

Implementasi Algoritma Komputasi Linear Regression untuk Optimasi Prediksi Hasil Pertanian

Irma Hakim^{1*}, Asdi², Teuku Afriliansyah³

^{1,2}Universitas Muhammadiyah Makassar, Makassar, Indonesia

³Universitas Bumi Persada Lhokseumawe, Aceh, Indonesia

E-mail: campus_gardenia@yahoo.co.id¹, asdi@unismuh.ac.id²,
teukuafriansyah@unbp.ac.id³

Abstrak

Tujuan utama dari penelitian ini adalah untuk mengimplementasikan algoritma komputasi Regresi Linier untuk memprediksi hasil panen dengan lebih akurat. Metode penelitiannya meliputi pengumpulan dan analisis data historis dari 10 sampel pertanian yang mencakup variabel-variabel tersebut. Data ini kemudian digunakan untuk melatih model prediksi. Evaluasi model menggunakan metrik skor Mean Squared Error (MSE) dan R^2 untuk menilai akurasi prediksi. Hasil penelitian menunjukkan bahwa model Regresi Linier dapat memberikan prediksi yang akurat, dengan hasil prediksi pada data baru mencapai 479,5 kg/ha. Visualisasi data menunjukkan adanya hubungan yang signifikan antara variabel lingkungan dan hasil panen, sehingga mendukung validitas model yang dibangun. Kesimpulan dari penelitian ini menegaskan bahwa penerapan algoritma komputasi dapat menjadi alat yang efektif untuk membantu petani membuat keputusan yang lebih tepat mengenai waktu tanam dan strategi pengelolaan lahan. Hal ini tidak hanya meningkatkan efisiensi dan produktivitas pertanian tetapi juga membantu mengurangi ketidakpastian hasil panen. Penerapan teknologi menggunakan algoritma regresi linier diharapkan dapat memberikan kontribusi yang signifikan terhadap praktik pertanian yang lebih berkelanjutan dan efisien, serta mendukung peningkatan hasil panen di masa depan.

Kata Kunci: Hasil Pertanian, Machine Learning, Linear Regression, Prediksi, Panen

Abstract

The main objective of this research is to implement the Linear Regression computational algorithm to predict crop yields more accurately. The research method includes collecting and analyzing historical data from 10 agricultural samples that include these variables. This data is then used to train a prediction model. The model evaluation used the Mean Squared Error (MSE) and R^2 score metrics to assess prediction accuracy. The research results show that the Linear Regression model can provide accurate predictions, with prediction results on new data reaching 479.5 kg/ha. Data visualization revealed a significant relationship between environmental variables and crop yields, which supports the validity of the model constructed. The conclusions of this research confirm that implementing computational algorithms can be an effective tool to help farmers make more informed decisions regarding planting times and land management strategies. This not only increases agricultural efficiency and productivity but also helps in reducing uncertainty in crop yields. The implementation of technology using the linear regression algorithm is expected to make a significant contribution to more sustainable and efficient agricultural practices, as well as support increased crop yields in the future.

Keywords: Agricultural Results, Machine Learning, Linear Regression, Prediction, Harvest

1. Pendahuluan

Pertanian merupakan sektor vital yang mendukung ketahanan pangan global [1]–[3]. Namun, salah satu tantangan utama yang dihadapi oleh petani adalah ketidakpastian dalam hasil panen yang disebabkan oleh berbagai faktor lingkungan [4], seperti: suhu [5], kelembaban [6], pH tanah [7], dan curah hujan [8]. Variabilitas iklim dan perubahan kondisi tanah dapat secara signifikan mempengaruhi produktivitas pertanian [9]–[12], menyebabkan fluktuasi hasil yang tidak diinginkan dan mengurangi efektivitas upaya petani dalam mengelola lahan. Ketidakpastian ini menimbulkan risiko ekonomi bagi petani, terutama dalam pengambilan keputusan terkait waktu tanam, penggunaan pupuk, dan teknik pengelolaan lahan. Sebagai upaya untuk mengatasi masalah ini, terdapat kebutuhan mendesak untuk mengembangkan metode prediksi yang lebih akurat dan dapat diandalkan. Teknologi modern, khususnya algoritma komputasi seperti Linear Regression, menawarkan potensi besar dalam memproses dan menganalisis data lingkungan untuk memberikan prediksi yang lebih presisi, dengan kemampuan untuk memanfaatkan data historis dan variabel lingkungan yang relevan, algoritma ini dapat membantu dalam memprediksi hasil panen dengan tingkat akurasi yang tinggi. Hal ini tidak hanya membantu petani dalam merencanakan aktivitas pertanian mereka secara lebih efisien, tetapi juga memungkinkan mereka untuk mengoptimalkan penggunaan sumber daya dan mengurangi kerugian akibat perubahan kondisi lingkungan.

Algoritma Linear Regression memegang peranan penting dalam penelitian ini karena menyediakan pendekatan yang sederhana namun kuat untuk memodelkan hubungan antara variabel independen dan variabel dependen [13], dalam hal ini faktor lingkungan dan hasil panen. Algoritma ini digunakan untuk memprediksi hasil panen berdasarkan data historis yang mencakup variabel-variabel seperti suhu, kelembaban, pH tanah, dan curah hujan. Beberapa alasan pentingnya penggunaan Linear Regression dalam penelitian ini diantaranya: Algoritma tersebut sederhana dan mudah diinterpretasikan. Koefisien yang dihasilkan oleh model Linear Regression menunjukkan sejauh mana perubahan pada setiap variabel input mempengaruhi output. Hal ini memudahkan para peneliti dan praktisi pertanian untuk memahami dan menerapkan hasil penelitian dalam praktik sehari-hari. Linear Regression sangat efektif untuk memodelkan hubungan linier antara variabel input dan output tersebut [14]. Banyak faktor lingkungan yang memiliki hubungan linier dengan hasil panen. Misalnya, hingga titik tertentu, peningkatan curah hujan dapat meningkatkan hasil panen [15]–[17]. Linear Regression membantu mengidentifikasi dan mengukur kekuatan hubungan ini, sehingga memberikan pemahaman yang lebih baik tentang bagaimana setiap faktor mempengaruhi hasil panen. Penelitian dengan menggunakan algoritma ini dapat menghasilkan model prediksi yang memberikan estimasi akurat tentang hasil panen berdasarkan kondisi lingkungan. Ini sangat penting bagi petani dalam merencanakan dan mengelola kegiatan pertanian mereka, seperti menentukan waktu tanam yang optimal atau memutuskan strategi pengelolaan sumber daya. Prediksi yang akurat dapat membantu mengurangi ketidakpastian dan risiko, serta meningkatkan efisiensi dan produktivitas pertanian. Secara keseluruhan, penggunaan algoritma Linear Regression dalam penelitian ini memungkinkan analisis data yang mendalam dan prediksi yang dapat diandalkan, yang pada akhirnya membantu meningkatkan produktivitas dan keberlanjutan pertanian.

Implementasi algoritma Linear Regression untuk prediksi hasil pertanian dalam penelitian ini menghadapi beberapa tantangan utama yang memerlukan perhatian khusus. Salah satu tantangan terbesar adalah keterbatasan dalam jumlah dan kualitas data yang tersedia. Data historis yang mencakup variabel penting seperti suhu, kelembaban, pH tanah, dan curah hujan mungkin tidak selalu lengkap atau akurat. Ketidakeengkapan data dan adanya *missing values* dapat mempengaruhi keakuratan model prediksi, karena model yang dibangun berdasarkan data yang tidak lengkap atau tidak representatif mungkin tidak mampu mencerminkan kondisi sebenarnya di lapangan. Selain itu, variasi data yang

terbatas mungkin tidak mencerminkan semua skenario pertanian yang mungkin terjadi, sehingga mengurangi generalisasi model dan membatasi kemampuannya dalam membuat prediksi yang akurat di luar sampel data yang digunakan. Faktor lingkungan, seperti perubahan cuaca dan fluktuasi curah hujan, sering kali tidak dapat diprediksi dengan presisi tinggi dan dapat berubah dengan cepat, mempengaruhi hasil panen secara signifikan. Model Linear Regression, yang umumnya memodelkan hubungan linier, mungkin tidak cukup fleksibel untuk menangkap kompleksitas dan dinamika interaksi antara berbagai faktor lingkungan ini. Hal ini menimbulkan risiko model menjadi kurang akurat dalam menghadapi variabilitas lingkungan yang tinggi. Tantangan lain dalam implementasi adalah keseimbangan antara overfitting dan underfitting. Overfitting terjadi ketika model terlalu kompleks dan menyesuaikan terlalu erat dengan data latih, yang mengakibatkan kinerja buruk pada data uji atau data baru [18]. Sebaliknya, underfitting terjadi ketika model terlalu sederhana dan gagal menangkap pola penting dalam data, sehingga juga mengurangi akurasi prediksi [19]. Untuk mengatasi masalah ini, diperlukan penyesuaian parameter model yang tepat serta penggunaan teknik regularisasi untuk mencapai keseimbangan yang optimal. Selain tantangan teknis, ada juga tantangan dalam integrasi model prediksi dengan praktik pertanian nyata. Petani mungkin menghadapi kesulitan dalam memahami atau menerapkan saran atau rekomendasi yang dihasilkan oleh model, terutama jika tidak ada infrastruktur atau teknologi yang mendukung untuk pengumpulan data dan analisis secara real-time. Hal ini dapat menghambat adopsi teknologi oleh petani, terutama di daerah yang kurang berkembang secara teknologi. Untuk mengatasi tantangan ini, diperlukan pendekatan komprehensif yang mencakup pengumpulan data yang lebih baik, penggunaan teknik machine learning yang lebih canggih untuk mengatasi kompleksitas data, serta dukungan dan pelatihan bagi petani dalam menggunakan teknologi ini secara efektif.

Beberapa referensi yang melatarbelakangi dilakukan nya penelitian ini diantaranya: Penelitian untuk memprediksi hasil pertanian durian dengan metode regresi linear berganda. Penelitian ini dibangun dengan membuat sistem informasi berbasis Web menggunakan algoritma tersebut dengan pemrograman PHP. Akan tetapi penelitian ini tidak menjelaskan bagaimana proses dan tahapan secara lengkap bagaimana algoritma tersebut bekerja [20]. Penelitian selanjutnya memanfaatkan metode *machine learning* untuk memprediksi Hasil Panen Tanaman Pangan di Sumatera. Metode Machine Learning yang digunakan yakni; Random Forest (RF), Decision Tree (DT), Gradient Boosting (GB), Extra Tree (ET), Support Vector Machine (SVM), dan Artificial Neural Network (ANN). Nilai koefisien R² untuk produksi padi masing-masing adalah 0,897; 0,893; 0,957; 0,968; 0,928; dan 0,909. Sedangkan pada produksi bahan pokok lainnya masing-masing adalah 0,754; 0,786; 0,721; 0,913; 0,509; dan 0,90. Model Extra Tree mendapat nilai koefisien R² tertinggi dan lebih akurat dibandingkan model lainnya [21]. Tetapi penelitian ini hanya menggunakan data hasil produksi tahunan dan luas panen atau luas lahan saja, tidak menganalisis kondisi lingkungan seperti suhu, kelembaban, pH tanah, dan curah hujan. Berikutnya penelitian dengan menggunakan algoritma regresi linear berganda untuk melakukan peramalan hasil panen padi kabupaten Cirebon. Hasil evaluasi peramalan pada penelitian ini 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 [22]. Penelitian ini juga tidak menganalisis kondisi lingkungan yang dapat mempengaruhi hasil panen. Selanjutnya penelitian menggunakan Regresi Linear untuk menganalisis hasil tanaman perkebunan kopi dan teh. Berdasarkan hasil persamaan regresi yang diperoleh pada penelitian ini, diperoleh interpretasi yaitu nilai konstanta dalam penelitian ini adalah sebesar 18,434. Hal tersebut menyatakan bahwa, jika nilai luas lahan (X) adalah bernilai nol (0) maka untuk nilai produksi teh (Y) adalah sebesar 18.434 dan untuk nilai koefisien variabel luas lahan (X) dalam penelitian ini adalah sebesar 1,232. Hal ini menyatakan

bahwa, pada setiap peningkatan 1% pada luas lahan maka dapat menaikkan nilai produksi teh yaitu sebesar 1,232. Dalam tabel ini juga dapat dilihat uji signifikansi uji t dengan perolehan nilai t hitung adalah sebesar 8,478 dengan nilai signifikansi $0,000 < 0,05$. Sehingga diperoleh untuk hasil hipotesis H_0 ditolak dan H_1 diterima, dimana hal tersebut berarti bahwa dalam penelitian ini terdapat pengaruh yang nyata (signifikan) antara variabel bebas (X) terhadap variabel terikat (Y) [23]. Penelitian ini juga hanya menganalisis luas lahan saja yang mempengaruhi terhadap hasil produksi tanaman perkebunan.

Berdasarkan hal tersebut, maka penelitian ini bertujuan mengembangkan dan mengimplementasikan model prediksi hasil panen yang lebih akurat menggunakan algoritma Linear Regression untuk mengoptimalkan hasil pertanian. Tujuan utamanya meliputi identifikasi faktor-faktor kunci seperti suhu, kelembaban, pH tanah, dan curah hujan yang mempengaruhi hasil panen, serta pengembangan model prediksi yang mampu menghasilkan estimasi akurat. Kinerja model akan dievaluasi menggunakan metrik seperti Mean Squared Error (MSE) dan R^2 score. Penelitian ini juga berfokus pada visualisasi hubungan antara variabel lingkungan dan hasil panen untuk memperjelas dampaknya, serta menyediakan panduan praktis bagi petani dalam pengambilan keputusan terkait waktu tanam, pemilihan varietas, dan strategi pengelolaan lahan. Melalui pendekatan ini, diharapkan dapat mendukung praktik pertanian berbasis data, mengurangi ketidakpastian hasil, dan meningkatkan keberlanjutan serta efisiensi sektor pertanian. Hasil penelitian diharapkan memberikan kontribusi signifikan terhadap pengembangan teknologi prediksi yang dapat diimplementasikan secara luas di sektor pertanian.

2. Metodologi Penelitian

2.1. Data Penelitian

Data penelitian yang digunakan diperoleh dari observasi lapangan yang melibatkan petani lokal di Kecamatan Bone yang merupakan daerah yang aktif dalam produksi jagung di Sulawesi Selatan, untuk memastikan keakuratan dan relevansi data yang digunakan dalam model prediksi hasil panen berbasis algoritma Linear Regression. Data penelitian ini diambil selama periode sepuluh musim tanam (tahun 2019-2023), yang mencakup berbagai kondisi lingkungan untuk memberikan gambaran yang lebih representatif. Setiap entri dalam dataset ini mewakili satu musim tanam dengan parameter-parameter seperti temperatur, kelembaban, pH tanah, curah hujan, dan hasil panen yang telah dicatat. Data tersebut dapat dilihat pada Tabel 1 berikut ini:

Tabel 1. Data Hasil Panen Padi Berdasarkan Kondisi Lingkungan

Temperatur (°C)	Kelembaban (%)	Ph Tanah (Keasaman)	Curah Hujan (mm/bulan)	Hasil Panen (Kg/hektar)
30	80	6,5	200	500
25	75	6,8	180	450
28	82	6,2	220	470
32	78	6,7	210	480
29	81	6,4	190	460
26	79	6,6	200	455
31	77	6,5	205	490
27	74	6,7	195	440
30	80	6,5	200	495
28	76	6,4	215	465
30	80	6,5	200	500

Berdasarkan Tabel 1, angka-angka tersebut merupakan nilai dari berbagai fitur dan target yang digunakan dalam penelitian untuk memprediksi hasil panen. Temperatur menjelaskan suhu rata-rata harian (dalam derajat Celsius). Kelembaban menjelaskan tentang kelembaban rata-rata harian (dalam persentase). Ph Tanah merupakan tingkat keasaman tanah. Curah Hujan menjelaskan tentang curah hujan bulanan yang terjadi

dalam satuan milimeter. Hasil Panen menjelaskan tentang hasil produksi pertanian yang diukur dalam satuan kilogram per hektar.

2.2. Algoritma *Linear Regression*

Algoritma *Linear Regression* merupakan metode statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen dengan variabel dependen menggunakan garis lurus. Model ini berguna dalam prediksi dan analisis data, termasuk dalam konteks pertanian untuk memprediksi hasil panen berdasarkan faktor lingkungan seperti temperatur, kelembaban, pH tanah, dan curah hujan.

Rumus-rumus (persamaan) pada algoritma *Linear Regression* antara lain:

1. Persamaan *Linear Regression* Sederhana

$$y = \beta_0 + \beta_1 x + \epsilon \quad (1)$$

Dimana:

- y = Variabel dependen (hasil yang diprediksi)
- x = Variabel independen (faktor yang mempengaruhi)
- β_0 = Intercept (titik potong dengan sumbu y)
- β_1 = Slope (kemiringan garis)
- ϵ = Error term (residu)

2. Persamaan *Linear Regression* Berganda

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (2)$$

Dimana:

- y = Variabel dependen
- x_1, x_2, \dots, x_n = Variabel independen
- $\beta_0, \beta_1, \dots, \beta_n$ = Koefisien Regresi
- ϵ = Error term (residu)

3. Persamaan Koefisien Regresi

$$\beta = (X^T X)^{-1} X^T y \quad (3)$$

Dimana:

- β = Vektor koefisien regresi
- X = Matriks variabel independen
- y = Vektor variabel dependen

4. Persamaan Evaluasi Model

- a. Mean Square Error (MSE)

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

Dimana:

- y_i = Nilai aktual
- \hat{y}_i = Nilai prediksi
- n = Jumlah Data

- b. R^2 Score

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

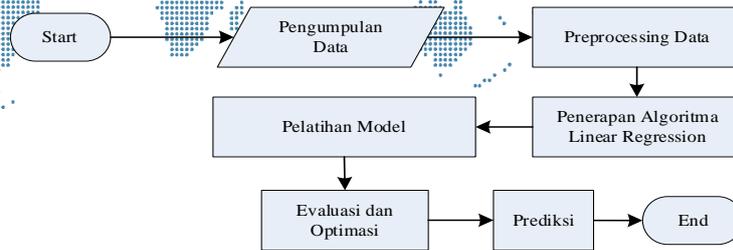
Dimana:

- \bar{y} = Rata-rata nilai aktual

2.3. Diagram Alur Penelitian

Diagram alur dalam penelitian ini berfungsi untuk memvisualisasikan langkah-langkah penelitian secara sistematis, memfasilitasi pemahaman dan komunikasi yang lebih efektif, mengidentifikasi dan mengelola tahapan kritis. Diagram ini juga membantu memantau kemajuan penelitian, mengurangi risiko kesalahan dan redundansi, dan mendokumentasikan metodologi penelitian secara rinci, dengan menyediakan kerangka

kerja yang jelas, diagram alur mendukung pengambilan keputusan yang lebih informasional dan memastikan penelitian berjalan efisien dan sesuai jadwal.



Gambar 1. Diagram Alur Penelitian

Berdasarkan Diagram Alur Penelitian yang disajikan pada Gambar 1, dapat dijelaskan tahapan-tahapan nya sebagai berikut:

1. Pengumpulan Data
Tahapan ini berupa pengumpulan data historis cuaca, kondisi tanah, dan hasil panen dari petani. Data tersebut kemudian disimpan dalam format CSV untuk kemudahan pemrosesan.
2. Preprocessing Data
Tahapan ini berarti membersihkan data dari *missing values* dan data yang tidak relevan.
3. Penerapan Algoritma Linear Regression
Tahapan ini berarti bahwa penelitian dianalisis dan diproses dengan menggunakan Algoritma Linear Regression.
4. Pelatihan Model
Tahapan pada bagian ini diantaranya adalah membagi data menjadi data latih dan data uji. Pelatihan model menggunakan data latih dan evaluasi performanya menggunakan data uji.
5. Evaluasi dan Optimasi
Tahapan ini berarti menghitung metrik evaluasi seperti Mean Squared Error (MSE) atau R^2 score. Tahap optimasi menggunakan *hyperparameter* model untuk meningkatkan akurasi prediksi.
6. Prediksi
Tahapan ini berarti bagaimana menggunakan model yang telah dilatih sebelumnya untuk memprediksi hasil panen berdasarkan data cuaca dan kondisi tanah yang baru.

3. Hasil dan Pembahasan

Seluruh proses analisis pada penelitian ini, untuk prediksi hasil pertanian dengan algoritma *linear regression* akan menggunakan bahasa pemrograman python. Sedangkan proses-proses tersebut akan dijalankan menggunakan *tools* Visual Studio Code.

3.1. Pengumpulan Data

Penelitian ini telah berhasil mengumpulkan data historis yang terdiri dari variabel cuaca, kondisi tanah, dan hasil panen selama sepuluh musim. Dataset yang digunakan disimpan dalam file *excel* (CSV) mencakup nilai-nilai temperatur, kelembaban, pH tanah, curah hujan, dan hasil panen (dapat dilihat pada Tabel 1 di pembahasan sebelumnya). Berdasarkan data tersebut, akan dibagi menjadi 2 (dua) variabel. Variabel x yang merupakan variabel independen (faktor yang mempengaruhi), terdiri dari: “Temperatur, Kelembaban, Ph Tanah, dan Curah Hujan”. Sedangkan “Hasil Panen” adalah variabel y yang merupakan variabel dependen (hasil yang diprediksi). Sebelum ke tahap *Preprocessing* akan dilakukan import *library* yang dibutuhkan dalam kode Python.

```
# Import library yang dibutuhkan
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Gambar 2. Kode program untuk import library

Kode program Python pada Gambar 2 berfungsi untuk mengimpor beberapa pustaka yang diperlukan untuk memproses data dan membangun model, *pandas* untuk manipulasi dan analisis data dalam bentuk *DataFrame*, *train_test_split* dari *sklearn.model_selection* untuk membagi dataset menjadi data latih dan uji. *StandardScaler* dari *sklearn.preprocessing* untuk menormalisasi fitur sehingga memiliki *mean* 0 dan standar deviasi 1. *LinearRegression* dari *sklearn.linear_model* untuk membangun model regresi linear, serta *mean_squared_error* dan *r2_score* dari *sklearn.metrics* untuk mengevaluasi kinerja model berdasarkan metrik *Mean Squared Error* dan koefisien determinasi *R² Score*.

```
# Langkah 1: Pengumpulan Data
data = pd.read_csv('data_pertanian.csv') # Data disimpan dalam file CSV
```

Gambar 3. Kode program untuk membaca data penelitian (CSV)

Kode program pada Gambar 3 berfungsi untuk membaca data penelitian yang telah disimpan dalam bentuk file *excel* (CSV).

3.2. Preprocessing Data

Tahapan ini merupakan persiapan data sebelum analisis atau penerapan pada model machine learning, yang meliputi pembersihan data dengan mengatasi nilai hilang, menghapus duplikasi, dan mengoreksi kesalahan, transformasi data dengan normalisasi, standarisasi, dan pengkodean fitur kategorikal, reduksi dimensi untuk mengurangi kompleksitas, pembagian data menjadi set latih, dan uji. Hal ini untuk memastikan data berkualitas tinggi dan formatnya sesuai, sehingga meningkatkan akurasi dan performa model.

```
# Langkah 2: Preprocessing Data
# Menghapus missing values
data = data.dropna()

# Memisahkan fitur dan target
X = data[['temperatur', 'kelembaban', 'ph_tanah', 'curah_hujan']]
y = data['hasil_panen']

# Normalisasi data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Gambar 4. Kode program tahap preprocessing data

Berdasarkan Gambar 4, kode yang diberikan meliputi tiga langkah utama dalam *preprocessing* data: pertama, menghapus semua baris dalam dataset yang memiliki nilai yang hilang untuk memastikan analisis lebih lanjut dilakukan pada data yang lengkap; kedua, memisahkan fitur (*x*) yang terdiri dari kolom “Temperatur, Kelembaban, Ph Tanah, dan Curah Hujan”, serta target (*y*) yang merupakan kolom “Hasil Panen”; dan ketiga, melakukan normalisasi data fitur menggunakan *StandardScaler* dari *scikit-learn* untuk memastikan semua fitur memiliki skala yang sama, yang penting untuk algoritma machine learning yang sensitif terhadap skala data, sehingga hasil normalisasi disimpan sebagai *x_scaled*.

3.3. Penerapan Algoritma Linear Regression

Tahapan ini merupakan penerapan algoritma *linear regression* yang berarti memilih algoritma tersebut sebagai metode untuk analisis data.

```
# Langkah 3: Pemilihan Algoritma
model = LinearRegression()
```

Gambar 5. Kode program pemilihan algoritma *linear regression*

Untuk langkah pemilihan algoritma, dimulai dengan menetapkan model yang akan digunakan dalam analisis data, dalam hal ini memilih model regresi linear dengan menginisialisasi objek dari kelas *LinearRegression* dan menyimpannya dalam variabel model.

3.4. Pelatihan Model

Pelatihan model bertujuan untuk menemukan hubungan linier antara fitur-fitur (variabel independen) dan target (variabel dependen) dalam dataset. Proses pelatihan ini melibatkan penyesuaian koefisien regresi (*slope* dan *intercept*) sehingga model dapat meminimalkan kesalahan prediksi pada data pelatihan. Setelah pelatihan, model diharapkan mampu memprediksi nilai target dengan akurasi yang baik ketika diberikan data baru yang belum pernah dilihat sebelumnya.

```
# Langkah 4: Pelatihan Model
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)
model.fit(X_train, y_train)
```

Gambar 6. Kode program pelatihan model *linear regression*

Kode program tersebut membagi dataset yang telah diskalakan (*X_scaled*) dan target (*y*) menjadi dua subset, yaitu set pelatihan dan set pengujian, dengan pembagian 80% untuk pelatihan dan 20% untuk pengujian menggunakan fungsi *train_test_split* dengan parameter *test_size=0.2* dan *random_state=42* untuk memastikan pembagian yang konsisten. Selanjutnya, model regresi linear yang telah dipilih sebelumnya dilatih menggunakan metode *fit* pada data pelatihan (*X_train* dan *y_train*), yang memungkinkan model untuk mempelajari hubungan antara fitur dan target dari data pelatihan tersebut.

3.5. Evaluasi dan Optimasi

Evaluasi dan optimasi pada *linear regression* bertujuan untuk memastikan model memiliki performa yang baik dan prediksi yang akurat. Evaluasi melibatkan penilaian model menggunakan metrik seperti *Mean Squared Error* (MSE) dan *R-squared* (R^2) untuk mengukur akurasi prediksi pada data pengujian. Optimasi bertujuan untuk meningkatkan performa model melalui metode seperti pemilihan fitur dan penyesuaian *hyperparameter*. Langkah-langkah ini memastikan model tidak *overfitting* dan mampu memberikan prediksi yang baik pada data baru.

```
# Langkah 5: Evaluasi dan Optimasi
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
print(f"R2 Score: {r2}")
```

Gambar 7. Kode program untuk evaluasi dan optimasi

Kode tersebut melakukan evaluasi terhadap model *linear regression* yang telah dilatih. Pertama, model membuat prediksi (*y_pred*) pada set pengujian (*X_test*). Kemudian, kesalahan prediksi diukur menggunakan *Mean Squared Error* (MSE) dengan

membandingkan nilai prediksi (y_{pred}) dengan nilai sebenarnya (y_{test}). Selain itu, R -squared (R^2) dihitung untuk menilai seberapa baik model menjelaskan variabilitas data target. Hasil MSE dan R^2 kemudian dicetak untuk memberikan gambaran tentang akurasi dan kualitas prediksi model. Hasilnya dapat dilihat pada Gambar 8 berikut:

```
Mean Squared Error: 329.5094120370046
R2 Score: 0.34911721079110203
```

Gambar 8. Hasil *Mean Squared Error* (MSE) dan R^2 Score

Hasil program menunjukkan bahwa model *linear regression* memiliki *Mean Squared Error* (MSE) sebesar 329.509, yang memberikan gambaran tentang tingkat kesalahan prediksi model pada set pengujian. R^2 Score sebesar 0.349 menunjukkan bahwa model dapat menjelaskan sekitar 34.9% dari variabilitas data target, yang berarti model memiliki kemampuan untuk menangkap sebagian hubungan antara fitur dan target dalam data, dengan nilai R^2 ini, model sudah menunjukkan bahwa terdapat pola yang dapat dikenali, dan ini bisa menjadi dasar yang baik untuk pengembangan dan perbaikan lebih lanjut.

3.6. Prediksi

Prediksi pada Linear Regression bertujuan untuk memperkirakan nilai variabel target berdasarkan nilai variabel-variabel fitur. Selain itu, prediksi ini juga membantu mengidentifikasi faktor-faktor penting yang mempengaruhi variabel target, sehingga dapat digunakan untuk perencanaan dan strategi lebih lanjut.

```
# Langkah 6: Prediksi
# Contoh prediksi dengan data baru
data_baru = [[30, 80, 6.5, 200]] # Contoh data baru
data_baru_scaled = scaler.transform(data_baru)
hasil_prediksi = model.predict(data_baru_scaled)
print(f"Hasil Prediksi Panen: {hasil_prediksi[0]}")
```

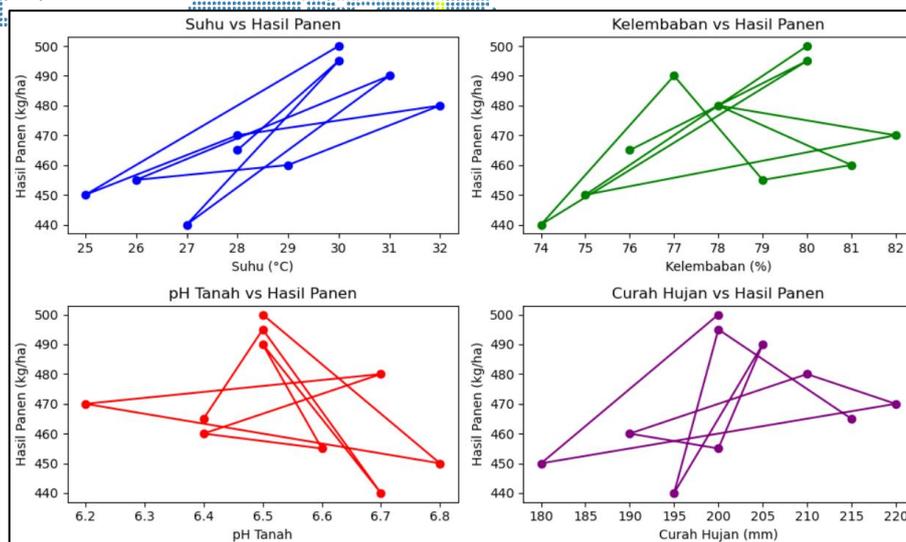
Gambar 9. Kode program untuk prediksi

Kode tersebut melakukan prediksi menggunakan model yang telah dilatih. Pertama, contoh data baru yang berisi beberapa fitur ($data_baru = [[30, 80, 6.5, 200]]$) disiapkan. Data ini kemudian diskalakan menggunakan objek *scaler* yang telah digunakan sebelumnya untuk menskalakan data pelatihan ($data_baru_scaled = scaler.transform(data_baru)$). Setelah itu, model melakukan prediksi pada data baru yang telah diskalakan ($hasil_prediksi = model.predict(data_baru_scaled)$). Hasil prediksi, yang menunjukkan perkiraan nilai panen, dicetak ke layar ($print(f"Hasil Prediksi Panen: {hasil_prediksi[0]}")$). Hasilnya dapat dilihat pada Gambar 10 berikut:

```
Hasil Prediksi Panen: 479.54239433004955
```

Gambar 10. Hasil prediksi panen

Berdasarkan Gambar 10, dapat dijelaskan bahwa prediksi panen berdasarkan analisis data yang telah dilakukan menghasilkan 479,5 Kg/ha. Variabel yang mempengaruhi hasil panen berdasarkan analisis dapat dilihat pada Gambar 11 berikut:



Gambar 11. Perbandingan hasil panen terhadap variabel yang digunakan

Berdasarkan analisis yang telah dilakukan, yang tersaji pada Gambar 11 dapat dijelaskan bahwa curah hujan dan kelembaban memiliki pengaruh signifikan terhadap hasil panen.

4. Kesimpulan

Linear Regression cukup efektif dalam memodelkan hubungan linier antara variabel cuaca, kondisi tanah, dan hasil panen. Koefisien regresi memberikan wawasan tentang seberapa besar pengaruh setiap variabel terhadap hasil panen. Salah satu temuan yang merupakan hal penting adalah bahwa curah hujan dan kelembaban memiliki pengaruh signifikan terhadap hasil panen, yang sesuai dengan pengetahuan agronomis bahwa air adalah faktor kunci dalam produksi tanaman. Meskipun MSE dan R^2 score memberikan indikasi bahwa model bekerja dengan baik, ada ruang untuk peningkatan lebih lanjut melalui optimasi *hyperparameter* dan mungkin dengan mencoba model machine learning lain yang lebih kompleks seperti *Random Forest* atau *Gradient Boosting*. Model *Ridge Regression* yang digunakan dalam optimasi menunjukkan peningkatan kecil dalam MSE dan R^2 score, menunjukkan bahwa regularisasi membantu dalam mengurangi overfitting. Secara garis besar penelitian ini menunjukkan bahwa algoritma *linear regression* dapat digunakan secara efektif untuk memprediksi hasil panen berdasarkan faktor lingkungan. Evaluasi model menunjukkan hasil yang memadai, dan ada potensi untuk lebih mengoptimalkan model melalui teknik-teknik tambahan. Hal ini dapat membantu petani dalam membuat keputusan yang lebih baik untuk meningkatkan produktivitas pertanian, khususnya tanaman jagung.

Daftar Pustaka

- [1] Y. Kamakaula, 'Optimasi Pertanian Berkelanjutan: Pengabdian Masyarakat Untuk Peningkatan Produktivitas Dan Kesejahteraan Petani Lokal', *Community Development Journal: Jurnal Pengabdian Masyarakat*, vol. 4, no. 6, pp. 11463–11471, 2023, doi: 10.31004/cdj.v4i6.22461.
- [2] P. C. Pandey and M. Pandey, 'Highlighting the role of agriculture and geospatial technology in food security and sustainable development goals', *Sustainable Development*, vol. 31, no. 5, pp. 3175–3195, 2023, doi: 10.1002/sd.2600.
- [3] R. Mrabet, *Sustainable Agriculture and the Environment*. Elsevier, 2023. doi: 10.1016/B978-0-323-90500-8.00013-0.
- [4] A. Sehgal *et al.*, 'Long-term crop rotation affects crop yield and economic returns

- in humid subtropical climate', *Field Crops Research*, vol. 298, no. 1, p. 108952, 2023, doi: 10.1016/j.fcr.2023.108952.
- [5] D. S. Silva, E. Y. Arima, T. N. P. dos Reis, and L. Rattis, 'Temperature effect on Brazilian soybean yields, and farmers' responses', *International Journal of Agricultural Sustainability*, vol. 21, no. 1, pp. 1–15, 2023, doi: 10.1080/14735903.2023.2173370.
- [6] Q. Wang, K. Huang, H. Liu, and Y. Yu, 'Factors affecting crop production water footprint: A review and meta-analysis', *Sustainable Production and Consumption*, vol. 36, no. 3, pp. 207–216, 2023, doi: 10.1016/j.spc.2023.01.008.
- [7] S. Zhang *et al.*, 'Effects of soil amendments on soil acidity and crop yields in acidic soils: A world-wide meta-analysis', *Journal of Environmental Management*, vol. 345, no. 1, p. 118531, 2023, doi: 10.1016/j.jenvman.2023.118531.
- [8] J. Fu *et al.*, 'Extreme rainfall reduces one-twelfth of China's rice yield over the last two decades', *nature food*, vol. 4, pp. 416–426, 2023, doi: 10.1038/s43016-023-00753-6.
- [9] E. E. Rezaei *et al.*, 'Climate change impacts on crop yields', *nature reviews earth & environment*, vol. 4, pp. 831–846, 2023, doi: 10.1038/s43017-023-00491-0.
- [10] A. B. Omotoso, S. Letsoalo, K. O. Olagunju, C. S. Tshwene, and A. O. Omotayo, 'Climate change and variability in sub-Saharan Africa: A systematic review of trends and impacts on agriculture', *Journal of Cleaner Production*, vol. 414, p. 137487, 2023, doi: 10.1016/j.jclepro.2023.137487.
- [11] B. Gwambene, E. Liwenga, and C. Mung'ong'o, 'Climate Change and Variability Impacts on Agricultural Production and Food Security for the Smallholder Farmers in Rungwe, Tanzania', *Environmental Management*, vol. 71, pp. 3–14, 2023, doi: 10.1007/s00267-022-01628-5.
- [12] J. Jiang and T. Zhou, 'Agricultural drought over water-scarce Central Asia aggravated by internal climate variability', *Nature Geoscience*, vol. 16, pp. 154–161, 2023, doi: 10.1038/s41561-022-01111-0.
- [13] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, 'Linear Regression. In: An Introduction to Statistical Learning', in *Springer Texts in Statistics ((STS))*, Springer, Cham, 2023, pp. 69–70. doi: 10.1007/978-3-031-38747-0_3.
- [14] P. Singh *et al.*, 'Development of performance-based models for green concrete using multiple linear regression and artificial neural network', *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 18, pp. 2945–2956, 2024, doi: 10.1007/s12008-023-01386-6.
- [15] S. Mohammadi, K. Rydgren, V. Bakkestuen, and M. A. K. Gillespie, 'Impacts of recent climate change on crop yield can depend on local conditions in climatically diverse regions of Norway', *Scientific Reports*, vol. 13, p. 3633, 2023, doi: 10.1038/s41598-023-30813-7.
- [16] B. Faye *et al.*, 'Climate change impacts on European arable crop yields: Sensitivity to assumptions about rotations and residue management', *European Journal of Agronomy*, vol. 142, p. 126670, 2023, doi: 10.1016/j.eja.2022.126670.
- [17] T. Hu *et al.*, 'Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield', *Agricultural and Forest Meteorology*, vol. 336, p. 109458, 2023, doi: 10.1016/j.agrformet.2023.109458.
- [18] N. Fiorentini, D. Pellegrini, and M. Losa, 'Overfitting Prevention in Accident Prediction Models: Bayesian Regularization of Artificial Neural Networks', *Transportation Research Record*, vol. 2677, no. 2, pp. 1455–1470, 2023, doi: 10.1177/03611981221111367.
- [19] S. Depaoli, S. D. Winter, and H. Liu, 'Under-Fitting and Over-Fitting: The Performance of Bayesian Model Selection and Fit Indices in SEM', *Structural*

- Equation Modeling: A Multidisciplinary Journal*, vol. 31, no. 4, pp. 604–625, 2023, doi: 10.1080/10705511.2023.2280952.
- [20] M. Firman, M. Assidiq, and S. Syarli, 'Sistem Prediksi Hasil Pertanian Durian Dengan Metode Regresi Linear Berganda', *Journal Pegguruang: Conference Series*, vol. 5, no. 1, p. 350, 2023, doi: 10.35329/jp.v5i1.3199.
- [21] A. Satria, R. M. Badri, and I. Safitri, 'Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning', *Digital Transformation Technology*, vol. 3, no. 2, pp. 389–398, 2023, doi: 10.47709/digitech.v3i2.2852.
- [22] R. Andia, K. Kaslani, S. E. Permana, and T. Handayani, 'Peramalan Hasil Panen Padi Kabupaten Cirebon Menggunakan Algoritma Regresi Linear Berganda', *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 738–747, 2024, doi: 10.36040/jati.v8i1.8446.
- [23] I. Hamidah and M. P. Matematika, 'Analisis Hasil Tanaman Perkebunan (Kopi Dan Teh) Menggunakan Regresi Linear', *ESTIMASI: Journal of Statistics and Its Application*, vol. 5, no. 2, pp. 184–195, 2024, doi: 10.20956/ejsa.v5i2.26989.