

Analisis Prediksi Produksi Tanaman Padi Berdasarkan Variabel Iklim Menggunakan Support Vector Regression

Sylvia ¹, Hendri Purnomo ², Rendi Estian Pambudi ³, Oki Arifin ⁴, Fathurrahman Kurniawan Ikhsan ⁴

¹ Politeknik Negeri Lampung

² IIB Darmajaya Bandar Lampung

³ IIB Darmajaya Bandar Lampung

⁴ Politeknik Negeri Lampung

⁵ Politeknik Negeri Lampung

Info Artikel

Riwayat Artikel:

Received May 26, 2025

Revised July 01, 2025

Accepted July 25, 2025

Corresponding Author:

Hendri Purnomo

Email: hendrialie@darmajaya.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract – Abstract in English

This study aims to analyze the prediction of rice crop production based on climate variables using the Support Vector Regression (SVR) algorithm. The dataset comprises data from several provinces in the Sumatra region, with variables including harvested area, rainfall, humidity, and average temperature. After preprocessing and data normalization, the SVR model was built using the Radial Basis Function (RBF) kernel with default parameters. Evaluation results showed a Root Mean Square Error (RMSE) of 970.144 and a coefficient of determination (R^2) of -0.065, indicating that the model failed to accurately capture the relationship between climate variables and rice production. Visualization comparing actual and predicted values revealed significant deviation. The study concludes that the SVR model with default parameters is not yet optimal for rice production prediction in Sumatra and requires parameter tuning and additional relevant features to improve performance.

Keywords: climate, prediction, rice production, Sumatra, Support Vector Regression.

Abstrak

Penelitian ini bertujuan untuk menganalisis prediksi produksi tanaman padi berdasarkan variabel iklim menggunakan algoritma Support Vector Regression (SVR). Data yang digunakan berasal dari beberapa provinsi di wilayah Sumatera dengan variabel yang meliputi luas panen, curah hujan, kelembapan udara, dan suhu rata-rata. Setelah dilakukan pra-pemrosesan dan normalisasi data, model SVR dibangun dengan kernel Radial Basis Function (RBF) dan parameter default. Hasil evaluasi menunjukkan nilai Root Mean Square Error (RMSE) sebesar 970.144 dan koefisien determinasi (R^2) sebesar -0.065, yang menunjukkan bahwa model belum mampu memodelkan hubungan antara variabel iklim dan produksi padi secara akurat. Visualisasi perbandingan antara nilai aktual dan prediksi mengindikasikan deviasi yang cukup signifikan. Penelitian ini menyimpulkan bahwa model SVR dengan parameter default belum optimal untuk prediksi produksi padi di wilayah Sumatera, sehingga perlu dilakukan tuning parameter dan penambahan fitur pendukung agar model dapat bekerja lebih efektif.

Kata Kunci: iklim, produksi padi, prediksi, Support Vector Regression, Sumatera

I. PENDAHULUAN

Pertanian merupakan sektor vital dalam perekonomian Indonesia, dengan padi sebagai komoditas utama yang menjadi sumber pangan pokok bagi sebagian besar penduduk. Namun, produksi padi seringkali dipengaruhi oleh variabilitas iklim, seperti curah hujan, suhu, dan kelembapan udara. Perubahan iklim yang tidak menentu dapat menyebabkan fluktuasi hasil panen, yang pada gilirannya mempengaruhi ketahanan pangan nasional. Oleh karena itu, diperlukan metode prediksi yang akurat untuk memperkirakan produksi padi berdasarkan variabel iklim [1] [2]

Support Vector Regression (SVR) merupakan salah satu metode dalam machine learning yang efektif untuk menangani masalah regresi non-linear. SVR memiliki kemampuan generalisasi yang baik dan telah digunakan dalam berbagai bidang, termasuk prediksi hasil pertanian. Dalam konteks produksi padi, SVR dapat memodelkan hubungan kompleks antara variabel iklim dan hasil panen, sehingga memungkinkan perencanaan yang lebih baik dalam sektor pertanian [3] [4].

Beberapa penelitian sebelumnya telah menerapkan SVR untuk memprediksi produksi padi. menggunakan SVR untuk meramalkan hasil panen padi di Kabupaten Malang dan memperoleh nilai Mean Absolute Percentage Error (MAPE) sebesar 10,133%, menunjukkan tingkat akurasi yang baik[5] . Namun, penelitian lain di Kabupaten Bogor menunjukkan bahwa regresi linier memiliki performa lebih baik dibandingkan SVR dalam memprediksi produksi padi, dengan nilai R^2 sebesar 1,000 untuk regresi linier dan -0,065 untuk SVR . Perbedaan hasil ini menunjukkan bahwa efektivitas SVR dapat bervariasi tergantung pada karakteristik data dan parameter yang digunakan. [6].

Selain itu, menerapkan SVR yang dioptimasi dengan Genetic Algorithm (GA) untuk meramalkan curah hujan sebagai pendukung kalender tanam padi di Bojonegoro. Hasil penelitian menunjukkan bahwa metode GA-SVR menghasilkan peramalan yang lebih akurat dibandingkan metode ARIMA dan SVR biasa, dengan nilai Root Mean Square Error (RMSE) dan Symmetric Mean Absolute Percentage Error (SMAPE) yang lebih kecil . Hal ini mengindikasikan bahwa pengoptimalan parameter SVR dapat meningkatkan akurasi prediksi[7]

Meskipun telah banyak penelitian yang menggunakan SVR untuk prediksi produksi padi, masih terdapat kebutuhan untuk mengeksplorasi penerapan metode ini pada data iklim di wilayah Sumatera, Indonesia. Wilayah ini memiliki karakteristik iklim dan kondisi pertanian yang unik, sehingga penting untuk mengkaji efektivitas SVR dalam konteks lokal. Dengan demikian, penelitian ini bertujuan untuk menganalisis prediksi produksi tanaman padi berdasarkan variabel iklim menggunakan Support Vector Regression, dengan fokus pada wilayah Sumatera[8][9].

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan model prediksi produksi padi yang akurat, yang dapat digunakan oleh petani, pengambil kebijakan, dan pihak terkait lainnya untuk merencanakan kegiatan pertanian secara lebih efektif. Dengan memanfaatkan data iklim dan metode SVR, diharapkan dapat meningkatkan ketahanan pangan dan kesejahteraan petani di Indonesia.[10][11]

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif-prediktif menggunakan algoritma Support Vector Regression (SVR). Tujuannya adalah untuk memodelkan dan memprediksi produksi tanaman padi berdasarkan variabel iklim di wilayah Sumatera, Indonesia[4][6].

Diagram berikut menyajikan representasi visual dari alur penelitian, mencakup tahapan dari identifikasi permasalahan hingga proses implementasi dan evaluasi[12]



Gambar 1 Alur Penelitian

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah *Dataset Tanaman Padi Sumatera* yang mencakup informasi dari beberapa provinsi di Sumatera dalam rentang waktu tertentu. Variabel yang tersedia meliputi [11]:

- Produksi padi (target variabel),
- Luas panen,
- Curah hujan,
- Kelembapan udara,
- Suhu rata-rata.

2. Pra-pemrosesan Data

Data mentah kemudian diproses untuk meningkatkan kualitas model:

- Menangani nilai hilang atau null,
- Normalisasi data menggunakan Min-Max Scaling atau Standard Scaler,
- Pengecekan outlier dan distribusi data.

3. Pemilihan Fitur

Dalam penelitian ini, fitur yang digunakan sebagai prediktor utama adalah:

- Curah hujan
- Kelembapan udara
- Suhu rata-rata

Fitur “luas panen” dapat diikutsertakan sebagai variabel tambahan dalam skenario tertentu untuk melihat pengaruhnya terhadap produksi.

4. Pemodelan dengan SVR

Model SVR dibangun menggunakan kernel Radial Basis Function (RBF), karena kernel ini mampu menangani hubungan non-linear antara variabel iklim dan hasil produksi. Penyesuaian parameter SVR seperti C, epsilon, dan gamma dilakukan menggunakan teknik Grid Search dan Cross Validation.

5. Evaluasi Model

Evaluasi dilakukan dengan membandingkan hasil prediksi dengan data aktual menggunakan metrik:

- Mean Absolute Error (MAE),

- Mean Squared Error (MSE),
- Root Mean Squared Error (RMSE),
- R-squared (R^2).

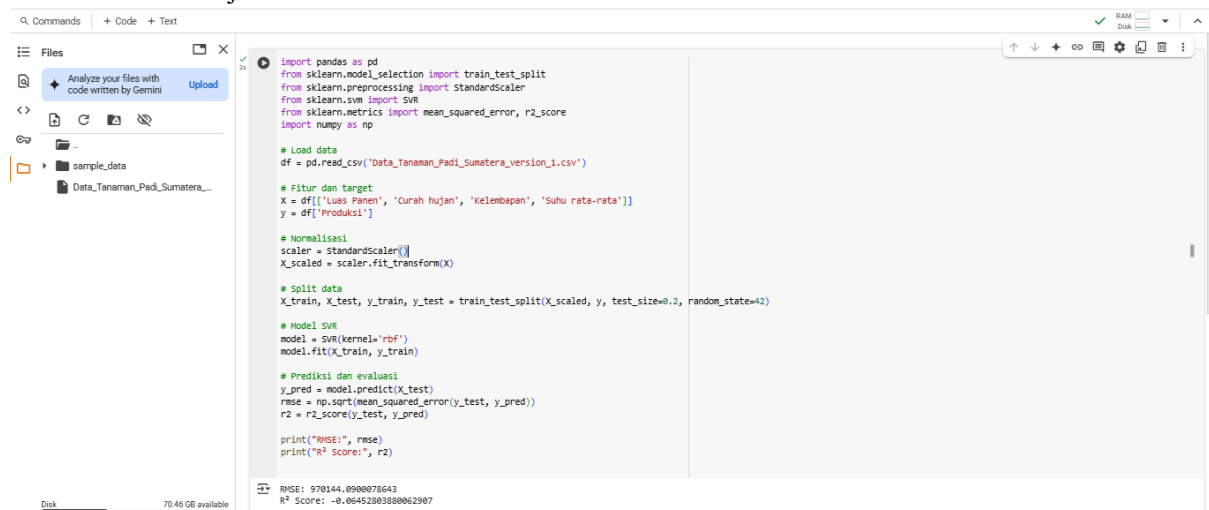
6. Interpretasi Hasil

Hasil prediksi dan evaluasi model dianalisis untuk memahami pengaruh masing-masing variabel iklim terhadap produksi padi. Hasil ini juga digunakan untuk memberikan rekomendasi kebijakan pertanian atau adaptasi iklim di masa depan.

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma *Support Vector Regression* (SVR) untuk melakukan prediksi terhadap produksi tanaman padi di wilayah Sumatera berdasarkan variabel-variabel iklim dan luas panen. Variabel input yang digunakan dalam pemodelan adalah *luas panen* (hektar), *curah hujan* (mm), *kelembapan udara* (%), dan *suhu rata-rata* ($^{\circ}\text{C}$). Sementara itu, variabel target yang diprediksi adalah *produksi padi* (ton).

Sebelum pemodelan, seluruh fitur numerik dilakukan normalisasi menggunakan metode *Standard Scaler* untuk memastikan setiap variabel memiliki skala yang setara. Selanjutnya, data dibagi menjadi dua bagian yaitu 80% data latih dan 20% data uji secara acak.



```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVR
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np

# Load data
df = pd.read_csv('Data_Tanaman_Padi_Sumatera_version_1.csv')

# Fitur dan target
X = df[['Luas Panen', 'Curah hujan', 'Kelembapan', 'Suhu rata-rata']]
y = df['Produksi']

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

# Split data
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# Model SVR
model = SVR(kernel='rbf')
model.fit(X_train, y_train)

# Prediksi dan evaluasi
y_pred = model.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

print("RMSE:", rmse)
print("R^2 Score:", r2)
```

RMSE: 970.144, 0.0645280388062907
R^2 Score: -0.0645280388062907

Gambar 2 proses Analisis SVR

Model SVR yang digunakan memiliki kernel default yaitu Radial Basis Function (RBF). Setelah proses pelatihan dan prediksi, hasil evaluasi performa model diperoleh sebagai berikut:

- Root Mean Square Error (RMSE): 970.144
- Koefisien Determinasi (R^2 Score): -0.065

Nilai RMSE yang cukup besar mengindikasikan bahwa terdapat selisih atau kesalahan prediksi yang signifikan antara nilai prediksi produksi dan nilai aktualnya. Sementara itu, nilai R^2 yang negatif menunjukkan bahwa model SVR yang digunakan tidak mampu menjelaskan variasi pada data target (produksi padi) dengan baik. Bahkan, performa model lebih buruk dibandingkan model dasar (baseline) yang hanya menggunakan rata-rata nilai produksi sebagai prediksi.

Hal ini mengindikasikan bahwa model SVR dalam bentuk dan konfigurasi saat ini belum mampu memodelkan hubungan antara variabel iklim dan luas panen terhadap produksi padi secara akurat. Beberapa faktor yang dapat menyebabkan rendahnya performa model antara lain:

1. Keterbatasan Fitur: Faktor-faktor lain yang berpengaruh terhadap produksi padi seperti jenis varietas, penggunaan pupuk, teknologi pertanian, serangan hama, dan kebijakan pemerintah tidak tersedia dalam dataset sehingga tidak dapat dimasukkan ke dalam model.
2. Kompleksitas Hubungan Data: Hubungan antara variabel iklim dan hasil produksi pertanian bisa sangat kompleks dan tidak selalu bersifat linier atau sederhana. Penggunaan parameter default pada SVR mungkin tidak cukup optimal untuk menangkap pola tersebut.

3. Variabilitas Spasial dan Temporal: Dataset mencakup beberapa provinsi dan kurun waktu yang panjang, tetapi model tidak diberi informasi eksplisit terkait perbedaan lokasi dan tahun, padahal faktor tersebut berpotensi sangat memengaruhi hasil produksi.
4. Kebutuhan Tuning Parameter: Model SVR sangat sensitif terhadap parameter seperti C, gamma, dan epsilon. Penggunaan nilai default tanpa tuning dapat menyebabkan performa yang tidak optimal.

Untuk melihat secara visual sejauh mana hasil prediksi mendekati nilai aktual, digunakan grafik perbandingan antara nilai produksi padi aktual dan hasil prediksi dari model Support Vector Regression (SVR). Visualisasi ini membantu dalam mengidentifikasi seberapa besar kesalahan prediksi yang terjadi pada setiap sampel data uji. Adapun script Python yang digunakan untuk menampilkan grafik tersebut adalah sebagai berikut:

```
import matplotlib.pyplot as plt

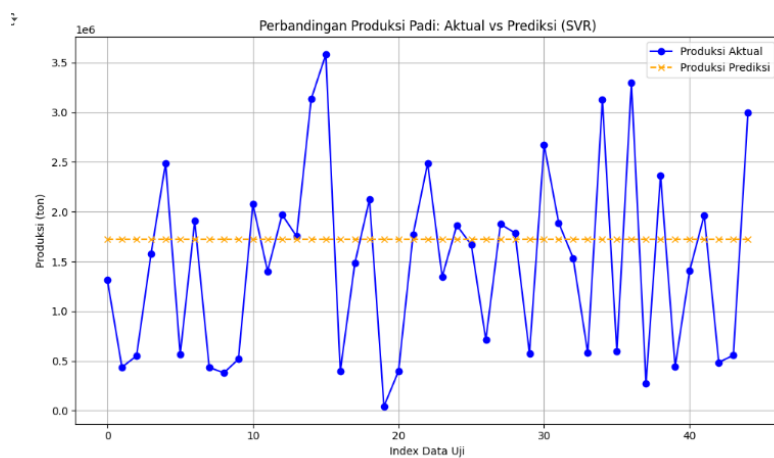
# Visualisasi: Prediksi vs Aktual
plt.figure(figsize=(10, 6))
plt.plot(range(len(y_test)), y_test.values, marker='o', label='Produksi Aktual', color='blue')
plt.plot(range(len(y_pred)), y_pred, marker='x', label='Produksi Prediksi', linestyle='--', color='orange')
plt.title('Perbandingan Produksi Padi: Aktual vs Prediksi (SVR)')
plt.xlabel('Index Data Uji')
plt.ylabel('Produksi (ton)')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

Gamabr 3 script Python

Script di atas menggunakan pustaka matplotlib.pyplot untuk membuat grafik. Dua garis ditampilkan:

- Garis pertama (warna biru) dengan marker bulat ('o') merepresentasikan data produksi padi aktual.
- Garis kedua (warna oranye) dengan marker silang ('x') dan garis putus-putus ('--') merepresentasikan hasil prediksi dari model SVR.

Selain evaluasi numerik menggunakan RMSE dan R^2 , visualisasi grafik juga dilakukan untuk melihat secara langsung perbandingan antara nilai produksi padi yang diprediksi oleh model dengan nilai produksi aktual pada data uji. Grafik ini ditampilkan pada Gambar 2 berikut:



Gambar 2 Grafik

Grafik di atas menunjukkan dua garis utama:

- Garis biru dengan titik bulat merepresentasikan nilai produksi aktual berdasarkan data uji.
- Garis oranye dengan tanda silang merepresentasikan nilai produksi prediksi yang dihasilkan oleh model SVR.

Dari grafik tersebut dapat diamati bahwa terdapat perbedaan yang cukup signifikan antara nilai aktual dan nilai prediksi pada beberapa titik data. Model belum mampu mengikuti pola fluktuasi data aktual dengan baik. Hal ini konsisten dengan hasil evaluasi kuantitatif sebelumnya yang menunjukkan nilai RMSE yang cukup tinggi dan R^2 yang bernilai negatif, yang berarti model tidak mampu menjelaskan variasi dalam data target secara optimal.

Grafik ini memberikan gambaran visual bahwa model masih perlu ditingkatkan, baik dari sisi tuning parameter maupun penambahan fitur yang lebih relevan untuk menghasilkan prediksi yang lebih akurat dan mendekati kenyataan.

IV. SIMPULAN

Penelitian ini telah menerapkan algoritma Support Vector Regression (SVR) untuk memprediksi produksi tanaman padi berdasarkan variabel iklim seperti curah hujan, kelembapan, suhu rata-rata, serta luas panen. Hasil evaluasi menunjukkan bahwa model menghasilkan nilai Root Mean Square Error (RMSE) sebesar 970.144 dan koefisien determinasi (R^2) sebesar -0.065, yang mengindikasikan bahwa model belum mampu memprediksi produksi padi dengan akurasi yang baik. Grafik perbandingan antara nilai aktual dan prediksi juga menunjukkan adanya deviasi signifikan, yang menandakan bahwa SVR dengan parameter default belum optimal dalam menangkap pola hubungan antar variabel dalam dataset yang digunakan.

DAFTAR PUSTAKA

- [1] D. P. L. Mardi Hardjianto, "Data Mining untuk Memprediksi Jumlah Penjualan Hasil Pertanian Menggunakan Algoritma Forecasting (Studi Kasus : Dinas Pertanian Kabupaten Banggai)," *Respati*, vol. 16, no. 2, p. 94, 2021, doi: 10.35842/jtir.v16i2.405.
- [2] M. F. Aziz, S. Defiyanti, and B. N. Sari, "Perbandingan Algoritma Cart Dan K-Nearest Neighbor Untuk Prediksi Luas Lahan Panen Tanaman Padi Di Kabupaten Karawang," *J. TAM (Technology Accept. Model.*, vol. 9, no. 2, pp. 74–78, 2018.
- [3] S. Rokhmah, A. Susilowati, and M. I. Permatasari, "Klasifikasi Data untuk Prediksi Hasil Panen Tanaman Padi di Wilayah Kabupaten Sukoharjo Menggunakan Algoritma C 45," *J. Rekayasa Teknol. Inf.*, vol. 6, no. 2, p. 134, 2022, doi: 10.30872/jurti.v6i2.8323.
- [4] A. A. Ilham, N. Harun, and Arwansyah, "Simulasi Penerapan Teknologi Data Mining Untuk Menghasilkan Model Pola Tanam Berkelanjutan," *Pros. Konf. Nas. Ilmu Komput. 2013*, pp. 124–132, 2013, [Online]. Available: <http://repository.unhas.ac.id/handle/123456789/9384>
- [5] D. A. Mardhika, B. D. Setiawan, and R. C. Wihandika, "Penerapan Algoritma Support Vector Regression Pada Peramalan Hasil Panen Padi Studi Kasus Kabupaten Malang," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 9402–9412, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [6] L. Febriyanti and H. Zakaria, "Implementasi Data Mining Untuk Memprediksi Produktivitas Pada Tanaman Kacang Tanah Menggunakan Metode Naive Bayes (Studi Kasus : Perkebunan Kacang Tanah Di Kota Bogor)," *Log. J. Ilmu Komput. dan Pendidik.*, vol. 1, no. 2, pp. 105–118, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- [7] E. A. Febriana Eka Dewi, Ari Kuntardina, "PENGARUH ONLINE CUSTOMER REVIEW DAN RATING TERHADAP KEPUTUSAN PEMBELIAN MELALUI MARKETPLACE SHOPEE (Studi Kasus Mahasiswa STIE CENDEKIA Bojonegoro angkatan 2017)," *JEMB J. Ekon. Manaj. dan Bisnis*, vol. 1, no. 2, pp. 90–99, 2022.
- [8] M. S. E. ,M. K. M. Y. S. K. ,M. K. Kandi Sri, "Penerapan Data Mining untuk Meprediksi Hasil Panen Pertanian Tanaman Padi Daerah Serdang Bedagai Menggunakan Metode C4.5 Pada Dinas Tanaman Pangan Dan Holtikultura Sumut," *J. Sains Manaj. Inform. dan Komput.*, vol. 1, no. 1, pp. 01–07, 2020.
- [9] R. Y. Rusdi, A. R. Tolangara, and H. Ahmad, "Jenis Tumbuhan Bertahan Hidup di Lahan Kering," *Techno J. Penelit.*, vol. 6, no. 02, p. 12, 2018, doi: 10.33387/tk.v6i02.564.
- [10] A. Bahtiar, "Jurnal Informatika Terpadu PREDIKSI HASIL PANEN PADI TAHUN 2023

- MENGGUNAKAN METODE REGRESI LINIER DI KABUPATEN INDRAMAYU,” *J. Inform. Terpadu*, vol. 9, no. 1, pp. 18–23, 2023, [Online]. Available: <https://journal.nurulfikri.ac.id/index.php/JIT>
- [11] S. Maesaroh and Kusriani, “Sistem Prediksi Produktifitas Pertanian Padi Menggunakan Data Mining,” *J. Energi*, vol. 7, no. 2, pp. 25–30, 2017, [Online]. Available: eprints.dinus.ac.id/16925/1/jurnal_16115.pdf
- [12] H. Panen, P. Di, and P. Jawa, “PERBANDINGAN METODE LINEAR REGRESSION , RANDOM FOREST & K-NEAREST NEIGHBOR UNTUK PREDIKSI PRODUKSI,” vol. 8, no. 3, pp. 3895–3900, 2024.