PENERAPAN REGRESI LINIER BERGANDA UNTUK MEMPREDIKSI HASIL PANEN KACANG KEDELAI (STUDI KASUS: KECAMATAN SURADE)

M. Fikri Ihsani Raehan, Asep Budiman Kusdinar, Didik Indrayana
Teknik Informatika, Universitas Muhammadiyah Sukabumi
Jl. R. Syamsudin, S.H. No. 50, Cikole, Kec. Cikole, Kota Sukabumi, Jawa Barat 43113
fikri5344@ummi.ac.id

ABSTRAK

Pertanian memainkan peran penting dalam pembangunan nasional Indonesia sebagai penyerap tenaga kerja, penyumbang PDB, dan pendorong pertumbuhan ekonomi. Produksi hasil pertanian, seperti kacang kedelai di Kecamatan Surade, Kabupaten Sukabumi, sangat penting untuk kebutuhan pangan. Namun, pendataan hasil panen kacang kedelai masih kurang optimal karena minimnya teknologi, dan prediksi hasil panen belum tersedia, mempengaruhi akurasi prediksi produksi tanaman pangan. Penelitian ini bertujuan mengembangkan model prediksi hasil panen kacang kedelai menggunakan regresi linier berganda. Dengan data dari balai penyuluhan pertanian Kecamatan Surade, penelitian ini mengintegrasikan algoritma regresi linier berganda dalam sebuah website untuk meningkatkan akurasi prediksi hasil panen. Website ini dirancang untuk memberikan pengalaman optimal bagi pengguna dengan efisiensi implementasi algoritma yang baik. Nilai Mean Square Error (MSE) model ini adalah 3,74412, menunjukkan rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Nilai Root Mean Square Error (RMSE) adalah 1,93755, menandakan tingkat kesalahan prediksi yang rendah dan akurasi model yang baik. Hasil penelitian ini diharapkan meningkatkan akurasi prediksi, efisiensi pengelolaan pertanian, serta ketahanan pangan di Kecamatan Surade. Dokumentasi proses pengembangan memudahkan replikasi di wilayah lain.

Kata kunci **:** Prediksi hasil panen, Regresi linier berganda, *Machine learning*, Kacang kedelai, *Website*

1. PENDAHULUAN

Pertanian berperan penting dalam pembangunan nasional sebagai penyerap tenaga kerja, penyumbang produk domestik bruto (PDB), dan mesin pertumbuhan sektor ekonomi lainnya [1].

Sektor pertanian terdiri dari subsektor seperti produksi pangan, perkebunan, pertanian, perikanan, dan kehutanan. Meski begitu kebutuhan pangan nasional terus meningkat, Indonesia berpotensi menjadi produsen produk hortikultura terbaik karena negara ini memiliki dua musim. Hortikultura, sebagai sumber penting tanaman pangan, telah berkembang pesat dalam beberapa dekade terakhir [2].

Kedelai atau dikenal juga dengan sebutan tanaman kedelai, populer di kalangan masyarakat Indonesia dan merupakan salah satu tanaman pangan yang banyak digunakan dalam bidang hortikultura. Tanaman kedelai juga dapat digolongkan sebagai tanaman kecil atau tanaman tahunan. Di Indonesia, kedelai merupakan salah satu bahan pangan penting yang kaya akan protein. Seiring dengan meningkatnya kebutuhan bahan baku kedelai, maka permintaan terhadap kedelai juga semakin meningkat [3].

Kabupaten Surade terletak di Provinsi Sukabumi dan mempunyai produktivitas pertanian yang tinggi khususnya hortikultura termasuk tanaman pangan seperti kedelai, karena mempunyai lahan pertanian yang luas dan kondisi iklim yang ideal untuk pertumbuhan tanaman. Saat menanam kedelai, penting untuk diingat bahwa banyak faktor, termasuk jenis tanaman itu sendiri, kondisi lingkungan di mana tanaman itu tumbuh, dan teknik budidaya yang digunakan, dapat sangat mempengaruhi hasil panen.

Meskipun hasil panen kedelai di kecamatan ini berfluktuasi dari tahun ke tahun, namun pendataan panen masih kurang optimal karena minimnya penggunaan teknologi [4].

Keterbatasan dalam pengumpulan data ini dapat mempengaruhi keakuratan perkiraan produksi pangan di masa depan. Oleh karena itu, perbaikan sistem pendataan dan peningkatan aksesibilitas data sangat penting untuk mengoptimalkan prakiraan hasil kedelai sebagai tolok ukur penilaian ketahanan pangan kedelai di wilayah Surade [5].

Dalam analisis data, penggunaan algoritma merupakan langkah penting dalam membuat prediksi tentang fenomena. Algoritma yang biasa digunakan untuk melakukan prediksi adalah algoritma regresi linier berganda. Algoritma ini merupakan metode analisis data yang digunakan untuk memprediksi variabel dan menguji hubungan beberapa variabel yang ada.

Memprediksi hasil panen kedelai sangat penting untuk memastikan keberlanjutan produksi pangan dan memenuhi kebutuhan masyarakat di masa mendatang. Prediksi yang akurat dapat membantu para petani dan pengambil kebijakan untuk merencanakan strategi yang tepat dalam menghadapi fluktuasi hasil panen, mengelola sumber daya, dan meningkatkan produktivitas. Selain itu, prediksi hasil panen juga berperan dalam mengoptimalkan distribusi pasokan pangan sehingga dapat mencegah kelangkaan atau

surplus yang berlebihan. Dengan demikian, prediksi hasil panen menjadi langkah kunci dalam memastikan ketahanan pangan kedelai di wilayah ini.

Oleh karena itu, dengan menggunakan algoritma tersebut, penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang mempengaruhi hasil panen kedelai. Hasil penelitian ini juga dapat memberikan panduan berharga bagi lembaga yang ingin mencapai hasil panen terbaik.

2. TINJAUAN PUSTAKA

2.1. Machine Learning

Machine Learning adalah cabang kecerdasan buatan (AI) yang menggunakan prinsip-prinsip ilmu komputer dan statistik untuk membuat model yang mencerminkan pola dalam data. Model tersebut diajarkan melalui berbagai algoritma dalam metode pembelajaran mesin sehingga dapat mengklasifikasikan entitas bencana [6].

2.2. Prediksi

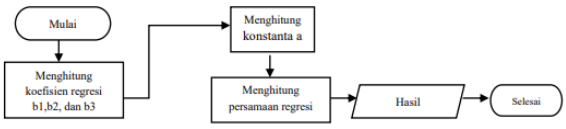
Prediksi adalah proses sistematis memperkirakan kemungkinan kejadian di masa depan berdasarkan informasi masa lalu dan masa kini. Tujuannya adalah untuk meminimalkan kesalahan, yaitu perbedaan antara hasil prediksi dan kenyataan. Penting untuk dicatat bahwa peramalan tidak selalu memberikan jawaban pasti tentang kejadian di masa depan, namun lebih merupakan upaya untuk memberikan perkiraan seakurat mungkin [4].

2.3. Regresi Linear Berganda

Regresi Linear Berganda merupakan metode statistik yang disebut regresi, digunakan untuk memprediksi hubungan antara suatu variabel terikat dan satu atau lebih variabel bebas [7].

Dalam analisis regresi sederhana, hubungan antar variabel bersifat linier, dan perubahan variabel X selalu mengakibatkan perubahan variabel Y [7].
Langkah-langkahnya adalah sebagai berikut:

a. Menentukan variabel independen dan dependen.
b. Menentukan nilai konstanta dan koefisien regresi.
c. Substitusikan nilai-nilai tersebut ke dalam persamaan.



Gambar 1. Flowcart Regresi Linear Berganda

2.4. Kedelai

Kedelai (Glycine max (L.)) merupakan tanaman pangan yang tumbuh tegak dari jenis liar Glycine ururiencis. Kedelai tumbuh jauh di atas 500 m di atas permukaan laut, terutama di daerah tropis dan 10573edelai10573e. Kedelai lebih tahan terhadap kondisi lingkungan dibandingkan jagung [8].

2.5. MSE & RMSE

MSE (mean squared error) adalah rata-rata selisih kuadrat antara nilai prediksi dan nilai observasi. Meskipun metode MSE terkadang menghasilkan perbedaan yang lebih besar, namun metode ini umumnya menghasilkan kesalahan yang lebih baik untuk kesalahan yang kecil [9].

Rumus berikut dapat digunakan untuk menentukan nilai MSE:

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

Root mean square error (RMSE) adalah metrik yang digunakan untuk mengevaluasi keakuratan prediksi model. Rata-rata kuadrat deviasi antara nilai 10573edela suatu variabel dan nilai prediksi disebut root mean square error atau RMSE. Semakin rendah nilai RMSE maka semakin dapat diandalkan hasil prediksinya [10].

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

2.6. Google Colaboratory

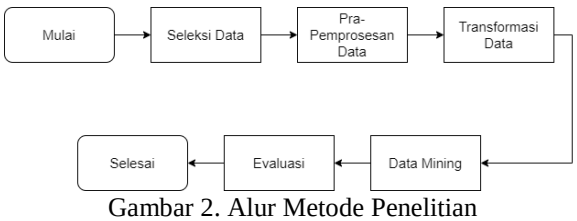
Google Colab merupakan platform pengembangan berbasis cloud yang memungkinkan pengguna melakukan pemrograman dan analisis data menggunakan Python [11].

2.7. Flask

Flask merupakan web framework yang ditulis dengan Python dan ditempatkan pada kategori microframeworks karena tidak memerlukan tools atau pustaka tertentu serta memiliki database yang terintegrasi tanpa memerlukan abstraksi lapisan database [12].

3. METODE PENELITIAN

Dengan menerapkan pendekatan KDD yang merupakan salah satu metode yang digunakan dalam pengelolaan data. Langkah-langkahnya meliputi pemilihan data, prapemrosesan data, transformasi data, penambangan data, dan evaluasi. Detil setiap tahapan dapat dilihat pada gambar yang tersedia, yang memberikan gambaran menyeluruh mengenai rencana pelaksanaan penelitian yang terstruktur dan terukur.



Gambar 2. Alur Metode Penelitian

3.1. Seleksi Data

Pada tahap ini dilakukan pemilihan data dengan menggunakan data hasil panen 10573edelai wilayah Soulad pada tahun 2013 hingga tahun 2023. Tujuan dari proses ini adalah untuk mendapatkan dataset yang akan digunakan untuk membangun model prediktif

dengan menggunakan metode Regresi Linear Berganda yang dipilih.

3.2. Pra-Pemrosesan Data

Pada 10574edelai ini, data diperiksa untuk mengidentifikasi dan menghapus data duplikat dan menyelesaikan nilai nol. Tujuan dari proses ini adalah untuk mengoptimalkan data 10574edelai yang akan diproses untuk meminimalkan potensi kesalahan dan memastikan bahwa hasil yang dihitung dengan algoritma regresi linier berganda sesuai dengan yang diharapkan.

3.3. Transformasi Data

Pada 10574edelai ini, data diubah ke dalam format yang diperlukan sesuai persyaratan dan persyaratan algoritma Regresi Linear Berganda yang digunakan oleh peneliti. Saat mentransformasikan dataset yang dipilih, data dibagi menjadi dua bagian, yaitu data latih dan data uji. Proses ini sangat penting karena membantu menjamin kualitas hasil tahap data mining.

3.4. Data Mining

Fase ini melibatkan proses eksplorasi data menggunakan metode regresi linier berganda. Dalam kerangka ini, variabel 10574edelai10574ent mengacu pada atribut-atribut yang dipilih sebagai 10574edela yang berpotensi mempengaruhi variabel dependen. Prosesnya dimulai dengan memilih atribut-atribut sebagai variabel 10574edelai10574ent dalam model regresi. Setelah itu, parameter model diestimasi untuk mencari hubungan linier antara variabel 10574edelai10574ent (atribut) dan variabel dependen.

3.5. Evaluasi

Langkah selanjutnya adalah mengevaluasi dan menyesuaikan model untuk mengurangi kesalahan prediksi. Hasil dari proses ini berupa persamaan regresi yang memberikan perkiraan nilai variabel terikat berdasarkan nilai variabel bebas yang dipilih. Dalam hal ini, variabel 10574edelai10574ent mencerminkan 10574edela-faktor yang dapat diubah atau dimanipulasi, sedangkan variabel dependen mewakili hasil yang diprediksi. Tujuan keseluruhan dari proses ini adalah untuk menciptakan model prediktif yang dapat digunakan untuk melakukan analisis mendalam terhadap data yang digunakan.

3.6. Deployment

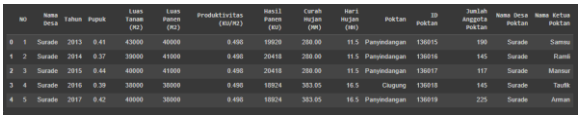
Untuk meningkatkan kemudahan penggunaan, para peneliti berhasil membuat antarmuka akhir berbasis website untuk memprediksi hasil 10574edelai di kecamatan Soulard. Saat mengembangkan 10574edela ini, peneliti menggunakan Flask sebagai editor teks untuk merancang dan mengatur struktur dan logika program. Secara spesifik, hasil pemodelan yang diperoleh dari data mining sebelumnya kemudian diintegrasikan ke dalam 10574edela berbasis website.

4. HASIL DAN PEMBAHASAN

Bab ini membahas tentang implementasi dan analisis penelitian prediksi hasil panen kedelai di Kecamatan Surade dengan menggunakan regresi linier berganda. Bab ini menjelaskan seluruh tahapan tersebut secara detail, mulai dari persiapan data, pelatihan model menggunakan regresi linier berganda, dan evaluasi hasil prediksi.

4.1. *Data Selection*

Pada tahap ini, perpustakaan Pandas digunakan untuk memuat data panen kedelai dari file Excel. Dataset awal panen kedelai di wilayah Soulad terdiri dari 134 data historis dari tahap pemilihan data. Interpretasi dari data yang dikumpulkan adalah sebagai berikut:



	NO	Nama Desa	Tahun	Pupuk	Luas Tanam (M2)	Luas Panen (M2)	Produktivitas (KU/M2)	Hasil Panen (KU)	Curah Hujan (MM)	Hari Hujan (HH)	Poktan	ID Poktan	Jumlah Anggota Poktan	Nama Desa Poktan	Nama Ketua Poktan
0	1	Surade	2013	0.41	43000	40000	0.498	19920	280.00	11.5	Panyindangan	136015	190	Surade	Samsu
1	2	Surade	2014	0.37	39000	41000	0.498	20418	280.00	11.5	Panyindangan	136016	145	Surade	Ramli
2	3	Surade	2015	0.44	40000	41000	0.498	20418	280.00	11.5	Panyindangan	136017	117	Surade	Mansur
3	4	Surade	2016	0.39	38000	38000	0.498	18924	383.05	16.5	Cugung	136018	145	Surade	Taufik
4	5	Surade	2017	0.42	40000	38000	0.498	18924	383.05	16.5	Panyindangan	136019	225	Surade	Arman

Gambar 2. Membaca Atribut Dataset

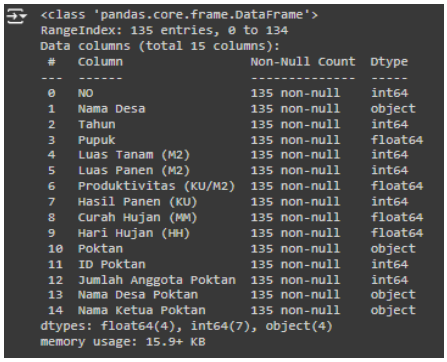
Berikut beberapa atribut-atribut yang terdapat pada dataset:

Tabel 1. Atribut Dataset

NO	Atribut
1	Nama Desa
2	Tahun
3	Luas Tanam
4	Luas Panen
5	Produktivitas
6	Hasil Panen
7	Pupuk
8	Curah Hujan
9	Hari Hujan
10	Poktan
11	ID Poktan
12	Jumlah Anggota Poktan
13	Nama Desa Poktan
14	Nama Ketua Poktan

4.2. *Preprocessing*

Langkah pertama melibatkan peninjauan informasi dasar tentang kumpulan data dan menentukan operasi apa yang perlu dilakukan pada setiap kolom yang akan digunakan.



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 135 entries, 0 to 134
Data columns (total 15 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   NO                     135 non-null    int64
 1   Nama Desa              135 non-null    object
 2   Tahun                  135 non-null    int64
 3   Pupuk                  135 non-null    float64
 4   Luas Tanam (M2)        135 non-null    int64
 5   Luas Panen (M2)        135 non-null    int64
 6   Produktivitas (KU/M2)  135 non-null    float64
 7   Hasil Panen (KU)       135 non-null    int64
 8   Curah Hujan (MM)       135 non-null    float64
 9   Hari Hujan (HH)        135 non-null    float64
 10  Poktan                 135 non-null    object
 11  ID Poktan              135 non-null    int64
 12  Jumlah Anggota Poktan  135 non-null    int64
 13  Nama Desa Poktan       135 non-null    object
 14  Nama Ketua Poktan      135 non-null    object
dtypes: float64(4), int64(7), object(4)
memory usage: 15.9+ KB
```

Gambar 3. Mencari Informasi Dataset

Selanjutnya mencari data bernilai duplikat. Seperti yang ditunjukkan pada gambar dibawah, terdapat data berisi baris duplikat, yaitu 3 baris data.

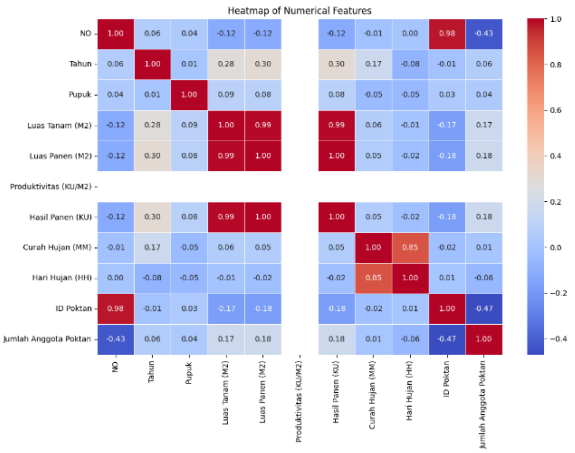
Gambar 4. Mencari Duplikasi Data

Dari uraian ini terlihat bahwa beberapa kolom berisi data kategorikal (objek), sedangkan kolom lainnya berisi data numerik (int64 dan float64).

```
NO                          int64
Nama Desa                  object
Tahun                       int64
Pupuk                     float64
Luas Tanam (M2)             int64
Luas Panen (M2)             int64
Produktivitas (KU/M2)     float64
Hasil Panen (KU)            int64
Curah Hujan (MM)          float64
Hari Hujan (HH)           float64
Poktan                     object
ID Poktan                   int64
Jumlah Anggota Poktan       int64
Nama Desa Poktan           object
Nama Ketua Poktan          object
dtype: object
```

Gambar 5. Memeriksa Tipe Data

Untuk memahami hubungan antara fitur-fitur ini, matriks korelasi dihitung dan divisualisasikan menggunakan Seaborn dan Matplotlib. Matriks korelasi memungkinkan kita menemukan hubungan antar variabel dan mengidentifikasi potensi pola atau tren dalam data.



Gambar 6. Visualisasi Heatmap Matriks Korelasi

Keterangan:

a. 0: Tidak ada korelasi linier
b. 0.0 hingga 0.3 (atau -0.0 hingga -0.3): Korelasi lemah
c. 0.3 hingga 0.7 (atau -0.3 hingga -0.7): Korelasi sedang
d. 0.7 hingga 1.0 (atau -0.7 hingga -1.0): Korelasi kuat

4.3. *Transformation*

Saat membuat model pembelajaran mesin, data perlu dipisahkan menjadi fitur (*x*)dan target (*y*).

```
1. data_digunakan =
   data_preprocessing[['Luas Tanam
   (M2)','Luas Panen
   (M2)','Produktivitas
   (KU/M2)','Hasil Panen (KU)']]
2. X =
   data_no_outliers.drop(columns=[
   'Hasil Panen (KU)'])
3. y = data_no_outliers['Hasil
   Panen (KU)']
```

Selanjutnya, Ssandardisasi data menggunakan StandardScaler diperlukan untuk memastikan semua fitur memiliki skala yang sama. Standarisasi ini membantu model pembelajaran mesin berperforma lebih baik dengan mengurangi perbedaan skala antara fitur-fitur yang ada.

	Luas Tanam (M2)	Luas Panen (M2)	Produktivitas (KU/M2)
0	0.740260	0.625	0.0
1	0.636364	0.650	0.0
2	0.662338	0.650	0.0
3	0.610390	0.575	0.0
4	0.662338	0.575	0.0
..	...	...	...
125	0.870130	0.825	0.0
126	0.792208	0.750	0.0
127	0.818182	0.725	0.0
128	0.766234	0.775	0.0
129	0.740260	0.700	0.0

[130 rows x 3 columns]

Gambar 7. Standarisasi Data

4.4. *Data Mining*

Menghitung Regresi Linear Berganda Menggunakan Colab. Dalam analisis data ini, kita menggunakan regresi linear berganda untuk memodelkan hubungan antara beberapa variabel independen dan variabel dependen. Kode berikut menggunakan pustaka *numpy* untuk manipulasi *array* dan *statsmodels* untuk analisis regresi. Data yang digunakan mencakup empat variabel: Luas Tanam, Luas Panen, Produktivitas, dan Hasil Panen. Tujuan dari model ini adalah untuk menentukan bagaimana variabel-variabel independen mempengaruhi Hasil Panen.

```
1. import numpy as np
2. import statsmodels.api as sm
3.
4. # Data: Luas Tanam, Luas Panen,
   Produktivitas, Hasil Panen
5. data = np.array([
6.     [43000, 40000, 0.498,
   19920],
7.     [39000, 41000, 0.498,
   20418],
8.     [40000, 41000, 0.498,
   20418],
9.     [38000, 38000, 0.498,
   18924],
10. [40000, 38000, 0.498, 18924]
11. ])
```

```
12.
13. # Memisahkan variabel independen (X) dan variabel dependen (y)
14. X = data[:, :-1]  # Luas Tanam, Luas Panen, Produktivitas
15. y = data[:, -1]   # Hasil Panen
16.
17. # Menambahkan kolom satu (intercept) ke X
18. X = sm.add_constant(X)
19.
20. # Membuat model regresi linear berganda
21. model = sm.OLS(y, X).fit()
22.
23. # Mendapatkan koefisien dari model
24. beta = model.params
25.
26. # Menampilkan koefisien
27. print("Intercept (β0):", beta[0])
28. print("Koefisien (β1) untuk Luas Tanam:", beta[1])
29. print("Koefisien (β2) untuk Luas Panen:", beta[2])
30. print("Koefisien (β3) untuk Produktivitas:", beta[3] if len(beta) > 3 else "Tidak ada")
31.
32. # Menghitung prediksi Hasil Panen berdasarkan model regresi
33. prediksi_hasil_panen = model.predict(X)
34.
35. # Menampilkan prediksi hasil panen
36. for i, prediksi in enumerate(prediksi_hasil_panen):
37.     print(f"Prediksi Hasil Panen untuk data {i+1}: {prediksi:.3f}")
38.
39. # Menampilkan ringkasan model
40. print(model.summary())
```

```
Intercept (β0): -2.0486000049113606e-16
Koefisien (β1) untuk Luas Tanam: 0.4980000000000022
Koefisien (β2) untuk Luas Panen: -1.8318502270631143e-10
Koefisien (β3) untuk Produktivitas: Tidak ada
Prediksi Hasil Panen untuk data 1: 19920.000
Prediksi Hasil Panen untuk data 2: 20418.000
Prediksi Hasil Panen untuk data 3: 20418.000
Prediksi Hasil Panen untuk data 4: 18924.000
Prediksi Hasil Panen untuk data 5: 18924.000
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 3.518e+27
Date:                Tue, 30 Jul 2024   Prob (F-statistic):           2.84e-28
Time:                        14:54:42   Log-Likelihood:                 118.90
No. Observations:                   5   AIC:                            -231.8
Df Residuals:                       2   BIC:                            -233.0
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1         -2.049e-16   4.99e-15     -0.041      0.971   -2.17e-14    2.13e-14
x2             0.4980   6.16e-15   8.09e+13      0.000       0.498       0.498
x3         -1.832e-10   5.45e-10     -0.336      0.769   -2.53e-09    2.16e-09
==============================================================================
Omnibus:                        nan   Durbin-Watson:                   0.102
Prob(Omnibus):                  nan   Jarque-Bera (JB):                0.638
Skew:                         0.000   Prob(JB):                        0.727
Kurtosis:                     1.250   Cond. No.                     3.81e+06
==============================================================================
```

Gambar 8. Perhitungan Regresi Linier Berganda Menggunakan Colab

Selanjutnya, membagi data *training* dan data *testing*. Tujuan dari pemisahan data pelatihan dan pengujian adalah untuk menguji kinerja model pada data yang belum pernah dilihat sebelumnya dan untuk memastikan bahwa model tersebut tidak hanya cocok dengan set pelatihan.

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

Selanjutnya, model regresi linier berganda dibuat menggunakan LinearRegression dari scikit-learn dan dilatih pada data pelatihan. Model tersebut menggabungkan satu variabel terikat dengan beberapa variabel bebas. Pickle digunakan untuk menyimpan model yang dilatih sehingga dapat digunakan kembali nanti tanpa pelatihan ulang.

```
1. model = LinearRegression()
2. model.fit(X_train, y_train)
3.
4. with open('linear_regresion.pkl', 'wb') as f:
5.     pickle.dump(model, f)
6.
7. y_pred = model.predict(X_test)
```

4.5. *Interpretation/Evaluation*

Performa model dievaluasi menggunakan berbagai metrik evaluasi seperti MSE & RMSE.

```
1. # Evaluasi model
2. rmse = np.sqrt(mean_squared_error(y_test, y_pred))
3. mae = mean_absolute_error(y_test, y_pred)
4. mse = mean_squared_error(y_test, y_pred)
5. r2 = r2_score(y_test, y_pred)
6.
7. rmse = np.sqrt(mean_squared_error(y_test, y_pred))
8. mae = mean_absolute_error(y_test, y_pred)
9. mse = mean_squared_error(y_test, y_pred)
10.
11. print(f"Root Mean Squared Error: {rmse}")
12. print(f"Mean Absolute Error: {mae}")
13. print(f"Mean Squared Error: {mse}")
14. print(f"R² Score: {r2}")
15.
```

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

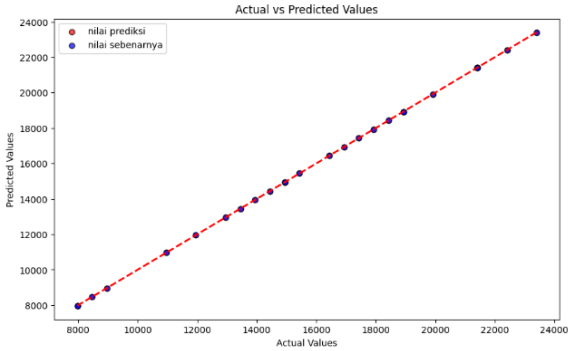
```
Root Mean Squared Error: 1.9375570689486726e-12
Mean Absolute Error: 1.2593003563009775e-12
Mean Squared Error: 3.754127395432972e-24
R² Score: 1.0
Koefisien:  [2.64591776e-12 1.99200000e+04 0.00000000e+00]
Intercept:  7469.999999999998
```

Gambar 9. Menguji Performa Model

Berikut ini penjelasan mengenai temuan evaluasi pola:

a. RMSE sebesar 1,93755 dalam kasus ini menunjukkan bahwa kesalahan prediksi model cukup kecil.
b. Prediksi model menyimpang dari nilai sebenarnya rata-rata sebesar 1,25930, sesuai dengan nilai MAE sebesar 1,25930.
c. Model ini memiliki mean square error yang sangat kecil, seperti yang ditunjukkan oleh nilai MSE sebesar 3,74412.
d. Nilai R2 sebesar 1,0 menunjukkan bahwa variasi data hampir seluruhnya dijelaskan oleh model.
e. Koefisien menunjukkan bagaimana setiap fitur mempengaruhi prediksi model dalam hubungannya dengan fitur lainnya. Koefisien fitur pertama sebesar 2,64591, fitur kedua sebesar 1,99200, dan seterusnya.
f. Intersep 7469,9 menunjukkan bahwa model memprediksi nilai 7469,9 ketika semua fitur bernilai nol

Kemudian plot sebar dibuat untuk membandingkan nilai aktual dengan nilai prediksi untuk memahami secara jelas keakuratan model dalam memprediksi hasil panen.



Gambar 10. Grafik Interpretasi Model

Plot ini menunjukkan tingkat kesesuaian antara nilai prediksi dan nilai aktual. Garis merah putus-putus mewakili garis ideal di mana nilai prediksi dan nilai aktual sama. Jika titik data tersebar di sepanjang garis ini, berarti model bekerja dengan baik.

4.6. *Deployment*

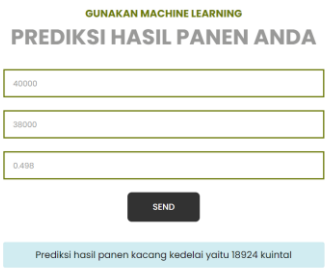
Tahapan ini membuat model pembelajaran mesin pada tahap *deployment* yang dapat memperkirakan hasil panen kedelai berdasarkan penerapannya di lapangan. Halaman "Beranda" ini menunjukkan bagaimana situs web Kecamatan Surade mengintegrasikan model prediksi hasil panen.

Gambar 11. Tampilan Beranda

Halaman "Tentang Kami" ini memberikan pengenalan singkat tentang Kecamatan Surade, dengan fokus pada budidaya kedelai dan sejarah pendirian lembaga tersebut.

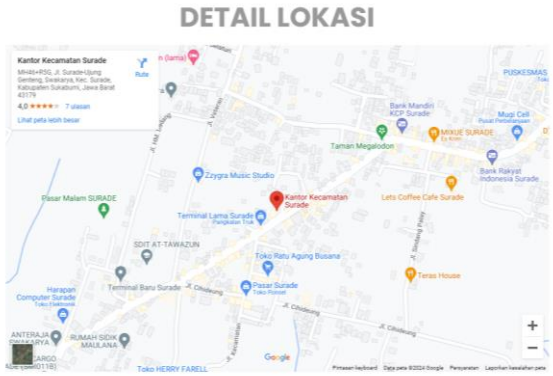
Gambar 12. Tampilan Tentang Kami

Petani dapat menggunakan halaman "Prediksi Hasil Panen" untuk memperkirakan hasil kedelai dengan cepat berdasarkan informasi seperti luas tanam, luas panen, dan produktivitas.



Gambar 13. Tampilan Prediksi Hasil Panen

Pengguna dapat dengan mudah menemukan Kecamatan Surade dengan bantuan peta interaktif di halaman "Lokasi". Integrasi Google Maps memungkinkan pengguna melihat rute secara langsung dan mendapatkan petunjuk arah.



Gambar 14. Tampilan Lokasi

Halaman "Contact" ini memberikan informasi lengkap tentang Kecamatan Surade.

Gambar 15. Tampilan Contact

4.7. Pengujian Fungsionalitas

Pengujian fungsionalitas merujuk pada proses untuk menilai apakah website yang dikembangkan sudah memenuhi standar kelayakan penggunaan. Berikut ini adalah hasil dari pengujian pada Website Prediksi Hasil Panen Kacang Kedelai yang telah dilaksanakan:

Tabel 2. Pengujian

Pengujian Halaman *Home*				
No	**Pengujian Fungsi**	**Deskripsi**	***Output***	**Hasil Pengujian**
1	Halaman *Home*	Peneliti melakukan *Klik* pada halaman *Home*	Jika meng klik *home* dengan benar, maka peneliti akan beralih ke halaman Website Prediksi hasil panen kacang kedalai	SUKSES
Pengujian Halaman Tentang kami				
2	Halaman Tentang kami	Peneliti melakukan *Klik* pada halamanTentang kami	Jika menekan Tentang kami maka Website menampilkan halaman Tentang kami	SUKSES
Pengujian Halaman Prediksi Hasil Panen				
3	Halaman Prediksi Hasil Panen	Peneliti melakukan *Klik* pada Prediksi Hasil Panen	Jika mengisi form luas lahan, luas panen dan produktivitas maka akan menghasilkan hasil dari prediksi	SUKSES
Pengujian Halaman Lokasi				
4	Halaman Lokasi	Peneliti melakukan *Klik* pada halaman Lokasi	Jika menekan Lokasi maka Website menampilkan halaman Lokasi	SUKSES
Pengujian Halaman *Contact*				
5	Halaman Contact	Peneliti melakukan *Klik* pada halaman *Contact*	Jika menekan *Contact* maka Website menampilkan halaman *Contact*	SUKSES

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan dan mengimplementasikan model prediksi hasil panen kacang kedelai menggunakan regresi linear berganda, yang menunjukkan performa yang cukup baik dengan nilai Mean Squared Error (MSE) sebesar 3,74412 dan Root Mean Squared Error (RMSE) sebesar 1,93755. Nilai ini menunjukkan bahwa model memiliki kesalahan prediksi yang rendah, sehingga dapat diandalkan untuk memprediksi hasil panen di Kecamatan Surade. Model ini juga telah diintegrasikan ke dalam tampilan web untuk memberikan manfaat langsung kepada para pemangku kepentingan.

Meskipun hasilnya menunjukkan akurasi yang baik, penelitian ini memiliki keterbatasan dalam hal penggunaan satu jenis algoritma. Untuk meningkatkan akurasi dan kinerja prediksi di masa mendatang, disarankan untuk membandingkan model regresi linear berganda dengan algoritma prediksi alternatif lainnya. Penelitian ini memberikan kontribusi yang signifikan baik secara praktis maupun akademis, dan diharapkan dapat membantu dalam perencanaan pertanian yang lebih efektif di wilayah ini.

DAFTAR PUSTAKA

[1] S. G. I. Pogaga, P. Kindangen, and R. A. M. Koleangan, "Analisis Pengaruh Produktivitas Pertanian Dan Pendidikan Terhadap Tendapatan Rumah Tangga Di Kabupaten Minahasa Tenggara," *J. Pembanguan Ekon. dan Keuang. Drh.*, vol. 20, no. 04, pp. 54–70, 2020.

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

[3] I. Zufria and H. Santoso, "Sistem Pakar Menggunakan Metode Backward Chaining Untuk Mengantisipasi Permasalahan Tanaman Kacang Kedelai Berbasis Web," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 1, pp. 20–28, 2021.

[4] E. Triyanto, H. Sismoro, and A. D. Laksito, "Implementasi Algoritma Regresi Linear

Berganda Untuk Memprediksi Produksi Padi Di Kabupaten Bantul," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 4, no. 2, pp. 66–75, 2019, doi: 10.36341/rabit.v4i2.666.

[5] Salahuddin, M. Khadafi, Huzeini, and M. Davi, "Rancang Bangun Aplikasi Machine Learning Prediksi Hasil Panen Buah Pinang (ArecaCatechu) Menggunakan Metode Regresi Linier Berganda," *Proceeding Semin. Nas. Politek. Negeri Lhokseumawe*, vol. 6, no. 1, pp. 181–186, 2022.

[6] N. Giarsyani, "Komparasi Algoritma Machine Learning dan Deep Learning untuk Named Entity Recognition : Studi Kasus Data Kebencanaan," *Indones. J. Appl. Informatics*, vol. 4, no. 2, p. 138, 2020, doi: 10.20961/ijai.v4i2.41317.

[7] A. J. Karlina, M. Irsyad, F. Insani, and E. P. Cynthia, "Estimasi Hasil Panen Ayam Pedaging Menggunakan Algoritma Regresi Linear Berganda," *KLIK Kaji. Ilm. ...*, vol. 3, no. 6, pp. 966–976, 2023, doi: 10.30865/klik.v3i6.920.

[8] D. P. dan P. K. Ngawi, "Klasifikasi Kedelai, Tanaman Segudang Manfaat," Website Dinas Pertanian dan Pangan Kabupaten Ngawi. [Online]. Available: https://pertanian.ngawikab.go.id/2023/01/02/klasifikasi-kedelai-tanaman-segudang-manfaat/

[9] S. Kurniasari, "OPTIMASI FUNGSI KEANGGOTAAN FUZZY TSUKAMOTO MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI KEJADIAN BANJIR DI KOTA BALIKPAPAN," pp. 6–23, 2020.

[10] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear : Studi Kasus pada Industri Otomotif di Indonesia," vol. 4, pp. 1199–1216, 2024.

[11] R. G. Guntara, "Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab," *J. Ilm. Multidisiplin*, vol. 2, no. 6, pp. 2091–2100, 2023.

[12] P. Banerjee, B. Kumar, A. Singh, R. Kumar, and 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, pp. 24890–24893, 2019, doi: 10.18535/ijecs/v8i12.4399.