PREDIKSI HASIL PANEN WORTEL MENGGUNAKAN ALGORITMA REGRESI LINEAR BERGANDA

Nandita Afrilia S, Fathia Frazna Az-Zahra, Prajoko
Teknik Informatika, Universitas Muhammadiyah Sukabumi
Jl. R. Syamsudin, S.H. No. 50, Cikole, Kec. Cikole, Kota Sukabumi, Jawa Barat 43113
nandita045@ummi.ac.id

ABSTRAK

Selama beberapa dekade terakhir, pertanian hortikultura telah berkembang pesat dan berkembang menjadi pemasok utama berbagai macam sayuran. Wortel (*daucus carota l*) yang merupakan sumber utama vitamin A merupakan salah satu sayuran unggulan dalam bidang hortikultura. Kandungan vitamin dan mineral pada wortel menjadikannya pilihan yang sangat baik untuk dikonsumsi sehari-hari karena sangat penting untuk kesehatan tubuh. Budidaya wortel adalah fokus utama Maya Wortel, sebuah bisnis pertanian. Perusahaan ini didirikan di Sukabumi dengan potensi wortel yang besar untuk memenuhi permintaan konsumen. Pemanenan adalah salah satu langkah penting dalam produksi tanaman. Akibat ketidakstabilan produksi, produksi wortel Maya tidak optimal sehingga menimbulkan risiko yang mempengaruhi pasokan tanaman untuk memenuhi permintaan pasar. Jika panen tidak dilakukan sebaik mungkin, maka produksi tertinggi tidak akan berarti apa-apa. Memperkirakan hasil panen wortel memerlukan prediksi guna menjamin produksi optimal pada musim panen. Untuk meramalkan hasil panen, algoritma *machine learning* yang dikenal sebagai Regresi Linear Berganda diterapkan. Regresi Linear Berganda merupakan salah satu teknik untuk menghitung pengaruh variabel independen terhadap variabel dependen. Hasil evaluasi kinerja model dengan menggunakan MSE sebesar 0,00228 dan RMSE sebesar 0,04788 yang digunakan dalam penelitian ini untuk menghasilkan model prediksi hasil panen wortel menunjukkan bahwa model prediksi tersebut cukup akurat dalam memperkirakan hasil panen wortel.

Kata kunci : *Machine Learning, Prediksi, Regresi Linear Berganda, MSE & RMSE*

1. PENDAHULUAN

Mayoritas masyarakat Indonesia masih bermatapencaharian dari subsektor pertanian antara lain tanaman pangan, perkebunan, peternakan, perikanan, dan kehutanan. Indonesia hampir tidak mempunyai persaingan dalam hal potensi sebagai produsen produk hortikultura unggulan karena mempunyai dua musim yang berbeda. Tanaman sayuran dapat diperoleh dari sektor pertanian hortikultura yang tumbuh signifikan dalam beberapa tahun terakhir. Tanaman sayuran wortel (Daucus carota L) yang kaya vitamin A merupakan salah satu tanaman yang banyak diminati di bidang hortikultura [1].

Sebuah perusahaan bernama Maya Carrot bekerja di industri pertanian, yang mengkhususkan diri pada sayuran wortel. Dalam hal ini, petani dan pemilik usaha bekerja sama untuk memproduksinya. Salah satu tugas tanaman yang paling krusial adalah memanen, potensi produksi tertinggi pun akan sia-sia jika pemanenan tidak dilakukan dengan kemampuan terbaiknya. Upaya optimalisasi produksi perlu dikelola dengan baik jika hasil panen tidak sesuai dengan hasil produksi [2].

Akibat ketidakstabilan produksi, produksi wortel Maya tidak optimal sehingga menimbulkan risiko yang dapat mempengaruhi pasokan tanaman untuk memenuhi permintaan pasar. Pada tahun 2018, produksi wortel sebesar 144 kuintal, dan pada tahun 2019 meningkat sebesar 85 kuintal, dengan produksi wortel sebesar 235 kuintal. Pada tahun 2019 terjadi surplus sebesar 25 kuintal, sedangkan total permintaan pada tahun tersebut sebesar 210 kuintal. Hal ini menyebabkan instansi tersebut memberikan sebagian hasil panen wortel kepada masyarakat lokal dan bahkan harus mencari konsumen untuk menjual hasil panen wortel di pasar lain.

Oleh karena itu, untuk memastikan apakah produksi di masa depan akan naik atau turun, perkiraan panen wortel harus dibuat di instansi tersebut. Prediksi umumnya terdiri dari tiga langkah utama: menilai data historis, memilih metode yang sesuai, dan memproyeksikan data historis menggunakan metode yang dipilih [3].

Salah satu teknik analisis data dalam data mining yang sering digunakan untuk memprediksi suatu variabel dan menguji hubungan beberapa variabel adalah algoritma regresi linier berganda [4]. Ketika memprediksi nilai variabel yang sesuai dengan nilai data numerik yang akan digunakan, Algoritma Regresi Linier Berganda merupakan algoritma yang dapat memberikan hasil yang lebih tepat. Oleh karena itu, penerapan Algoritma Regresi Linier Berganda diharapkan dapat memberikan dasar yang kuat untuk meningkatkan produktivitas dan keberlanjutan sektor pertanian hortikultura.

2. TINJAUAN PUSTAKA

2.1. Wortel

Wortel (Daucus Carota L) merupakan tanaman sayuran perdu. Wortel mengandung nutrisi yang dibutuhkan oleh tubuh manusia terutama vitamin dan mineral [5].

Tanaman wortel umumnya tumbuh baik di daerah dataran tinggi pada ketinggian 1200m hingga 1500m. Tanaman wortel membutuhkan tanah yang ideal - subur, gembur, kaya humus, tidak basah, pH 6,0-6,8, suhu antara 22 hingga 24 derajat Celcius - untuk menghasilkan umbi berkualitas tinggi [6].

2.2. *Knowledge Discovery in Database* (KDD)

Istilah "*Knowledge Discovery in Database*" atau "*knowledge mining*" merupakan istilah lain yang digunakan untuk merujuk pada *data mining*, yaitu proses penambangan data dalam jumlah besar [7]. Proses KDD terdiri dari 5 tahap, yaitu sebagai berikut:

a. *Data Selection* Proses pengambilan informasi di KDD memerlukan pemilihan data dari kumpulan data operasi sebelum tahap dimulai [8] .
b. *Preprocessing* Operasi pembersihan harus dilakukan untuk menghilangkan data yang berlebihan, memverifikasi konsistensi data, dan memperbaiki kesalahan dalam data [8].
c. *Transformation* Proses pengkodean data tertentu agar sesuai untuk teknik data mining [8].
d. *Data Mining* Proses pencarian pola atau informasi yang menarik pada data terpilih dengan menggunakan teknik atau strategi khusus [8].
e. *Interpretation/Evaluation* Langkah ini menentukan apakah informasi atau pola yang ditemukan bertentangan dengan fakta atau teori yang diketahui [8].

2.3. Estimasi

Teknik menghitung suatu nilai populasi dari suatu nilai sampel dapat disebut dengan estimasi. Estimasi diperlukan untuk mengambil keputusan, merencanakan tugas, menghitung durasi, dan menjelaskan biaya [9]. Estimasi adalah proses yang digunakan untuk menghasilkan nilai tertentu untuk suatu parameter. Data yang digunakan untuk membuat estimasi Parameter adalah contoh yang digunakan estimator selama pengembangannya untuk menghasilkan nilai parameter. Estimasi parameter pertama kali digunakan untuk memperkirakan populasi dari suatu sampel. Estimasi merupakan tahapan terpenting dalam menentukan model peluang dari suatu dataset secara akurat [10].

2.4. Regresi Linear Berganda

Analisis regresi berganda dapat menguji hubungan antara variabel terikat dan bebas dengan menggunakan koefisien regresi parsial [3].

Berikut beberapa langkah yang perlu dilakukan saat menggunakan algoritma Regresi Linier Berganda:

a. Identifikasi variabel independen (x_1, x_2,…x_n) dan dependen (y)
b. Identifikasi nilai konstanta ($b0$) dan koefisien regresi ($b1$, $b2$, …, bn)
c. Substitusikan nilai-nilai tersebut ke dalam persamaan:

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (1)$$

2.5. *Machine Learning*

Machine Learning adalah cabang ilmu komputer dan kecerdasan buatan yang berfokus pada penggunaan data dan algoritme untuk meniru cara manusia belajar dan secara bertahap meningkatkan akurasi. Ketika algoritma *machine learning* yang lebih baik digunakan, keputusan yang diambil menjadi lebih baik. Pelatihan, pembelajaran, atau training ini merupakan salah satu ciri dari *machine learning*. Oleh karena itu, data yang perlu diperiksa berfungsi sebagai data pelatihan (*training set*) untuk *machine learning*. *Machine learning* dibagi menjadi tiga kategori: *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning* [7].

2.6. *Mean Square Error* & *Root Mean Square Error*

Mean Square Error (MSE) merupakan metode penghitungan kesalahan dan tidak digunakan untuk estimasi model melainkan untuk menghitung kesalahan pengambilan sampel data. Teknik yang biasa digunakan untuk menghitung error model dari prediksi data adalah *Root Mean Square Error* (RMSE) [7].

Adapun rumus meghitung MSE & RMSE sebagai berikut:

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

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

2.7. Google Colaboratory

Sebuah *platform* komputasi berbasis *cloud* gratis yang disebut Google Colaboratory menyediakan lingkungan *notebook* interaktif tanpa *server* yang memungkinkan pengguna membuat, berbagi, dan menjalankan kode. Jupyter Notebook, yang kompatibel dengan bahasa pemrograman populer seperti Python, R, dan Julia, adalah platform tempat Colab dijalankan. Colab menyediakan sumber daya komputasi canggih seperti mesin virtual jarak jauh yang menjalankan Linux hingga 12 jam dan dilengkapi dengan unit pemrosesan grafis (GPU) yang canggih. Oleh karena itu, ini adalah platform yang berguna untuk banyak aplikasi, seperti pembelajaran mesin dan tugas pembelajaran mendalam [11].

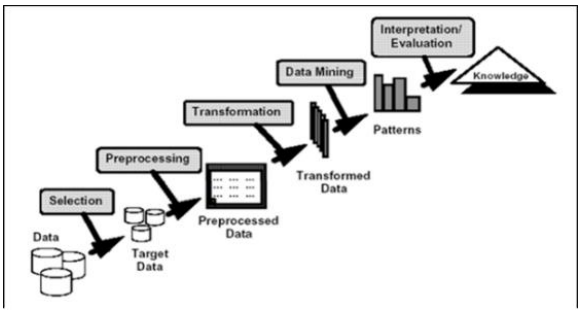
2.8. Flask

Flask berbasis Python merupakan *framework* yang memiliki *database* yang terintegrasi di dalamnya dan tidak memerlukan *tools* atau pustaka tambahan, ini tergolong dalam *micro framework*. Ekstensi tersedia untuk berbagai teknologi autentikasi terbuka, pemetaan relasional objek, validasi formulir, penanganan unggahan, dan alat untuk kerangka kerja populer. Ekstensi lebih sering diperbarui daripada program Flask itu sendiri [12].

3. METODE PENELITIAN

Metode penelitian yang digunakan peneliti untuk merencanakan penyelesaian penelitian adalah metode *Knowledge Discovering in Database*. Pada tahap ini

diusulkan rencana kegiatan komprehensif yang menguraikan semua langkah yang perlu diambil untuk melaksanakan penelitian.



Gambar 1. Tahapan Penelitian KDD

3.1. *Data Selection*

Pada tahap ini data produksi wortel dikumpulkan dengan cara memahami data tersebut untuk selanjutnya dibuat proses pemodelan prediktif dengan menggunakan metode yang dipilih yaitu Regresi Linier Berganda.

3.2. *Preprocessing*

Pada tahap ini data produksi wortel diperiksa, data duplikat dan data bernilai null dibersihkan, dengan tujuan untuk mengoptimalkan data masukan yang akan diolah, meminimalkan terjadinya kesalahan, dan agar hasil yang dihitung dengan algoritma memberikan hasil yang wajar.

3.3. *Transformation*

Tahap ini dilakukan dengan mengubah format data produksi wortel ke bentuk yang diperlukan sesuai kebutuhan dan algoritma yang digunakan. Pada fase transformasi, kumpulan data yang dipilih diubah, seperti dengan memilih atribut. Fase ini penting karena berperan penting dalam memastikan kualitas hasil selama proses data mining.

3.4. *Data Mining*

Pada fase ini dilakukan proses membangun model prediktif menggunakan metode regresi linier multivariat, yang bertujuan untuk mengekstrak pola atau wawasan berharga dari data. Proses pembuatan model pada tahap data mining menggunakan Google Colaboratory.

3.5. *Interpretation/Evaluation*

Pada tahap ini dilakukan evaluasi untuk mengetahui apakah nilai akurasi yang dihasilkan konsisten, relevan, dan sejalan dengan tujuan pembuatan model prediksi menggunakan metode yang dipilih. Tahap ini sangat penting untuk penyampaian prediksi yang menentukan besarnya kinerja. Proses ini melibatkan penentuan apakah informasi atau pola yang ditemukan bertentangan dengan fakta atau prinsip yang diketahui sebelumnya. Pendekatan yang digunakan adalah metode *mean square error* (MSE) dan *root mean square error* (RMSE).

3.6. *Deployment*

Pada titik ini, peneliti mengembangkan antarmuka akhir berbasis website yang dapat memperkirakan hasil panen wortel di Maya Wortel. Proses pembuatan sistem berbasis website ini menggunakan Visual Studio Code sebagai editor teks dan Flask untuk memasukkan temuan model ke dalam sistem.

4. HASIL DAN PEMBAHASAN

4.1. *Data Selection*

Pada tahap *data selection*, kumpulan data awal produksi panen wortel di Maya Wortel memiliki total 108 data historis. Adapun atribut-atribut pada data yaitu sebagai berikut:

	tahun	musim_tanam	curah_hujan	total_per_tahun	luas_lahan_tanam	luas_lahan_panen	hasil_panen
0	2015	1 s/d 3	331	NaN	2500	2115	10.52
1	2015	2 s/d 4	388	NaN	2704	3136	15.63
2	2015	3 s/d 5	433	NaN	4356	4096	20.39
3	2015	4 s/d 6	435	185.85	2400	2500	12.45
4	2015	5 s/d 7	173	NaN	3480	3720	18.53

Gambar 2. Membaca Atribut Dataset

Tabel 1. Atribut Dataset

NO	Atribut
1	Tahun
2	Musim Tanam
3	Curah Hujan
4	Total per Tahun
5	Luas Laha Tanam
6	Luas Lahan Panen
7	Hasil Panen
8	Pemupukan Buka Lahan
9	Pemupukan Dasar
10	Pemupukan Dasar Cair
11	Pemupukan Bulan Pertama
12	Pemupukan Bulan Kedua
13	Jumlah Petani

4.2. *Preprocessing*

Pemrosesan awal data adalah komponen kedua dari proses KDD, yang melibatkan pembersihan data untuk meningkatkan kualitasnya. Pada gambar 3, terlihat bahwa beberapa atribut berisi data numerik (int64 dan float64), sedangkan yang lain berisi data kategori (*object*).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 108 entries, 0 to 107
Data columns (total 13 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   tahun                   108 non-null    int64
 1   musim_tanam             108 non-null    object
 2   curah_hujan             108 non-null    int64
 3   total_per_tahun         9 non-null      float64
 4   luas_lahan_tanam        108 non-null    int64
 5   luas_lahan_panen        108 non-null    int64
 6   hasil_panen             108 non-null    float64
 7   pemupukan_buka_lahan    108 non-null    object
 8   pemupukan_dasar         108 non-null    object
 9   pemupukan_dasar_cair    108 non-null    object
 10  pemupukan_bulan_pertama 108 non-null    object
 11  pemupukan_bulan_kedua   108 non-null    object
 12  jumlah_petani           108 non-null    int64
dtypes: float64(2), int64(5), object(6)
memory usage: 11.1+ KB
```

Gambar 3. Mencari Informasi Dataset

Dalam mencari nilai duplikat, nilai yang dihasilkan adalah 0, yang berarti tidak ada baris duplikat dalam data tersebut seperti pada gambar 4.

```
0
```

Gambar 4. Mencari Nilai Duplikat

Terdapat 99 *missing value* pada total_per_tahun, hal ini terjadi karena data pada total_per_tahun merupakan atribut tambahan yang berisi jumlah hasil panen untuk setiap tahunnya.

```
tahun                    0
musim_tanam              0
curah_hujan              0
total_per_tahun         99
luas_lahan_tanam         0
luas_lahan_panen         0
hasil_panen              0
pemupukan_buka_lahan     0
pemupukan_dasar          0
pemupukan_dasar_cair     0
pemupukan_bulan_pertama  0
pemupukan_bulan_kedua    0
jumlah_petani            0
dtype: int64
```

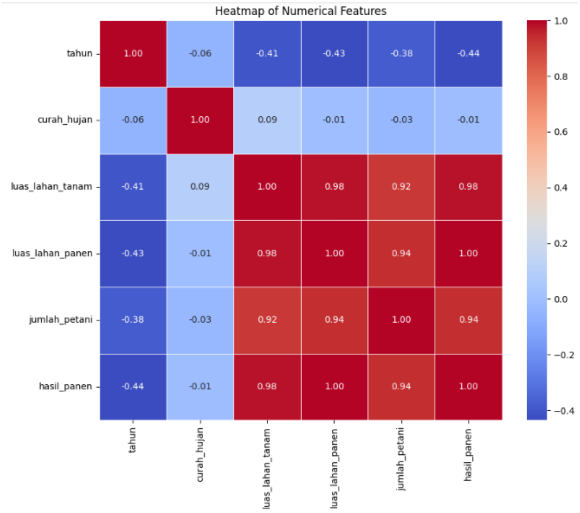
Gambar 5. Mencari Nilai Missing Value

Setelah itu, fitur numerik yang relevan dipilih untuk analisis lebih lanjut, seperti tahun, curah hujan, luas lahan tanam, luas lahan panen, jumlah petani, dan hasil panen. Adapun atribut yang akan digunakan yaitu sebagai berikut:

Tabel 2. Detail Penggunaan Atribut

Atribut	Detail Penggunaan
Tahun	×
Musim Tanam	×
Curah Hujan	√
Total Per Tahun	×
Luas Lahan Tanam	√
Luas Lahan Panen	√
Hasil Panen	√
Pemupukan Buka Lahan	×
Pemupukan Dasar	×
Pemupukan Dasar Cair	×
Pemupukan Bulan Pertama	×
Pemupukan Bulan Kedua	×
Jumlah Petani	×

Untuk memahami hubungan antara atribut-atribut numerik tersebut, matriks korelasi dihitung dan ditampilkan menggunakan seaborn dan matplotlib.



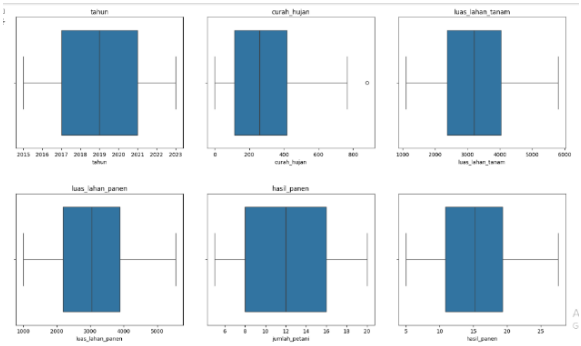
Gambar 6. Visualisasi Heatmap Nilai Matriks Korelasi

Berikut penjelasan koefisien korelasi Pearson dalam bentuk tabel:

Tabel 3. Koefisien Korelasi Pearson

Nilai	Korelasi
0	Tidak ada korelasi linier.
0.0 hingga 0.3 (atau -0.0 hingga -0.3)	Korelasi lemah.
0.3 hingga 0.7 (atau -0.3 hingga -0.7)	Korelasi sedang.
0.7 hingga 1.0 (atau -0.7 hingga -1.0)	Korelasi kuat.

Untuk mengetahui *outlier*, kemudian dibuat visualisasi boxplot untuk setiap atribut numerik. Hasilnya terdapat 1 *outlier* pada curah hujan.



Gambar 7. Visualisasi Boxplot Mencari *Outlier*

Selanjutnya dilakukan metode *Interquartile Range* (IQR) yang digunakan untuk menghilangkan *outlier* yang terdeteksi, memastikan bahwa data yang akan dianalisis bersih dan tidak terpengaruh oleh nilai ekstrim. Kode berikut menunjukkan cara menghilangkan outlier

```
(107, 6)
```

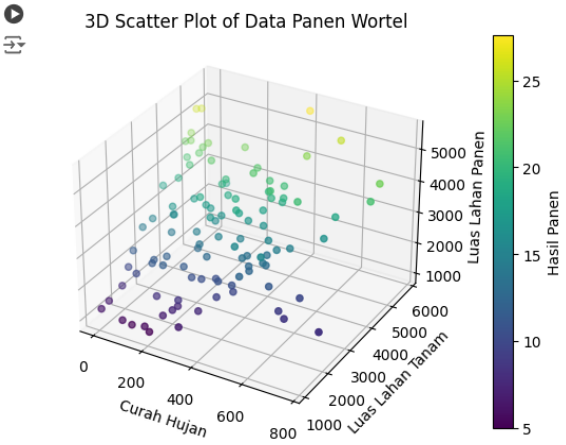
Gambar 8. Menghilangkan Outlier

4.3. *Transformation*

Fase ketiga dari proses KDD disebut transformasi data, tahap ini melibatkan konversi data yang telah diproses sebelumnya ke dalam format yang dapat digunakan untuk data mining. Selanjutnya untuk membangun model *machine learning* memerlukan pemisahan data menjadi fitur (x) dan target (y).

```
data_digunakan =
numeric_data[['curah_hujan',
'luas_lahan_tanam', 'luas_lahan_panen',
'hasil_panen']]
```

Sebelum menentukan fitur (x) dan target (y) pada variabel numerik, scatter plot 3D dibuat dan ditampilkan untuk memvisualisasikan hubungan antara curah_hujan, luas_lahan_tanam, dan luas_lahan_panen, dengan hasil_panen yang direprentasikan oleh warna titik-titik pada plot.



Gambar 9. Scatter Plot 3D

Setelah itu, data tersebut dipisahkan menjadi dua bagian: fitur (X) dan target (y). Fitur (x) mencakup curah_hujan, luas_lahan_tanam, dan luas_lahan_panen, sementara target (y) yaitu hasil_panen.

```
X =
data_digunakan.drop(columns=['hasil_pane
n'])
y = data_digunakan['hasil_panen']
```

Untuk menjamin konsistensi skala data, data tersebut dinormalisasi menggunakan *scaler* yang telah dilatih sebelumnya.

	curah_hujan	luas_lahan_tanam	luas_lahan_panen
0	0.247691	-0.636264	-0.878743
1	0.542282	-0.446716	0.068984
2	0.774855	1.088251	0.960089
3	0.785191	-0.729179	-0.521373
4	-0.568897	0.274310	0.611073
...	...	...	...
102	-1.463008	-1.937083	-1.635254
103	-1.463008	-1.658336	-1.588842
104	-1.013368	-1.937083	-1.774489
105	-0.868656	-1.844167	-1.913724
106	-1.158080	-1.147300	-0.985490

107 rows × 3 columns

Gambar 10. Standarisasi Data

4.4. *Data Mining*

Langkah keempat dari proses KDD adalah penambangan data, yang melibatkan penggunaan algoritma untuk menemukan pola atau koneksi dalam data. Dengan menggunakan fungsi train_test_split scikit-learn, data dibagi menjadi set pelatihan dan pengujian, dengan 80% data digunakan untuk pelatihan dan 20% sisanya untuk pengujian.

```
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=42)
```

a. train_test_split adalah fungsi dari *library*
b. sklearn.model_selection yang digunakan untuk membagi dataset menjadi dua subset: satu untuk pelatihan (X_train, y_train) dan satu untuk pengujian (X_test, y_test).
c. X dan y adalah data fitur dan label target, masing-masing.
d. test_size=0.2 berarti 20% dari data akan digunakan untuk pengujian, dan sisanya (80%) untuk pelatihan.
e. random_state=42 memastikan bahwa pembagian data bersifat reproducible, artinya jika Anda menjalankan kode ini lagi dengan random_state yang sama, hasil pembagian data akan tetap sama.

Selanjutnya, model regresi linier berganda dibuat dan dilatih.

```
model = LinearRegression()
model.fit(X_train, y_train)

with open('linear_regression_model.pkl',
'wb') as f:
    pickle.dump(model, f)


y_pred = model.predict(X_test)
```

a. 'LinearRegression' adalah kelas dari pustaka sklearn.linear_model yang digunakan untuk membuat model regresi linear.
b. model.fit(X_train, y_train) melatih model menggunakan data pelatihan (X_train dan y_train). Model belajar dari data pelatihan untuk membuat prediksi.
c. pickle adalah pustaka Python untuk serialisasi objek, memungkinkan Anda menyimpan model yang telah dilatih ke dalam file sehingga dapat dimuat kembali di lain waktu tanpa perlu melatih ulang.
d. 'linear_regression_model.pkl' adalah nama file di mana model akan disimpan.
e. wb adalah mode untuk membuka file dalam bentuk binary write.
f. pickle.dump(model, f) menyimpan model yang telah dilatih ke dalam file.
g. model.predict(X_test) menggunakan model yang telah dilatih untuk memprediksi nilai target (y_pred) berdasarkan data fitur pengujian (X_test).

4.5. *Interpretation/Evaluation*

Langkah kelima dalam proses KDD adalah evaluasi pola, yang dilakukan untuk memastikan legitimasi dan signifikansinya.

```
rmse =
np.sqrt(mean_squared_error(y_test,
y_pred))
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Root Mean Squared Error:
{rmse}")
print(f"Mean Squared Error (MSE):
{mse}")
print(f"R² Score: {r2}")

print("Koefisien: ", model.coef_)
print("Intercept: ", model.intercept_)
```

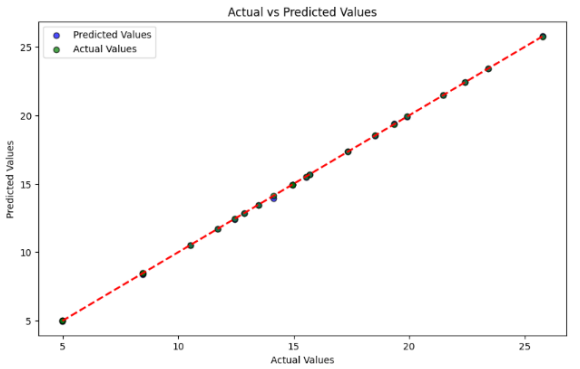
```
Root Mean Squared Error: 0.04778854616283905
Mean Squared Error (MSE): 0.0022837451443577985
R² Score: 0.9999257451790042
Koefisien:  [-0.02678665  0.0147414   5.35312837]
Intercept:  15.223171714925416
```

Gambar 11. Menguji Performa Model

Penjelasan mengenai hasil dari evaluasi pola adalah sebagai berikut:

a. Dalam konteks ini, RMSE sebesar 0.04778 menunjukkan bahwa model memiliki kesalahan prediksi yang relatif kecil.
b. Nilai MSE sebesar 0.00228 menunjukkan bahwa model memiliki kesalahan kuadrat rata-rata yang sangat kecil.
c. Nilai R^2 sebesar 0.9999 menunjukkan bahwa model hampir sepenuhnya menjelaskan variasi dalam data.
d. Koefisien menunjukkan dampak relatif dari setiap fitur pada prediksi model. Fitur pertama memiliki koefisien -0.02678, fitur kedua 0.01474, dan seterusnya.
e. Intercept sebesar 15.22317 menunjukkan bahwa, ketika semua fitur bernilai nol, model memprediksi nilai 15.22317.

Tahapan selanjutnya adalah presentasi pengetahuan, dimana pengetahuan yang telah ditemukan dikomunikasikan. Untuk membandingkan nilai aktual dengan prediksi, dengan cara membuat plot sebar yang memberikan gambaran jelas tentang seberapa baik model memprediksi hasil panen.



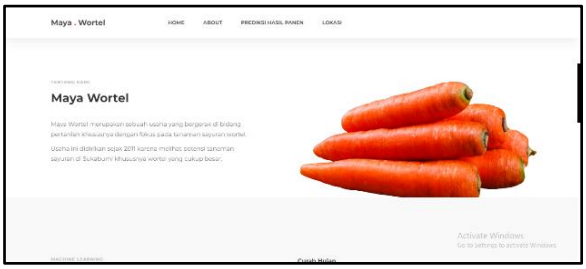
Gambar 12. Grafik Interpretasi Model

4.6. *Deployment*

Pada saat penerapan, dibuatlah model pembelajaran mesin yang diimplementasikan pada website yang dapat memperkirakan hasil panen wortel. Halaman utama ini menampilkan penerapan model prediksi hasil panen yang terintegrasi di situs web "Maya Wortel".

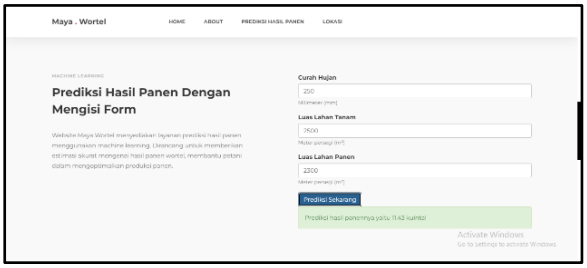
Gambar 13. Tampilan *Home*

Halaman "*About*" ini memberikan deskripsi singkat tentang perusahaan Maya Wortel, yang berfokus pada pertanian wortel dan sejarah pendiriannya.



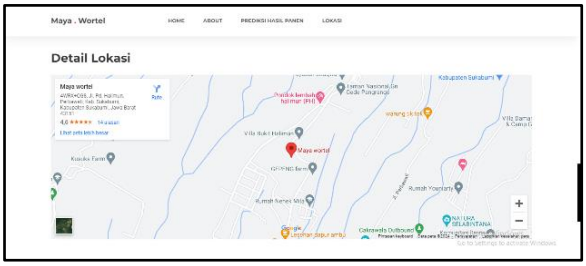
Gambar 14. Tampilan *About*

Dengan bantuan halaman "Prediksi Panen", petani dapat dengan mudah memperkirakan hasil panen wortel berdasarkan input tertentu seperti luas tanam, luas panen, dan curah hujan. Dengan penggunaan teknologi *machine learning*, fitur ini bertujuan untuk membantu petani dalam mengambil keputusan yang lebih tepat mengenai produksi hasil panen dan pengelolaan lahan dengan menawarkan perkiraan yang tepat.



Gambar 15. Tampilan Prediksi Hasil Panen

Peta interaktif pada halaman "Lokasi" ini memudahkan pengguna menemukan lokasi Maya Wortel. Pengguna dapat melihat rute dan menerima petunjuk arah secara langsung melalui fitur integrasi Google Maps.



Gambaran 16. Tampilan Lokasi

4.7. Pengujian Fungsionalitas Sistem

Setelah tahap *deployment* dilakukan, maka diperlukan pengujian fungsionalitas sistem terhadap fitur yang disediakan pada tampilan *website* yang dihasilkan.

Tabel 4. Pengujian Fungsionalitas Sistem

NO	Tampilan	*Input*	*Output*	Hasil Pengujian
1	Halaman *Home*	Pengguna menekan tombol "*Home*"	Tampil halaman "Home"	SUKSES
2	Halaman *About*	Pengguna menekan tombol "*About*"	Tampil halaman "*About*"	SUKSES
3	Halaman Prediksi Hasil Panen	Pengguna menekan tombol "Prediksi Hasil Panen"	Tampil halaman "Prediksi Hasil Panen"	SUKSES
4	Form Prediksi Hasil Panen	Pengguna mengisi form, kemudian menekan tombol "Prediksi Sekarang"	Tampil hasil Prediksi Hasil Panen	SUKSES
5	Halaman Lokasi	Pengguna menekan tombol "Lokasi"	Tampil halaman "Lokasi"	SUKSES

5. KESIMPULAN DAN SARAN

Penelitian ini menghasilkan model prediksi menggunakan algoritma Regresi Linear Berganda dengan nilai MSE sebesar 0.00228, yang menunjukkan bahwa model memiliki kesalahan kuadrat rata-rata yang sangat kecil. Serta nilai RMSE sebesar 0.04778, yang menunjukkan bahwa perbedaan antara hasil panen yang diprediksi oleh model dan hasil panen aktual relatif kecil. Penelitian ini juga menghasilkan model prediksi yang telah diintegrasikan ke dalam tampilan website. Adapun saran untuk penelitian selanjutnya yaitu sertakan faktor-faktor tambahan seperti jenis tanah, metode irigasi, dan penggunaan pupuk yang mungkin berdampak pada hasil panen.

DAFTAR PUSTAKA

[1] I. Komang *et al.*, "Pengembangan Sentra Pertanian Tomat Dengan Sistem Polikultur Hortikultura Berteknologi Digital Di Desa Pinggan, Kintamani," hal. 997–1003, 2022.

[2] A. Lili, H. Cipta, dan S. Widodo, "Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok Menggunakan Algoritma K-Means," *J. Mach. Learn. Data Anal.*, vol. 01, no. 01, hal. 45–54, 2022.

[3] E. Triyanto, H. Sismoro, dan A. D. Laksito, "IMPLEMENTASI ALGORITMA REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI PRODUKSI PADI DI KABUPATEN BANTUL," vol. 4, no. 2, hal. 73–86, 2019.

[4] I. B. K. D. S. Negara, I. P. K. Negara, dan N. Y. Arso, "PREDIKSI HASIL PANEN PADI DI KABUPATEN JEMBRANA MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," hal. 260–265, 2023.

[5] J. Bakoil, "Effect of Carrot Flour Addition on Biscuit Quality," *Sci. J. Educ. Forum*, vol. 8, no. 5, hal. 305–309, 2022, doi: 10.5281/zenodo.6474260.

[6] M. Idris, N. Khoiriyah, dan A. D. S. Syathori, "Pendapatan usahatani wortel di Desa Ngabab Kecamatan Pujon Kabupaten Malang," *J. Sos. Ekon. dan Pertan.*, vol. 9, no. 1, hal. 1–9, 2021.

[7] K. Puteri dan A. Silvanie, "Machine Learning untuk Model Prediksi Harga Sembako," *J. Nas. Inform.*, vol. 1, no. 2, hal. 82–94, 2020.

[8] S. Z. Harahap dan A. Nastuti, "Teknik Data Mining Untuk Penentuan Paket Hemat Sembako," *J. Ilm. Fak. Sains dan Teknol.*, vol. 7, no. 3, hal. 111–119, 2019.

[9] A. Z. Siregar, "Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 3, hal. 133–137, 2021, [Daring]. Tersedia pada: https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/73

[10] M. Berek, N. Rizky, dan R. Saputra, "ANALISIS PERBANDINGAN MODEL REGRESI LINIER DAN SUPPORT VECTOR MACHINE UNTUK ESTIMASI KINERJA CPU," vol. 11, no. 1, hal. 1–5, 2024.

[11] E. Supriyadi *et al.*, "Pendampingan Guru Matematika Di Cianjur Dalam Penggunaan Bahasa Pemrograman Python Dengan Google Collaboratory," vol. 3, no. 2, hal. 39–44, 2023.

[12] P. Banerjee, B. Kumar, A. Singh, R. Kumar, dan R. Kumar, "Comparative performance analysis of optimized round robin scheduling(ORR) using dynamic time quantum with round robin scheduling using static time quantum in Real Time System," *Int. J. Eng. Comput. Sci.*, vol. 8, no. 12, hal. 24890–24893, 2019, doi: 10.18535/ijecs/v8i12.4399.