

Analisis Luas Panen dalam memprediksi Produksi Padi di Kota Binjai menggunakan Regresi Linear

Malawika¹, Reza Alamsyah², Irwan Jani Tarigan¹

¹ Sistem Informasi, STMIK Methodist Binjai, Binjai, Indonesia

² Manajemen informatika, Politeknik Ganesha Medan, Medan, Indonesia

Email: ¹malawika431@email.com, ²rezaalamsyah@gmail.com, ¹irwanjantarigan@email.com

Email Penulis Korespondensi: malawika431@email.com

Abstrak– Penelitian ini bertujuan menganalisis hubungan antara luas panen dengan produksi padi serta membangun model prediksi menggunakan algoritma regresi linear sederhana. Pendekatan penelitian yang digunakan adalah kuantitatif dengan data sekunder dari Badan Pusat Statistik (BPS) Kota Binjai periode 2020–2024. Variabel prediktor yang digunakan adalah luas panen (hektar), sedangkan variabel target adalah produksi padi (ton). Pengolahan data dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab. Dataset dibagi menjadi data latih (80%) dan data uji (20%), kemudian dievaluasi menggunakan Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), serta Koefisien Determinasi (R^2). Hasil penelitian menunjukkan bahwa luas panen memiliki hubungan linear yang sangat kuat dengan produksi padi, dengan persamaan regresi $Y = -8,97 + 6,07X$. Evaluasi model menghasilkan nilai $R^2 = 0,9990$, MAE = 56,87, MSE = 5.092,47, dan RMSE = 71,36. Nilai tersebut membuktikan bahwa model memiliki tingkat akurasi yang tinggi. Dengan demikian, regresi linear sederhana efektif digunakan untuk memprediksi produksi padi di Kota Binjai, serta dapat menjadi acuan bagi petani maupun pemangku kepentingan dalam perencanaan strategi pertanian ke depan.

Kata Kunci: Produksi padi, Luas panen, Prediksi, Regresi Linear Sederhana, Google Colab.

Abstract–This study aims to analyze the relationship between harvested area and rice production and build a prediction model using a simple linear regression algorithm. The research approach employed is quantitative, utilizing secondary data from the Central Statistics Agency (BPS) of Binjai City for the period from 2020 to 2024. The predictor variable used is harvested area (in hectares), while the target variable is rice production (in tons). Data processing was carried out using the Python programming language on the Google Colaboratory platform. The dataset was divided into training data (80%) and test data (20%), then evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). The results showed that the harvested area has a very strong linear relationship with rice production, with a regression equation of $Y = -8.97 + 6.07X$. The model evaluation yielded R^2 values of 0.9990, MAE of 56.87, MSE of 5,092.47, and RMSE of 71.36. These values demonstrate the model's high accuracy. Therefore, simple linear regression is effective in predicting rice production in Binjai City and can serve as a reference for farmers and stakeholders in planning future agricultural strategies

Keywords: Rice production, Harvested Area, Prediction, Simple Linear Regression, Google Colab.

1. PENDAHULUAN

Padi menjadi komoditas strategis dalam sistem pangan nasional, Indonesia sebagai negara yang mayoritas penduduknya bergantung pada pertanian, menjadikan sektor ini kunci utama ketahanan pangan [1]. Kota Binjai[2], [3], [4], salah satu wilayah di Sumatera Utara, memiliki potensi besar di bidang pertanian, khususnya padi. Akan tetapi, jumlah produksi padi di wilayah tersebut tidak selalu stabil dari tahun ke tahun. Data dari tahun 2020 hingga 2024 menunjukkan bahwa meskipun secara total produksi padi di Kota Binjai mengalami peningkatan, terdapat perbedaan yang cukup mencolok antar kecamatan[5]. Beberapa kecamatan memiliki luas panen yang tinggi dan hasil produksi yang besar, sedangkan beberapa lainnya tidak menunjukkan aktivitas panen sama sekali karena keterbatasan lahan, seperti yang terjadi di daerah perkotaan.

Tingkat produksi padi ditentukan oleh berbagai faktor, salah satunya adalah luas panen, luas panen merupakan indikator penting yang secara langsung berkorelasi dengan jumlah hasil panen yang diperoleh [6]. Namun dalam praktiknya, banyak petani di Kota Binjai belum mengetahui secara pasti berapa hasil panen yang akan mereka peroleh berdasarkan luas lahan yang mereka miliki. Kondisi ini menyebabkan adanya ketidakpastian dalam perencanaan tanam, distribusi hasil, hingga komunikasi dengan pemasok atau pembeli. Kemampuan untuk memprediksi hasil panen berdasarkan luas lahan sangat penting[7].

Dengan adanya prediksi, petani dapat merencanakan strategi produksi secara lebih baik, menyesuaikan kebutuhan input seperti pupuk dan tenaga kerja, serta memperkirakan pendapatan yang akan diterima [8]. Lebih dari itu, prediksi yang akurat juga dapat meningkatkan rasa percaya diri petani dalam melakukan komunikasi atau negosiasi dengan pemasok atau pembeli hasil panen.

Salah satu pendekatan yang dapat digunakan adalah dengan membangun model prediksi produksi padi berdasarkan luas panen menggunakan algoritma regresi linear sederhana. Pendekatan ini memungkinkan analisis hubungan antara dua variabel dan memberikan gambaran prediktif berdasarkan data yang ada. Dengan bantuan bahasa pemrograman Python dan library pendukung, proses perhitungan dapat dilakukan secara efisien dan akurat[9].

Berdasarkan permasalahan yang telah di temukan, maka telah di tentukan judul “Analisis Luas Panen dalam memprediksi Produksi Padi di Kota Binjai menggunakan Regresi Linear Sederhana” Penelitian ini,

menerapkan algoritma regresi linear sederhana untuk memprediksi produksi padi berdasarkan luas panen di Kota Binjai dengan data dari tahun 2020 hingga 2024.

2. METODOLOGI PENELITIAN

Dalam penelitian ini, regresi linear sederhana dipilih sebagai metode prediksi karena kesederhanaannya dan kemampuannya untuk memberikan gambaran awal yang baik tentang hubungan antara luas panen dengan hasil produksi padi [10]. Pendekatan analisis data kuantitatif telah digunakan dalam berbagai penelitian pertanian, mulai dari peramalan produksi, analisis hubungan antar variabel iklim dan hasil panen, hingga pengembangan sistem informasi pertanian berbasis teknologi. Dalam era digital, pemanfaatan teknologi komputasi seperti Python dalam analisis prediktif telah menjadi solusi yang efektif karena fleksibilitas dan kapabilitasnya dalam mengelola dan menganalisis data dalam skala besar.

2.1 Tahapan Penelitian

a. Mulai

Tahapan awal dimulai dengan penetapan topik dan tujuan penelitian, yaitu untuk membangun model prediksi. Penelitian ini dilakukan sebagai upaya penerapan algoritma regresi linear sederhana menggunakan *Python* untuk mempermudah proses estimasi produksi padi.

b. Identifikasi Masalah

Pada bagian ini, peneliti menelaah serta menemukan masalah yang terjadi di lapangan maupun dalam data. Permasalahan yang diangkat dalam penelitian ini adalah ketidakpastian dalam memperkirakan hasil panen padi yang dapat berdampak pada kebijakan pertanian. Oleh karena itu, dibutuhkan suatu metode prediksi yang sederhana dan efektif, salah satunya menggunakan regresi linear sederhana.

c. Pengumpulan Data

Tahap pengumpulan data mencakup pengambilan data sekunder terkait luas panen dan produksi padi selama tahun 2020 sampai 2024. Sumber data berasal dari Badan Pusat Statistik (BPS) Kota Binjai.

d. Persiapan Data

Data yang telah dikumpulkan kemudian dipersiapkan melalui proses *pre-processing*. Tahapan ini bertujuan untuk mempersiapkan data agar siap digunakan dalam proses pemodelan. Beberapa proses yang dilakukan meliputi: [11][12]

1. Penghapusan duplikasi data

```
print("Jumlah Baris Duplikat")
print(df.duplicated().sum())
```

```
Jumlah Baris Duplikat
```

```
0
```

Gambar 2. Penghapusan data duplikat

Hasil `df.duplicated().sum()` juga menunjukkan nilai 0, yang berarti tidak ada baris data yang terduplikasi.

2. Penanganan nilai yang hilang (*missing values*)

```
print("Missing Values")
print(df.isnull().sum())
```

```
Missing Values
```

```
Kecamatan      0
```

```
Luas_Panen     0
```

```
Produksi       0
```

```
Tahun          0
```

```
dtype: int64
```

Gambar 3. Cek Missing Value

Selanjutnya pada gambar 4.4 dilakukan pengecekan terhadap keberadaan nilai kosong (*missing values*), Hasil pengecekan menggunakan fungsi `df.isnull().sum()` menunjukkan bahwa tidak terdapat nilai kosong pada seluruh atribut dalam *dataset*.

3. Penghapusan data yang tidak valid

```
# Hapus data dengan luas panen = 0
df = df[df['Luas_Panen'] > 0]
```

Gambar 4. Hapus data yang tidak valid

Proses ini dilakukan dengan menyaring data agar hanya baris dengan nilai `Luas_Panen > 0` yang dipertahankan. Hal ini bertujuan untuk meningkatkan akurasi model dan menghindari hasil prediksi yang bias akibat keberadaan data kosong atau tidak valid dalam variabel independen (X).

e. Pembagian Dataset

Dataset dibagi menjadi dua subset, yaitu *training data* dan *testing data*, dengan tujuan untuk membangun model prediksi serta menilai tingkat akurasinya.

f. Pelatihan Model

Model regresi linear sederhana dibangun menggunakan data latih. Model ini akan mempelajari hubungan antara variabel independen (luas panen) terhadap variabel dependen (produksi padi).

g. Perhitungan Parameter Model

Parameter model (koefisien regresi dan *intercept*) dihitung menggunakan metode statistik (seperti metode kuadrat terkecil / *least squares*), baik secara manual dan dengan bantuan program *Python*.

h. Prediksi Produksi Padi

Model yang telah dilatih kemudian digunakan untuk memprediksi nilai produksi padi berdasarkan nilai-nilai luas panen tertentu pada data uji.

i. Evaluasi Model

Setelah model selesai dilatih, dilakukan evaluasi terhadap model tersebut menggunakan data uji. Evaluasi dilakukan dengan menghitung beberapa metrik, seperti: [13][14], [15], [16]

- Mean Absolute Error* (MAE)
- Mean Squared Error* (MSE)
- Root Mean Squared Error* (RMSE)
- Koefisien Determinasi* (R^2)

2.2 Data Mining

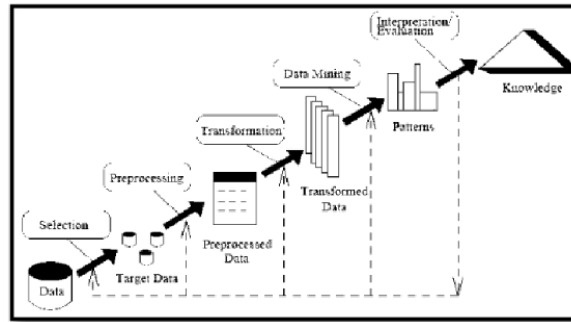
Data mining merupakan proses ekstraksi atau penyaringan data dari himpunan data berukuran besar melalui tahapan-tahapan sistematis guna memperoleh informasi yang bernilai. Proses ini berperan dalam menggali pola dan pengetahuan tersembunyi dari data, yang selanjutnya dapat dimanfaatkan untuk mendukung pengambilan keputusan di berbagai bidang [17][18], [19], [20], [21].

Data mining merupakan salah satu tahapan penting dalam keseluruhan proses KDD yang bertujuan untuk mengekstraksi pola atau model dari data dengan menggunakan algoritma tertentu [22]. Proses KDD melibatkan beberapa tahap. Tahap awal adalah seleksi data (*selection*), di mana data yang relevan dipilih untuk diproses lebih lanjut. Selanjutnya adalah pra-pemrosesan data, yaitu tahap pembersihan data untuk menghilangkan data yang tidak diperlukan atau duplikat.

Tahap berikutnya adalah transformasi data, yaitu mengubah data mentah menjadi format yang sesuai dengan kebutuhan analisis, misalnya dengan melakukan pengkodean atau normalisasi. Setelah itu, dilakukan tahap inti berupa data mining, yaitu penerapan algoritma tertentu pada data untuk menemukan pola. Terakhir, proses diakhiri dengan evaluasi, yang bertujuan untuk menilai apakah pola atau hasil yang diperoleh dari algoritma sudah optimal dan akurat.

2.3 Knowledge Discovery in Database (KDD)

Knowledge Discovery In Database (KDD) merupakan suatu proses untuk menentukan suatu informasi yang berguna serta menentukan suatu pola - pola yang ada didalam suatu data. Mencakup serangkaian tahapan, mulai dari pengumpulan data, pembersihan data, transformasi data, penambangan data (data mining), hingga interpretasi dan evaluasi hasil [23].



Gambar 5. Proses Knowledge Discovery In Database

2.4 Regresi Linear Sederhana (Simple Linear Regression)

Analisis regresi linear sederhana adalah hubungan secara linear antara satu variabel independen (X) dengan variabel dependen (Y). Analisis ini mengetahui arah hubungan antara variabel independen dengan variabel dependen apabila variabel independen mengalami kenaikan atau penurunan [24]. Model regresi linier sederhana dikatakan sebagai model dengan regresor tunggal yang memiliki hubungan linier dengan respons [25]. Model regresi linier sederhana adalah,

$$y = a + bX$$

Dimana:

- Y = variabel respons/output
- X = variabel prediktor/input
- a = konstanta
- b = kemiringan garis regresi

Dalam konteks machine learning, prediksi adalah proses memperkirakan nilai atau hasil di masa depan berdasarkan data dan pola yang telah dipelajari oleh model. Prediksi ini dapat berupa nilai numerik, kategori, atau distribusi probabilitas, tergantung pada jenis masalah dan data yang digunakan. Prediksi algoritmik dapat memberikan dukungan keputusan yang berharga dengan memprediksi secara akurat apakah seorang individu adalah Y, tetapi mengandalkan prediksi tersebut secara eksklusif untuk alokasi sumber daya mungkin menimbulkan kekurangan epistemic [26].

Dalam konteks penelitian ini, prediksi merujuk pada upaya untuk memperkirakan hasil produksi padi berdasarkan data historis tentang luas panen dan faktor-faktor lainnya yang relevan, seperti curah hujan, suhu, dan kelembapan. Prediksi ini dilakukan dengan menggunakan berbagai metode statistik atau algoritma machine learning yang dapat menangkap pola atau hubungan antar variabel [27].

Prediksi sangat penting dalam sektor pertanian karena dapat membantu petani, pembuat kebijakan, dan pihak terkait merencanakan tindakan yang lebih efektif. Misalnya, dengan adanya prediksi yang akurat tentang hasil produksi padi, pemerintah dapat lebih siap dalam mengelola distribusi beras, sementara petani dapat menentukan langkah-langkah yang tepat dalam mengelola lahan mereka.

Ada berbagai jenis metode yang digunakan dalam proses prediksi, di antaranya:

- a. Model Statistik, seperti regresi linear sederhana, yang mengasumsikan hubungan linier antara variabel input dan output.
- b. Metode Machine Learning, seperti pohon keputusan dan jaringan syaraf tiruan, yang lebih kompleks dan dapat menangkap pola yang tidak linier.
- c. Metode Time Series, yang digunakan jika data yang tersedia merupakan data waktu atau bersifat urut waktu, seperti data tahunan atau bulanan.

Dalam penelitian ini, regresi linear sederhana dipilih sebagai metode prediksi karena kesederhanaannya dan kemampuannya untuk memberikan gambaran awal yang baik tentang hubungan antara luas panen dengan hasil produksi padi. Pendekatan analisis data kuantitatif telah digunakan dalam berbagai penelitian pertanian, mulai dari peramalan produksi, analisis hubungan antar variabel iklim dan hasil panen, hingga pengembangan sistem informasi pertanian berbasis teknologi [28].

Dalam era digital, pemanfaatan teknologi komputasi seperti Python dalam analisis prediktif telah menjadi solusi yang efektif karena fleksibilitas dan kapabilitasnya dalam mengelola dan menganalisis data dalam skala besar.

3. HASIL DAN PEMBAHASAN

3.1 Proses pengolahan Data

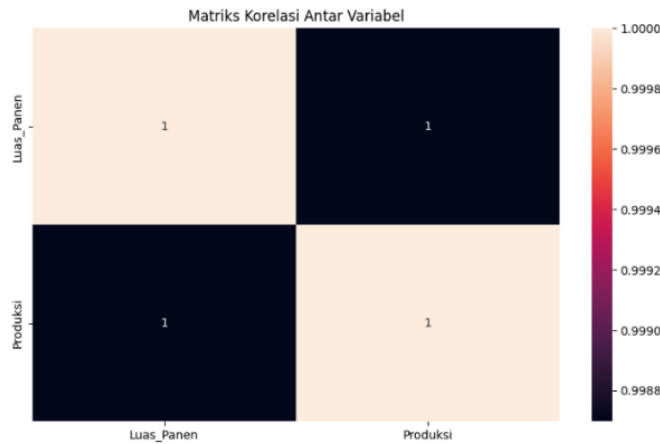
Proses prediksi dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab.

Tabel 2. Data yang sudah bersih

Kecamatan	Luas_Panen	Produksi	Tahun
Binjai Selatan	1046	6413	2020
Binjai Timur	300	1842	2020
Binjai Utara	1139	6983	2020
Binjai Barat	58	354	2020
Binjai Selatan	751	4631	2021
Binjai Timur	338	2077	2021
Binjai Utara	1097	6811	2021
Binjai Barat	45	279	2021
Binjai Selatan	1120	6660	2022
Binjai Timur	311	1827	2022
Binjai Utara	1388	8307	2022
Binjai Barat	46	275	2022
Binjai Selatan	1173	7195	2023
Binjai Timur	321	1958	2023
Binjai Utara	1488	9228	2023
Binjai Barat	53	322	2023
Binjai Selatan	1304	7276	2024
Binjai Timur	333	1892	2024
Binjai Utara	1749	10866	2024
Binjai Barat	45	290	2024

3.2 Pembuatan Model Regresi Linear Sederhana

Gambar 6. Visualisasi Heatmap antar Variabel



Untuk melihat hubungan linier antar variabel, dilakukan analisis korelasi menggunakan matriks korelasi [29]. Pada gambar 6 menunjukkan korelasi antara variabel Luas Panen dan Produksi Padi. Dimana hasil korelasi menunjukkan bahwa nilai korelasi antara Luas Panen dan Produksi mencapai 1.00, yang berarti terdapat hubungan linier yang sangat kuat antara kedua variabel tersebut [16]. Hal ini mendukung penggunaan algoritma regresi linear sederhana dalam penelitian ini.

```
# Pisahkan variabel prediktor (X) dan target (Y)
x = df['Luas_Panen']
y = df['Produksi']

# Bagi data jadi data latih dan uji
x_train, x_test, y_train, y_test = train_test_split(
    x, y, test_size=0.2, random_state=42
)
```

Gambar 7. Pembagian data *Training* dan data *Testing*

Pada tahap ini, dataset dibagi menjadi dua bagian, yaitu data pelatihan (training) dan data pengujian (testing) dengan proporsi 80:20. Sebagian besar data, yaitu 80%, digunakan untuk melatih model agar mampu mempelajari hubungan antara variabel input dan output. Sementara 20% sisanya dialokasikan untuk menguji performa model terhadap data yang belum dikenali sebelumnya,

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

Gambar 8. Model Regresi linear sederhana

Setelah data dibagi, proses selanjutnya adalah membangun model prediktif menggunakan algoritma Regresi Linear Sederhana. Proses ini dilakukan dengan memanggil objek `LinearRegression()` dari *library scikit-learn*, kemudian melatihnya menggunakan data latih.

```
print("Intercept (a):", model.intercept_)
print("Slope (b):", model.coef_[0])
print(f"Persamaan regresi: Y = {model.intercept_:.2f} + {model.coef_[0]:.2f}X")

Intercept (a): -8.972974726091707
Slope (b): 6.074562967398163
Persamaan regresi: Y = -8.97 + 6.07X
```

Gambar 9. Hasil Persamaan Regresi

Setelah proses pelatihan model regresi linear sederhana dilakukan, diperoleh nilai intercept (a) sebesar -8.97 dan slope (b) sebesar 6.07. Dengan demikian, persamaan regresi linear sederhana yang terbentuk adalah sebagai berikut:

$$Y = -8.97 + 6.07X$$

Persamaan regresi ini menunjukkan bahwa setiap penambahan satu hektar luas panen diperkirakan akan meningkatkan produksi padi sebesar 6.07 ton. Nilai intercept yang negatif menunjukkan bahwa jika tidak terdapat luas panen ($X = 0$), maka produksi padi juga diperkirakan mendekati nol, yang secara praktis memang sesuai dengan logika di lapangan.

```
y_pred = model.predict(x_test)

# Hitung metrik evaluasi
r2 = r2_score(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
mae = mean_absolute_error(y_test, y_pred)

print("R-squared:", r2)
print("MSE:", mse)
print("RMSE:", rmse)
print("MAE:", mae)

R-squared: 0.99902081517072
MSE: 5092.468756892473
RMSE: 71.3615355587542
MAE: 56.86545655143546
```

Gambar 10. Hasil Evaluasi Model

Pengujian dilakukan menggunakan data testing untuk mengetahui sejauh mana model mampu memprediksi data yang sebelumnya tidak tersedia bagi model. Adapun hasil evaluasi model dapat dilihat pada

Tabel 3. Hasil Evaluasi

Metrik	Nilai
R ² Score	0.9990
MSE	5092.47
RMSE	71.36
MAE	56.87

Dari evaluasi yang dilakukan, terbukti bahwa model regresi linear sederhana bekerja dengan sangat baik. Nilai R² sebesar 0,9990 memperlihatkan bahwa 99,90% variasi pada data produksi padi dapat diterangkan oleh variabel luas panen, yang menegaskan kuatnya hubungan linier di antara keduanya.

Selain itu, nilai MSE, RMSE, dan MAE yang rendah menunjukkan bahwa tingkat kesalahan prediksi model tergolong sangat kecil. Hal ini mengindikasikan bahwa model cukup akurat dan dapat diandalkan untuk melakukan prediksi produksi padi di masa yang akan datang.

3.3 Inteprestasi Hasil

Tabel 4. Produksi sebenarnya dan produksi prediksi

No	Kecamatan	Luas Panen	Produksi	Produksi Prediksi	Error
1	Binjai Selatan	1046	6413	6345.92	67.98

2	Binjai Timur	300	1842	1813.36	28.64
3	Binjai Utara	1139	6983	6909.95	73.05
4	Binjai Barat	58	354	343.35	10.65
5	Binjai Selatan	751	4631	4553.03	77.97
6	Binjai Timur	338	2077	2044.23	32.77
7	Binjai Utara	1097	6811	6654.83	156.17
8	Binjai Barat	45	279	264.38	14.62
9	Binjai Selatan	1120	6660	6794.54	-133.54
10	Binjai Timur	311	1827	1880.22	-53.22
11	Binjai Utara	388	3887	8422.52	-1535.52
12	Binjai Barat	46	275	270.46	4.54
13	Binjai Selatan	1173	7195	7116.49	78.51
14	Binjai Timur	321	1958	1940.96	17.04
15	Binjai Utara	1488	9228	9030.00	198.02
16	Binjai Barat	53	322	312.98	9.02
17	Binjai Selatan	1304	7276	7912.26	-636.26
18	Binjai Timur	333	1892	2013.86	-121.86
19	Binjai Utara	1749	10866	10615.44	250.56
20	Binjai Barat	45	290	264.38	25.62

Berdasarkan hasil yang disajikan pada tabel, dapat diamati bahwa model regresi linear sederhana yang digunakan dalam penelitian ini menunjukkan tingkat prediksi yang cukup baik terhadap nilai produksi padi aktual di Kota Binjai. Hal ini tercermin dari kedekatan antara nilai prediksi (Produksi Prediksi) dan nilai aktual (Produksi), khususnya pada data dengan nilai luas panen yang tinggi.

Sebagai contoh, pada data Kecamatan Binjai Utara dengan luas panen sebesar 1.749 hektar, model memprediksi produksi padi sebesar 10.615,4 satuan, sedangkan nilai aktualnya adalah 10.866, sehingga menghasilkan selisih (error) sekitar -250,56. Nilai error ini termasuk kecil, yang mengindikasikan bahwa model mampu menghasilkan prediksi yang cukup akurat dalam kondisi nyata.

Selain itu, performa model tampak lebih stabil saat memprediksi data dengan luas panen yang besar, seperti yang terjadi di Binjai Utara dan Binjai Timur. Namun demikian, terdapat pula beberapa data yang menunjukkan perbedaan cukup signifikan antara nilai prediksi dan aktual.

Misalnya, pada Kecamatan Binjai Selatan dengan luas panen sebesar 1.304 hektar, model memprediksi nilai yang lebih rendah dari aktualnya, sehingga menghasilkan error negatif sebesar -636,27.

Hal ini menunjukkan bahwa model melakukan underestimate terhadap nilai produksi sebenarnya. Selain itu, pada data dengan luas panen yang kecil, seperti di Kecamatan Binjai Barat dengan luas panen sebesar 58 hektar, diperoleh error sebesar -10,64. Meskipun nilai ini terlihat kecil secara absolut, namun secara proporsional dapat berdampak lebih besar terhadap tingkat kesalahan prediksi.

Walaupun demikian, secara keseluruhan error tersebut masih dalam batas yang dapat diterima. Dari hasil-hasil tersebut, dapat disimpulkan bahwa model regresi linear sederhana lebih stabil dan akurat dalam memprediksi produksi padi pada data dengan luas panen besar.

Sebaliknya, pada data dengan luas panen kecil, akurasi model cenderung menurun, meskipun secara keseluruhan error yang dihasilkan tidak dominan atau signifikan dalam mempengaruhi performa model secara umum. Model ini dapat dijadikan acuan awal dalam merancang strategi prediksi produksi padi yang lebih efektif di masa mendatang.



Gambar 11. Perbandingan nilai aktual dan prediksi

Dari Gambar 11 di atas memperlihatkan hubungan antara nilai aktual produksi padi (sumbu X) dan nilai prediksi hasil model regresi (sumbu Y) pada data uji. Titik-titik berwarna merah mewakili hasil prediksi dari model regresi linear sederhana, sedangkan garis biru putus-putus menunjukkan garis ideal di mana prediksi sama persis dengan nilai aktual ($y = x$).

Dari visualisasi tersebut, dapat dilihat bahwa sebagian besar titik prediksi berada sangat dekat dengan garis biru, yang menandakan bahwa hasil prediksi memiliki tingkat kesalahan (*error*) yang rendah. Artinya, model mampu memprediksi nilai produksi padi dengan cukup akurat. Titik yang berada tepat di garis menunjukkan prediksi yang sangat tepat, sedangkan titik yang sedikit menjauh menunjukkan adanya sedikit error, namun tetap dalam batas wajar.

Secara keseluruhan, kedekatan titik-titik terhadap garis diagonal mengindikasikan bahwa model memiliki performa prediktif yang baik pada data uji, sehingga cocok digunakan untuk memprediksi produksi padi berdasarkan variabel input (luas panen).

```
print("\nPrediksi Produksi Padi Tahun 2025 di Seluruh Kecamatan Kota Binjai :")
print(df_2025)
```

Prediksi Produksi Padi Tahun 2025 di Kota Binjai (pakai rata-rata luas panen 2020-2024):			
	Kecamatan	Luas_Panen_2025	Produksi_Prediksi_2025
0	Binjai Barat	49.4	291.110436
1	Binjai Selatan	1078.8	6544.265555
2	Binjai Timur	320.6	1938.531913
3	Binjai Utara	1372.2	8326.542329
4	TOTAL	2821.0	17100.450232

Gambar 12. Prediksi produksi padi di tahun 2025

Berdasarkan hasil implementasi regresi linear sederhana dengan menggunakan rata-rata luas panen periode 2020–2024, diperoleh prediksi produksi padi di Kota Binjai pada tahun 2025 sebesar 17.100 ton gabah kering panen (GKP). Prediksi tersebut diperoleh dari estimasi rata-rata luas panen sebesar 2.821 hektar. Jika dilihat lebih rinci per kecamatan, produksi padi pada tahun 2025 diprediksi sebagai berikut:

- Kecamatan Binjai Barat diperkirakan menghasilkan sekitar 291 ton dengan rata-rata luas panen 49,4 hektar.
- Kecamatan Binjai Selatan memiliki prediksi produksi sebesar 6.544 ton dengan rata-rata luas panen 1.078,8 hektar.
- Kecamatan Binjai Timur diperkirakan menghasilkan sekitar 1.939 ton dengan rata-rata luas panen 320,6 hektar.
- Kecamatan Binjai Utara memiliki prediksi produksi paling tinggi, yaitu sekitar 8.327 ton dengan rata-rata luas panen 1.372,2 hektar.

Dari hasil tersebut dapat diketahui bahwa Kecamatan Binjai Utara dan Binjai Selatan menjadi daerah penyumbang produksi padi terbesar di Kota Binjai, sementara Kecamatan Binjai Barat memiliki kontribusi yang paling kecil.

4. KESIMPULAN

Luas panen memiliki hubungan yang sangat kuat dengan produksi padi di Kota Binjai, yang ditunjukkan oleh nilai korelasi sebesar 1,00 sehingga mengindikasikan adanya hubungan linier yang hampir sempurna antara kedua variabel. Model regresi linear sederhana yang dibangun menghasilkan persamaan $Y = -8,97 + 6,07X$, yang berarti setiap kenaikan 1 hektar luas panen akan meningkatkan produksi padi sekitar 6,07. Selain itu, hasil evaluasi model menunjukkan kinerja yang sangat baik dengan nilai koefisien determinasi $R^2 = 0,9990$, yang berarti sebesar 99,90% variasi produksi dapat dijelaskan oleh variabel luas panen. Serta didukung oleh nilai kesalahan prediksi yang relatif rendah, yaitu MSE sebesar 5092,47, RMSE sebesar 71,36, dan MAE sebesar 56,87. Dengan demikian, secara keseluruhan regresi linear sederhana terbukti mampu memberikan prediksi produksi padi yang akurat di Kota Binjai berdasarkan data luas panen periode 2020–2024.

REFERENCES

- [1] H. U. A. Rezvi *et al.*, "Rice and food security: Climate change implications and the future prospects for nutritional security," *Food Energy Secur.*, 2022, doi: 10.1002/fes3.430.
- [2] M. N. Panggabean, R. Alamsyah, S. Nasution, and T. S. Alasi, "Sistem Informasi Kas Kecil Berbasis Microsoft Excel VBA Pada Puskesmas HAH Hasan Payaroba Binjai," *J. Tekno Log*, vol. 1, no. 1, 2024.
- [3] J. Damanik, T. S. Alasi, and V. J. Sianipar, "Smart City Untuk Kemudahan Pelayanan Publik Kota Binjai dengan Pendekatan Sistem Terdistribusi dan Kecerdasan Buatan," *JURNAL RAMBUTAN*, vol. 1, no. 1, pp. 46–52, 2025.
- [4] T. S. Alasi, E. M. Padilla, I. Taufik, and H. A. Tambunan, "Lomba Buah Rambutan Terbaik untuk Warga Kota Binjai dalam Upaya Budi Daya Buah Unggul dengan Metode SMART," *JURNAL RAMBUTAN*, vol. 1, no. 1, pp. 27–34, 2025.
- [5] T. S. Alasi and P. Fitriani, "e-Planning Menerapkan Algoritma Boyer Moore untuk Penyesuaian Usulan Berdasarkan Tupoksi Organisasi Perangkat Daerah," in *SINTAKS (Seminar Nasional Teknologi Informasi Komputer dan Sains 2019)*, 2019, pp. 128–137.
- [6] D. Wulandari and R. Rumini, "Pemodelan dan Prediksi Produksi Padi Menggunakan Regresi Linear," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 12, no. 4, 2023, doi: 10.30591/smartcomp.v12i4.5905.
- [7] R. B. Pickson, G. He, and E. Boateng, "Impacts of climate change on rice production: evidence from 30 Chinese provinces," *Environ. Dev. Sustain.*, vol. 24, pp. 3907–3925, 2021, doi: 10.1007/s10668-021-01594-8.
- [8] S. K. S. Durai and M. D. Shamili, "Smart farming using Machine Learning and Deep Learning techniques," *Decision Analytics Journal*, 2022, doi: 10.1016/j.dajour.2022.100041.
- [9] Runimeirati, Abdul Muis, and Figur Muhammad, "Pelatihan Text Mining Menggunakan Bahasa Pemrograman Python," *Abdimas Langkanae*, vol. 3, no. 1, pp. 36–46, 2023, doi: 10.53769/abdimas.3.1.2023.83.
- [10] W. T. Bhirawa, "Proses Pengolahan Data Dari Model Persamaan Regresi Dengan Menggunakan Statistical Product and Service Solution (SPSS)," *Statistika*, pp. 71–83, 2020.
- [11] M. Guo *et al.*, "Normal Workflow and Key Strategies for Data Cleaning Toward Real-World Data: Viewpoint," *Interact. J. Med. Res.*, vol. 12, 2022, doi: 10.2196/44310.
- [12] S. Hosseini and S. R. Sardo, "Data mining tools -a case study for network intrusion detection Content courtesy of Springer Nature , terms of use apply . Rights reserved . Content courtesy of Springer Nature , terms of use apply . Rights reserved .," pp. 4999–5019, 2021.
- [13] H. Nuha, "Mean Squared Error (MSE) dan Penggunaannya," *Papers.Ssrn.Com*, vol. 52, pp. 1–1, 2023, [Online]. Available: <https://ssrn.com/abstract=4420880>
- [14] F. I. Sanjaya and D. Heksaputra, "Prediksi Rerata Harga Beras Tingkat Grosir Indonesia dengan Long Short Term Memory," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 2, pp. 163–174, 2020, doi: 10.35957/jatisi.v7i2.388.
- [15] S. Robeson and C. Willmott, "Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," *PLoS One*, vol. 18, 2023, doi: 10.1371/journal.pone.0279774.
- [16] F. Kumayas, A. G. Kumenaung, and H. F. D. Siwu, "Pengaruh Jumlah Penduduk, Tingkat Pendidikan Dan Tingkat Pengangguran Terhadap Kemiskinan Di Kabupaten Minahasa," *Jurnal Berkala Ilmiah Efisiensi*, vol. 24, no. 4, pp. 71–89, 2024.

- [17] S. Asyuti and A. A. Setyawan, "Data Mining Dalam Penggunaan Presensi Karyawan Denga Cluster Means," *Jurnal Ilmiah Sains Teknologi Dan Informasi*, vol. 1, no. 1, pp. 01–10, 2023, [Online]. Available: <https://jurnal.alimspublishing.co.id/index.php/JITI/article/download/6/6>
- [18] A. S. Sembiring, T. S. Alasi, and others, "Penerapan Data Mining Menggunakan Algoritma Apriori Pada Peminjaman Buku di Perpustakaan Pada Pesantren Babul Ulum," *Jurnal Armada Informatika*, vol. 7, no. 2, pp. 323–327, 2023.
- [19] T. S. Alasi, "Ilmu komputer," 2024, *Media Publikasi Idpress*.
- [20] S. Y. Prayogi, T. S. Alasi, and R. F. Rahmat, "Pengantar Machine Learning," 2025, *Media Publikasi Idpress*.
- [21] T. S. Alasi, "Pelatihan Perancangan Media Pembelajaran Berbasis Macromedia Flash untuk Guru SMA Sumatera Utara," *Jurnal Pengabdian Masyarakat Variasi*, vol. 1, no. 1, pp. 5–8, 2024.
- [22] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning.," *Soc. Sci. Res.*, vol. 110, p. 102817, 2022, doi: 10.1016/j.ssresearch.2022.102817.
- [23] S. S. Lubis, S. Defit, and Sumijan, "Implementasi Algoritma K-Means Guna Pengelompokkan Data Penjualan Berdasarkan Pembelian," *Jurnal KomtekInfo*, 2024, doi: 10.35134/komtekinfo.v11i4.557.
- [24] N Kholifah, "Analisis Standar Pelayanan Minimal Pada Instalasi Rawat Jalan di RSUD Kota Semarang," *IAIN parepare*, vol. 3, p. 65, 2021.
- [25] T. Daniya, M. Geetha, B. Santhosh Kumar, and R. Cristin, "Least square estimation of parameters for linear regression," *International Journal of Control and Automation*, vol. 13, no. 2, pp. 447–452, 2020.
- [26] S. Holm, "Statistical evidence and algorithmic decision-making," *Synthese*, vol. 202, pp. 1–16, 2023, doi: 10.1007/s11229-023-04246-8.
- [27] C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," *Jurnal Portal Data*, vol. 2, no. 3, pp. 1–12, 2022, [Online]. Available: <http://portaldata.org/index.php/portaldata/article/view/107>
- [28] M. A. R. Siregar, "Peningkatan Produktivitas Tanaman Padi Melalui Penerapan Teknologi Pertanian Terkini," *Jurnal Agribisnis*, vol. 1, no. 1, pp. 1–11, 2023.
- [29] A. Dudáš, A. Michalíková, and R. Jašek, "Fuzzy Masks for Correlation Matrix Pruning," *IEEE Access*, vol. 13, pp. 35387–35400, 2025, doi: 10.1109/ACCESS.2025.3544027.