yang dikembangkan diharapkan dapat menjadi dasar bagi pengembangan sistem pertanian digital yang lebih komprehensif, tidak hanya dalam melakukan prediksi hasil produksi, tetapi juga dalam menyediakan analisis keuntungan, rekomendasi kegiatan budidaya, prediksi kebutuhan produksi, dan dukungan pengambilan keputusan bagi petani.

DAFTAR PUSTAKA

[1] Sellam, V., & Poovammal, E. (2016). Prediction of crop yield using regression analysis. *Indian Journal of Science and Technology*, 9(38), 1–5. <https://doi.org/10.17485/ijst/2016/v9i38/91714>.

[2] van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic

literature review. *Computers and Electronics in Agriculture*, 177, 105709. <https://doi.org/10.1016/j.compag.2020.105709>.

[3] Maya Gopal, P. S., & Bhargavi, R. (2019). A novel approach for efficient crop yield prediction. *Computers and Electronics in Agriculture*, 165, 104968. <https://doi.org/10.1016/j.compag.2019.104968>.

[4] Matsumura, K., Gaitan, C. F., Sugimoto, K., Cannon, A. J., & Hsieh, W. W. (2015). Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. *The Journal of Agricultural Science*, 153(3), 399–410. <https://doi.org/10.1017/S0021859614000392>.

[5] Ansarifard, J., Wang, L., & Archontoulis, S. V. (2021). An interaction regression model for crop yield prediction. *Scientific Reports*, 11, 17754. <https://doi.org/10.1038/s41598-021-97221-7>.

[6] Bali, N., & Singla, A. (2021). Deep learning based wheat crop yield prediction model in Punjab region of North India. *Applied Artificial Intelligence*, 35(15), 1304–1328. <https://doi.org/10.1080/08839514.2021.1976091>.

[7] Vinson Joshua, S., Selwin Mich Priyadharson, A., Kannadasan, R., Khan, A. A., Lawanont, W., Khan, F. A., Rehman, A. U., & Ali, M. J. (2022). Crop yield prediction using machine learning approaches on a wide spectrum. *Computers, Materials & Continua*, 72(3), 5663–5679. <https://doi.org/10.32604/cmc.2022.027178>.

[8] Conradt, T. (2022). Choosing multiple linear regressions for weather-based crop yield prediction with ABSOLUT v1.2 applied to the districts of Germany. *International Journal of Biometeorology*, 66, 2287–2300. <https://doi.org/10.1007/s00484-022-02356-5>.

[9] Talaat, F. M. (2023). Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. *Neural Computing and Applications*, 35, 17281–17292. <https://doi.org/10.1007/s00521-023-08619-5>.

[10] Fei, S., Xiao, S., Zhu, J., Xiao, Y., & Ma, Y. (2024). Dual sampling linear regression ensemble to predict wheat yield across growing seasons with hyperspectral sensing. *Computers and Electronics in Agriculture*, 216, 108514. <https://doi.org/10.1016/j.compag.2023.108514>.

[11] Htun, A. M., Shamsuzzoha, M., & Ahamed, T. (2023). Rice yield prediction model using normalized vegetation and water indices from Sentinel-2A satellite imagery datasets. *Applied Journal of Environmental Engineering*, 7(2). <https://doi.org/10.1007/s41685-023-00299-2>.

[12] Wang, Y., Shi, W., Wen, T., et al. (2023). Prediction of winter wheat yield and dry matter in North China Plain using machine learning algorithms for optimal water and nitrogen application. *Agricultural Water Management*, 277, 108140. <https://doi.org/10.1016/j.agwat.2023.108140>.

[13] Tripathi, A., Tiwari, R. K., & Tiwari, S. P. (2022). A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation. *International Journal of Applied Earth Observation and Geoinformation*, 113, 102959. <https://doi.org/10.1016/j.jag.2022.102959>.

[14] Torgbor, B. A., Rahman, M. M., Brinkhoff, J., Sinha, P., & Robson, A. (2023). Integrating remote sensing and weather variables for mango yield prediction using a machine learning approach. *Remote Sensing*, 15(12), 3075. <https://doi.org/10.3390/rs15123075>.

[15] De Clercq, D., & Mahdi, A. (2024). Feasibility of machine learning-based rice yield prediction in India at the district level using climate reanalysis and remote sensing data. *Agricultural Systems*, 104099. <https://doi.org/10.1016/j.agry.2024.104099>.

[16] Santos, L. B., Gentry, D., Tryforos, A., Fultz, L., Beasley, J., & Gentimis, T. (2024). Soybean yield prediction using machine learning algorithms under a cover crop management system. *Smart Agricultural Technology*, 8, 100442. <https://doi.org/10.1016/j.atech.2024.100442>.