

# Penerapan Algoritma *Linear Regression* dan *Support Vector Regression* dalam Prediksi Temperatur Udara di Malang

*Application of Linear Regression and Support Vector Regression Algorithms in Air Temperature Prediction in Malang*

Karnisih<sup>1</sup>, Sunarno<sup>2</sup>, Iqbal<sup>3</sup>, Djuniadi<sup>4</sup>, Feddy Setio Pribadi<sup>5</sup>

<sup>1,2</sup> Fisika/ Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang

<sup>3</sup> Badan Meteorologi Klimatologi dan Geofisika

<sup>4,5</sup> Teknik Elektro, Universitas Negeri Semarang

E-mail: <sup>1</sup>karniwigianto@students.unnes.ac.id, <sup>2</sup>narnophysics@mail.unnes.ac.id,

<sup>3</sup>iqbal@bmgk.go.id, <sup>4</sup>djuniadi@mail.unnes.ac.id, <sup>5</sup>feddy.setyo@mail.unnes.ac.id

## Abstrak

Perubahan iklim global dan peningkatan variabilitas cuaca membuat prediksi temperatur udara menjadi salah satu kebutuhan penting di berbagai sektor. Temperatur udara merupakan parameter penting dalam meteorologi yang mempengaruhi berbagai aspek kehidupan manusia. Predisi temperatur udara saat ini banyak memanfaatkan algoritma *machine learning*, namun nilai akurasi masih belum optimal. Tujuan dari penelitian ini untuk meningkatkan akurasi prediksi temperatur udara rata-rata dengan menggunakan pendekatan berbasis *machine learning*. Metode dalam penelitian ini menggunakan algoritma *Linear Regression* dan *Support Vector Regression* (linier dan gaussian non linear) karena memiliki akurasi prediksi data yang cukup baik di berbagai bidang termasuk bidang hidrologi. Penelitian ini menggunakan data dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) lokasi Stasiun Klimatologi Jawa Timur periode data tahun 2019-2023 dengan parameter cuaca temperatur rata-rata (TAV), kelembaban udara (HAV), kecepatan angin (WAV), curah hujan (RR), tekanan udara (PPP), Penyinaran matahari (SUN) dan titik embun (DEW\_POINT). Kinerja model dievaluasi menggunakan pengukuran metrik MSE, RMSE, MAE, MAPE dan R<sup>2</sup>. Hasil pengukuran kinerja model algoritma Gaussian *support vector Regression* (non linier SVR) lebih baik dibanding dengan *linear support vector Regression* (linear SVR) dan algoritma *linear regression* dengan nilai yang lebih tinggi R<sup>2</sup> sebesar  $0,9891 \pm 0,0011$  dan nilai error yang lebih rendah pada semua metrik pengukuran.

*Kata kunci: Prediksi temperatur udara, machine learning, Linear Regression, Suport Vektor Regression*

## Abstract

*Global climate change and the increasing variability of weather patterns have made air temperature prediction an essential requirement across various sectors. Air temperature is a critical meteorological parameter that influences numerous aspects of human life. Current air temperature predictions often utilize machine learning algorithms; however, their accuracy remains suboptimal. The aim of this study is to improve the accuracy of average air temperature predictions using a machine learning-based approach. The methods in this study involve employing Linear Regression and Support Vector Regression (SVR) algorithms (both linear and Gaussian non-linear) due to their robust predictive accuracy across various fields, including hydrology. The study utilizes data from the Meteorological, Climatological, and Geophysical Agency (BMKG) at the East Java Climatology Station, covering the period from 2019 to 2023. The dataset includes weather parameters such as average temperature (TAV), humidity (HAV), wind speed (WAV), rainfall (RR), air pressure (PPP), solar radiation (SUN), and dew point (DEW\_POINT). The model's performance was evaluated using various metrics, including MSE, RMSE, MAE, MAPE, and R<sup>2</sup>. The results show that the Gaussian Support Vector Regression (non-linear SVR) algorithm outperformed both Linear Support Vector Regression (linear SVR) and the*

*Linear Regression algorithm. The Gaussian SVR achieved a higher  $R^2$  value of  $0.9891 \pm 0.0011$  and lower error values across all evaluation metrics.*

*Keywords: Air temperature prediction, machine learning, Linear Regression, Support Vector Regression.*

## 1. PENDAHULUAN

Perubahan iklim global dan peningkatan variabilitas cuaca membuat prediksi temperatur udara menjadi salah satu kebutuhan penting di berbagai sektor. Kota Malang, sebagai wilayah dengan karakteristik geografis yang kompleks, menghadapi tantangan dalam menghasilkan prediksi temperatur yang akurat. Temperatur udara merupakan parameter penting dalam meteorologi yang mempengaruhi berbagai aspek kehidupan manusia. Faktor-faktor seperti tekanan udara, kelembapan, dan intensitas radiasi matahari berperan dalam perubahan temperatur yang sering kali sulit diprediksi [1], [2].

Metode tradisional seperti analisis statistik sederhana memiliki keterbatasan dalam menangkap hubungan kompleks antar variabel cuaca. Oleh karena itu, pendekatan berbasis machine learning, seperti *Linear Regression* (LR) dan *Support Vector Regression* (SVR) menjadi alternatif yang menjanjikan karena mampu menangani data yang besar dan variabel yang saling bergantung [1], [2].

Sutaryani, dkk. menunjukkan bahwa prediksi temperatur udara menggunakan beberapa variabel cuaca dengan metode regresi linear mempunyai performa yang sangat akurat di Kota Semarang dengan nilai MAE = 0,099 dan nilai  $R^2 = 0,990$  [3]. Penelitian ini menunjukkan bahwa algoritma regresi linier mampu memprediksi temperatur udara dengan sangat akurat menggunakan 7 variabel cuaca, meskipun belum dibahas dengan lebih detail terkait pengaruh arah dan kekuatan hubungan linier antara variabel dengan variabel prediktor terhadap akurasi prediksi model. Selain itu penelitian ini belum cukup menjelaskan tentang alasan penggunaan algoritma decision tree sebagai pembandingan algoritma *linear regression*.

Penelitian lain terkait dengan penggunaan algoritma *linear regression* diantaranya adalah Yulistiani, dkk. yang menunjukkan bahwa prediksi produktivitas padi menggunakan data cuaca di Indonesia menunjukkan tingkat galat *Root Mean Square Error* (RMSE) bernilai 26.4763 dan *Mean Absolute Error* (MAE) bernilai 21.7043 [4]. Selain itu juga ada Limehuwey, dkk pada prediksi suhu harian yang menghasilkan RMSE dengan nilai  $1.293 \pm 0.0000$  yang berarti bahwa kesalahan rata-rata prediksi sekitar 1.293 [5].

Prasetyo, dkk. Dan Rahmabudhi menjelaskan bahwa terdapat perbedaan karakteristik hubungan linier antara suhu udara dengan kelembapan udara dan curah hujan di beberapa wilayah di Pulau Jawa [1], [2]. Hal ini menandakan bahwa arah dan kekuatan hubungan linier antara temperatur udara dengan variabel cuaca lainnya dapat berbeda-beda di berbagai wilayah di Pulau Jawa.

Penelitian sebelumnya terkait prediksi temperatur udara menggunakan algoritma regresi linier memiliki nilai akurasi yang sangat baik, namun performa tersebut belum dapat dipastikan apakah akan sama baiknya jika diterapkan di wilayah lain dengan variabel prediktor yang sama seperti pada penelitian Sutaryani, dkk karena terdapat perbedaan karakteristik hubungan linier temperatur udara dengan variabel cuaca lain [3]. Dengan demikian, perlu dilakukan langkah pengujian untuk memilih fitur-fitur yang paling sesuai sebagai variabel prediktor pada model regresi linier untuk melakukan prediksi temperatur udara sehingga akurasi model yang didapat lebih maksimal.

Selain itu, untuk mengatasi kemungkinan pengaruh penurunan akurasi prediksi temperatur udara pada model regresi linier akibat perbedaan karakteristik temperatur udara dengan variabel cuaca lainnya disuatu wilayah perlu dilakukan pengujian prediksi temperatur udara menggunakan algoritma machine learning yang lebih sesuai [4].

Penelitian yang dilakukan oleh Siregar mengenai prediksi curah hujan di Kota Medan, ditemukan bahwa metode Support Vector Regression (SVR) sangat efektif dalam memprediksi curah hujan. Efektivitas ini ditunjukkan melalui pengujian nilai Root Mean Square Error (RMSE)

pada parameter-parameter model SVR. Hasilnya menunjukkan nilai RMSE adalah 0,038800637 dengan gamma bernilai 0,0005, C bernilai 0,0001, dan nilai epsilon 1.[6].

Lesnusa, dkk. menunjukkan bahwa hasil evaluasi performa prediksi suhu udara di kota Palu dengan model *Support Vector Regression* (SVR) memiliki RMSE sebesar 0.6302, sedangkan *linear regression* memiliki RMSE adalah 0.6328 [7].

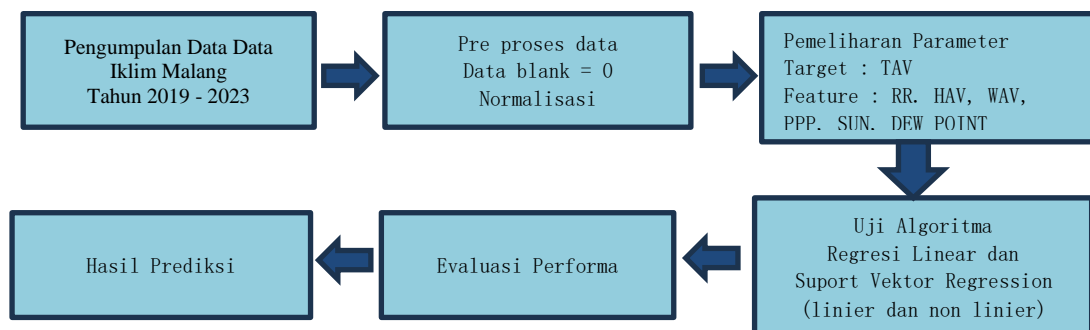
Zahroh, dkk. menunjukkan bahwa metode yang digunakan untuk prediksi kecepatan angin di kabupaten Kupang adalah *Support Vector Regression* (SVR) dengan melakukan pengujian terhadap tiga jenis kernel yaitu linear, polynomial, gaussian. Berdasarkan model SVR terbaik diperoleh dengan kernel gaussian yang menunjukkan nilai MAPE sebesar 20,703 dengan epsilon 0,010 dan bias 0,284.[8].

Berdasarkan penelitian sebelumnya, algoritma SVR (linier dan gaussian non linear) memiliki akurasi yang cukup baik untuk mengatasi kasus prediksi data-data meteorologi seperti curah hujan, suhu udara, dan kecepatan angin [6], [7]. SVR memiliki dua pilihan kernel yaitu kernel linier dan non-linier sehingga dapat digunakan untuk menentukan hubungan antar data baik secara linier maupun non-linier [8]. Dengan demikian, algoritma SVR dapat digunakan sebagai metode usulan untuk mengatasi kemungkinan pengaruh penurunan akurasi prediksi temperatur udara pada model *linear regression* akibat perbedaan karakteristik temperatur udara dengan variabel cuaca lainnya di suatu wilayah.

Penelitian ini memiliki tujuan mengetahui fitur-fitur yang paling sesuai sebagai variabel prediktor untuk data temperatur udara rata-rata pada model *linear regression* sehingga mendapat performa akurasi prediksi temperatur udara rata-rata model *linear regression* yang maksimal dan mengetahui model prediksi temperatur udara rata-rata berbasis *machine learning* yang sesuai untuk mengatasi kemungkinan penurunan tingkat akurasi prediksi temperatur udara model *linear regression* yang disebabkan oleh variabilitas karakteristik hubungan linier temperatur udara rata-rata dengan variabel cuaca di suatu wilayah. Algoritma machine learning SVR (*linier dan gaussian non linear*) merupakan metode yang diusulkan sebagai pembandingan metode *linear regression* dalam prediksi temperatur udara rata-rata di Malang.

## 2. METODE PENELITIAN

Rancangan penerapan algoritma *machine learning* dalam penelitian ini meliputi beberapa hal yaitu mengumpulkan data, melakukan Pre proses data, pemilihan parameter yang akan digunakan, dan melakukan pembagian data dengan komposisi 80 % sebagai data pelatihan dan 20 % sebagai data uji, selanjutnya dilakukan pengujian menggunakan metode *machine learning Linear Regression* dan *Support Vector Regression* (linier dan gaussian non linear) dengan evaluasi performa model menggunakan pengukuran metrik *R-square*  $R^2$ , *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percent Error* (MAPE). Berikut desain penelitian ditunjukkan pada gambar 1.



Gambar 1 Desain Penelitian

### 2.1 Pengumpulan Data

Dataset dalam penelitian menggunakan data iklim yang terdiri dari data Temperatur Udara rata-rata (TAV), Kelembaban Udara (HAV), Curah hujan (RR), Tekanan Udara (PPP), Penyinaran matahari (SUN), Titik embun (DEW\_POINT) dan Kecepatan angin (WAV) yang bersumber dari data BMKG Pusat untuk lokasi di Stasiun Klimatologi Jawa Timur Malang. Dataset yang digunakan sebagai target adalah temperatur udara rata-rata (TAV) sedangkan variabel yang akan digunakan untuk memprediksi nilai target atau variabel prediktor adalah data HAV, RR, PPP, SUN, DEW\_POINT, WAV. Dataset sebelum dilakukan preprocessing berjumlah 1826 row. Dataset dapat ditunjukkan pada Tabel 1.

Tabel 1 Dataset 2019-2023

NO	DEW_POINT	TAV	HAV	RR	SUN	WAV	PPP
1	21.5	25.0	81		7.2	0.6	1011.1
2	21.2	25.2	79	18.5	3.8	1.6	1011.2
3	20.8	22.7	89	0.8	4.4	0.5	1013.4
4	20.7	24.5	80	8.7	0	1.3	1012.9
5	21.9	26.0	78	0	6.8	2.2	1013.0
6	21.7	25.0	82	0	9.4	0.8	1013.2
7	21.8	26.4	76	0	3.7	1.1	1012.0
8	20.9	26.4	72	0	7.4	1.7	1012.0
9	21.3	27.1	71	0	7.6	2	1011.1
10	22.4	26.5	78	0	10.9	1.7	1011.2
11	21.5	25.1	81	0.1	6.8	1.4	1011.0
12	21.5	24.5	83	22.2	6.3	1.1	1011.5
13	21.5	26.1	76	41.5	5.5	1.6	1010.3
14	21.3	25.0	80	0.4	9.2	1.1	1009.7
15	22.0	25.7	80	6.7	7.1	1.5	1008.8
16	21.9	23.8	89	3.3	5.5	0.9	1009.5
17	21.8	23.9	88	1.7	0	0.6	1009.7
18	22.1	25.3	82	29.5	2	1	1008.9
19	22.4	24.6	87	1.6	4.6	0.6	1010.0
20	22.1	25.1	83	2.5	1.8	0.7	1010.1
21	22.2	23.9	90	41.6	6.2	1.1	1010.5
22	.....	.....	.....	.....	.....	.....	.....
23	.....	.....	.....	.....	.....	.....	.....
24	.....	.....	.....	.....	.....	.....	.....
1824	23.2	27.1	79	27.5	5.4	1.9	1011.2
1825	22.8	26.1	82	1.4	8.7	1.5	1010.7
1826	23.1	26.6	81	34	4	1.8	1011.2

### 2.2 Pre processing Data

Preprocessing data merupakan tahapan penting dalam menyiapkan dataset untuk menghasilkan analisis yang valid dalam memprediksi temperatur udara rata-rata pada dataset yang besar. Tahapan ini meliputi pembersihan data untuk menghilangkan informasi yang tidak relevan, penanganan data yang hilang dengan melakukan penghapusan pada data yang hilang, normalisasi data, serta pembagian data guna evaluasi yang akurat. Teknik seperti imputasi untuk nilai yang hilang dan deteksi outlier meningkatkan kualitas dataset, membantu mengurangi overfitting dan mengoptimalkan kinerja model [9].

Adapun untuk penanganan data yang hilang dengan melakukan penghapusan pada data yang hilang dapat ditunjukkan pada Gambar 2.

```
[ ] # creating a dataframe from dictionary
data = pd.DataFrame(data)
# Menghapus Baris yang mengandung nilai Null
data.dropna()
```

	DEW_POINT	TAV	HAV	RR	SUN	WAV	PPP
1	21.2	25.2	78.7	18.5	3.8	1.6	1011.2
2	20.8	22.7	88.8	0.8	4.4	0.5	1013.4
3	20.7	24.5	79.6	8.7	0.0	1.3	1012.9
4	21.9	26.0	78.2	0.0	6.8	2.2	1013.0
5	21.7	25.0	82.3	0.0	9.4	0.8	1013.2

Gambar 2 Penanganan data Hilang

Sedangkan normalisasi data yang merupakan proses untuk menangani perbedaan rentang nilai data dapat ditunjukkan pada Gambar 3.

```
Hasil Standarisasi (5 baris pertama):
```

	DEW_POINT	TAV	HAV	RR	SUN	WAV	PPP
0	0.856115	0.463768	0.731579	0.127586	0.336283	0.245614	0.496063
1	0.827338	0.101449	0.908772	0.005517	0.389381	0.052632	0.669291
2	0.820144	0.362319	0.747368	0.060000	0.000000	0.192982	0.629921
3	0.906475	0.579710	0.722807	0.000000	0.601770	0.350877	0.637795
4	0.892086	0.434783	0.794737	0.000000	0.831858	0.105263	0.653543

Gambar 3 Normalisasi Data

### 2.3 Pemilihan Parameter

Setelah proses pre prosesing data selesai selanjutnya dilakukan pemilihan data untuk menentukan data mana yang akan menjadi target dan data mana yang akan menjadi Fitur. Dalam penelitian ini untuk target digunakan data Temperatur Udara rata-rata (TAV) sedangkan fitur digunakan data Curah hujan (RRR), Kelambaban udara (HAV), Tekanan Udara (PPP), Penyinaran matahari (SUN), Dew Poin(Dew\_Point), Kecepatang Angin (WAV).

### 2.4 Modelling

Pada penelitian ini menggunakan Algoritma *Linear Regression* dan *Support Vector Regression* (linier dan gaussian non linear), dengan bahasa Pemograman Phyton dan memanfaatkan aplikasi *Google Colab* untuk melakukan pengolahan pada dataset tersebut

#### 2.4.1 Algoritma *Linear Regression*

Analisis regresi adalah bersifat asimetris atau dua arah. Secara umum, analisis regresi merupakan studi mengenai hubungan antara satu variabel, yang disebut sebagai variabel terikat, dengan satu atau lebih variabel bebas yang menjelaskan hubungan tersebut. Variabel terikat dikenal juga sebagai variabel respon, sedangkan variabel bebas sering disebut sebagai variabel prediktor [10]. Regresi linier adalah metode yang digunakan untuk memperkirakan nilai variabel terikat berdasarkan nilai variabel bebas. Metode ini menggunakan persamaan linear yang menghubungkan variabel bebas dengan variabel terikat. Persamaan regresi linear [11] dapat dinyatakan sebagai berikut:

$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k$  (1)  
 dimana,  $B_0$  adalah konstanta,  $B$  adalah koefisien variabel  $X_1, X_2, \dots, X_k$ ,  $Y$  adalah variabel yang diduga (variabel dependen),  $X_i$  adalah variabel penduga (variabel independen)

Model penduganya adalah

$$Y = b + b_1X_1 + b_2X_2 + \dots + b_kX_k$$
 (2)

#### 2.4.2 Algoritma Support Vector Regression

Support Vector Machine (SVM) ditemukan oleh Vapnik pada tahun 1979 dimana SVM mencakup Super Vector Classification (SVC) dan Support Vector Regression (SVR). Pada penelitian ini menggunakan SVR. Algoritma SVR akan mendapatkan fungsi  $f(x)$  sebagai hyperplane (garis pemisah) berupa fungsi regresi yang sesuai dengan data input [12]. Metode SVR dapat fungsi symmetrical loss [13]. Besar nilai  $\epsilon$ -insensitive dapat menentukan toleransi kesalahan mengatasi permasalahan data linear maupun non linear menggunakan  $\epsilon$ -insensitive sebagai nilai prediksi dan nilai keluaran sebenarnya yang dapat berpengaruh pada jumlah support vector [14]. Metode SVR dapat dirumuskan menggunakan Persamaan berikut: [8].

$$f(x) = w^T \varphi(x) + b$$
 (3)

dimana  $w$  adalah dimensi vektor bobot,  $b$  adalah bias, dan  $\varphi(x)$  adalah titik di ruang fitur pada dimensi yang lebih tinggi berdasarkan hasil pemetaan input vektor di ruang input yang berdimensi rendah [15]. Model SVR dengan permasalahan non linear dirumuskan pada Persamaan berikut [16].

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(X_i, X_j) + b$$
 (4)

dimana  $\alpha_i - \alpha_i^*$  adalah pengali Lagrange,  $k$  adalah fungsi kernel,  $(X_i, X_j)$  adalah input vektor ke- $i$  dan ke- $j$ ,  $b$  adalah bias.

Fungsi kernel memiliki peran penting dalam algoritma SVR [17]. Tujuan dari fungsi kernel yaitu memetakan data input ke dalam ruang fitur yang memiliki dimensi lebih tinggi melalui  $\varphi$  [18]. Rumus fungsi kernel dapat dilihat dibawah ini:

Linear	$x \cdot x^T$
Gaussian	$\exp(-\frac{1}{2\sigma^2} \ x - x_i\ ^2)$

Dimana parameter  $d > 0$  adalah konstanta,  $\sigma$  adalah nilai parameter dari SVR, dan  $\|x - x_i\|$  adalah jarak euclidean [19].

#### 2.5 Evaluasi Kinerja Algoritma Machine Learning

Hasil pengolahan dengan kedua metode tersebut perlu dilakukan pengujian dengan melakukan evaluasi kinerja model menggunakan metrik tertentu seperti  $R$ -square  $R^2$ , Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE) merupakan ukuran yang menunjukkan pengaruh antara nilai sebenarnya terhadap nilai prediksi dalam suatu model [20]. Dengan rumus:

$$R^2 = 1 - \frac{\sum_{i=1}^n (z_i - z'_i)^2}{\sum_{i=1}^n (z_i - \bar{z})^2}$$
 (5)

dimana,  $R^2$  adalah rata – rata dari data aktual,  $z_i$  adalah data aktual,  $z'_i$  adalah data prediksi,  $\hat{z}$  adalah rata-rata dari data aktual,  $n$  adalah jumlah data.

*Mean Square Error* (MSE) adalah ukuran statistik yang menghitung rata-rata kuadrat kesalahan antara nilai yang diprediksi dan nilai yang diamati. MSE digunakan untuk mengevaluasi seberapa baik model prediktif dalam memprediksi nilai-nilai aktual. MSE sering digunakan sebagai kriteria untuk menyetel parameter [17].

$$MSE = \frac{1}{n} \sum_{i=1}^n |x_i - x'_i|^2 \quad (6)$$

Dimana  $x_i$  data sebenarnya,  $x'_i$  data prediksi,  $n$  adalah Jumlah data.

*Root Mean Square Error* (RMSE) adalah evaluasi keakuratan prediksi memerlukan perbandingan data aktual. Beberapa metode umum digunakan untuk memperoleh nilai kesalahan prediksi, diantaranya adalah RMSE. Metode ini merupakan metode populer pada penilaian hasil prediksi. RMSE menghitung selisih nilai yang diprediksi dan nilai yang sebenarnya terjadi oleh suatu model [21]. Nilai RMSE yang rendah, mendekati 0, menunjukkan tingkat akurasi yang tinggi [22]. RMSE merupakan total dari selisih kuadrat atau perbedaan dari nilai aktual dengan nilai prediksi yang dinyatakan pada persamaan:

$$RMSE = \sqrt{\frac{\sum_{j=i}^n (y^i - y)^2}{n}} \quad (7)$$

dimana,  $n$  adalah Error rate,  $i$  adalah jumlah keseluruhan data,  $y_i$  adalah nilai keluaran atau output (prediksi),  $y$  adalah nilai aktual atau sebenarnya

*Mean Absolute Error* (MAE) merupakan teknik yang sering diaplikasikan dalam menilai keakuratan model prediksi. MAE menghitung nilai dari rerata kesalahan mutlak antara nilai sebenarnya dengan prediksi. Tingkat keakuratan prediksi yang dihasilkan dari metode MAE dimanfaatkan untuk mengukur tingkat keakuratan suatu model prediksi [23]. MAE merupakan hasil perhitungan dari rerata perbedaan mutlak nilai aktual terhadap nilai prediksi dinyatakan pada persamaan:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (8)$$

dimana,  $f_i$  adalah nilai hasil prediksi,  $y_i$  adalah nilai sebenarnya, dan  $n$  adalah jumlah data.

*Mean Absolute Percent Error* (MAPE) adalah ukuran tingkat keakuratan pada sebuah prediksi. Nilai MAPE dihitung berdasarkan rata-rata absolute dari persentase error [24]. Rumus perhitungan dapat dinyatakan pada persamaan berikut:[25].

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{Y_i - P_i}{Y_i} \right| \times 100\%}{n} \quad (9)$$

Dimana  $Y_i$  nilai aktual ke- $i$ ,  $P_i$  nilai prediksi ke- $i$ , dan  $n$  adalah jumlah pengamatan.

### 3. HASIL DAN PEMBAHASAN

Penerapan algoritma *Linear Regression* dan *Support Vector Regression* (linier dan gaussian non linear) dengan menggunakan dataset sebanyak 1755 baris. Selanjutnya membagi menjadi 1404 data atau 80 % sebagai data training dan sebanyak 351 data atau 20 % sebagai data uji.

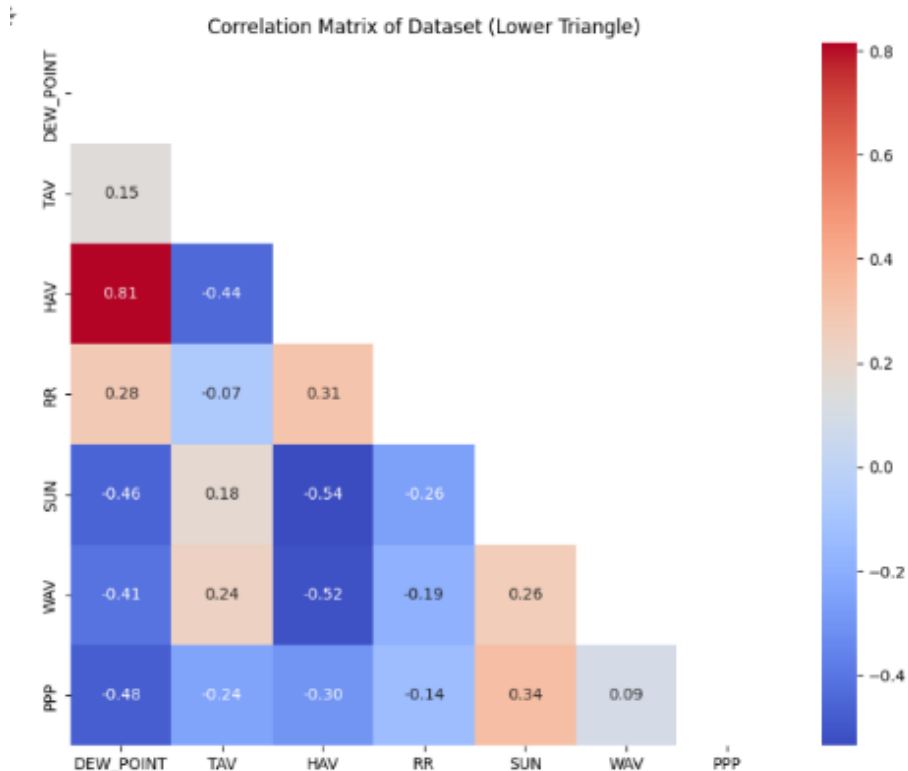
### 3.1 Penerapan Machine Learning Algoritma Linear Regression

Hasil pengujian yang dilakukan adalah merupakan pengujian terhadap fitur untuk Algoritma Linear Regression. Fitur- fitur yang paling sesuai untuk digunakan di model LR adalah (RR, WAV, HAV, SUN, PPP, DEW\_POINT) dengan nilai error paling kecil dan R<sup>2</sup> paling besar 98 %, sehingga semua parameter atau variable yang tersedia akan di gunakan sebagai fitur input pada model Linear Regression. Pengujian parameter dapat dtunjukkan pada Tabel 2.

Tabel 2 Pengujian Parameter algoritma Linear Regression

NO	Parameter Berpengaruh	MSE	RMSE	MAE	MAPE	R <sup>2</sup>
1	RR, WAV, HAV, SUN, PPP, DEW_POINT (Semua Parameter)	<b>0,016</b>	<b>0,126</b>	<b>0,091</b>	<b>0,004</b>	<b>0,985</b>
2	RR, WAV, HAV, SUN, PPP	0,701	0,837	0,655	0,026	0,339
3	RR, WAV, SUN, PPP, DEW_POINT	0,802	0,896	0,723	0,029	0,242
4	RR, HAV, SUN, PPP, DEW_POINT	<b>0,016</b>	0,127	<b>0,091</b>	<b>0,004</b>	<b>0,985</b>
5	RR, WAV, HAV, PPP, DEW_POINT	<b>0,016</b>	0,128	0,092	<b>0,004</b>	<b>0,985</b>
6	RR, WAV, HAV, SUN, DEW_POINT	<b>0,016</b>	0,127	<b>0,091</b>	<b>0,004</b>	<b>0,985</b>
7	WAV, HAV, SUN, PPP, DEW_POINT	<b>0,016</b>	0,128	<b>0,091</b>	<b>0,004</b>	<b>0,985</b>

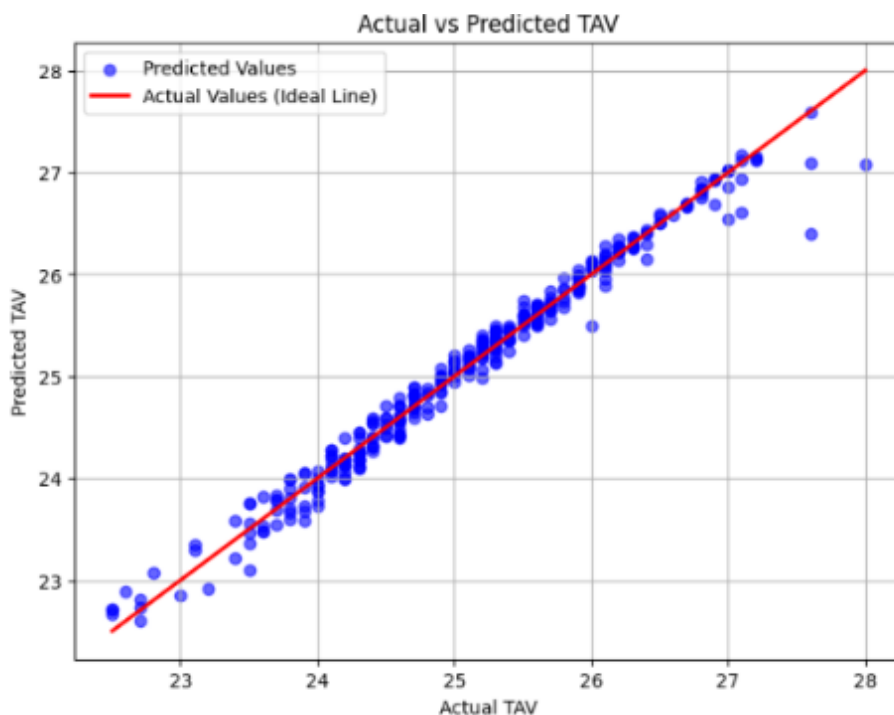
Adapun hasil korelasi antara Fitur DEW\_POINT, HAV, RR, SUN, WAV, PPP terhadap target TAV dapat di uraikan yaitu, DEW\_POINT sebesar 0,15, HAV sebesar -0,44, RR sebesar -0,07, SUN sebesar 0,18, WAVsebesar 0,24, dan PPP sebesar -0,24. Korelasi fitur terhadap TAV tersebut dapat di tunjukan pada Gambar 4.



Gambar 4 Korelasi Fitur terhadap Target



Hasil prediksi dengan algoritma *Linear Regression* untuk prediksi temperatur udara data-rata ditunjukkan Gambar 5.



Gambar 5 Hasil Prediksi Temperatur Udara Rata-rata dengan *Linear Regression*

### 3.2 Perbandingan Penerapan Algoritma *Machine Learning Linear Regression* dan *Support Vektor Regression* (linier dan gaussian non linier)

Pengujian dilakukan secara berulang menggunakan Algoritma *Machine Learning Linear Regression* menunjukkan hasil dimana MSE bernilai  $0,0204 \pm 0,0039$ , RMSE bernilai  $0,1425 \pm 0,0139$ , MAE bernilai  $0,0972 \pm 0,005$ , MAPE bernilai  $0,0039 \pm 0,0002$  dan  $R^2$  bernilai  $0,9810 \pm 0,0033$ . Algoritma *Support Vector Regression* (linear SVR) dengan nilai MSE  $0,0257 \pm 0,0027$ , RMSE bernilai  $0,1603 \pm 0,0086$ , MAE bernilai  $0,0962 \pm 0,0023$ , MAPE bernilai  $0,0038 \pm 0,0001$  dan  $R^2$  bernilai  $0,9761 \pm 0,0022$  dan Algoritma *Gaussian support vector Regression* (non linier SVR) menunjukkan hasil MSE bernilai  $0,0117 \pm 0,0014$ , RMSE bernilai  $0,1082 \pm 0,0064$ , MAE bernilai  $0,0714 \pm 0,0006$ , MAPE bernilai  $0,0029 \pm 0,0000$  dan  $R^2$  bernilai  $0,9891 \pm 0,0011$ , dapat di tujukan pada tabel 3.

Tabel 3 Perbandingan Performa algoritma *Linear Regression* dan *Support Vector Regression* (linier dan gaussian non linier)

NO	Parameter Berpengaruh	Model	MSE	RMSE	MAE	MAPE	$R^2$
1	Semua Parameter (RR, WAV, HAV, SUN, PPP, DEW_POINT)	Linear Regression	$0,0204 \pm 0,0039$	$0,1425 \pm 0,0139$	$0,0972 \pm 0,0055$	$0,0039 \pm 0,0002$	$0,9810 \pm 0,0033$
2	Semua Parameter (RR, WAV, HAV, SUN, PPP, DEW_POINT)	Linear SVR	$0,0257 \pm 0,0027$	$0,1603 \pm 0,0086$	$0,0962 \pm 0,0023$	$0,0038 \pm 0,0001$	$0,9761 \pm 0,0022$
3	Semua Parameter (RR, WAV, HAV, SUN, PPP, DEW_POINT)	Gaussian SVR (non linier)	$0,0117 \pm 0,0014$	$0,1082 \pm 0,0064$	$0,0714 \pm 0,0006$	$0,0029 \pm 0,0000$	$0,9891 \pm 0,0011$

Adapun perbandingan hasil prediksi temperatur udara rata-rata menggunakan algoritma *Linear*

Regression dan Support Vector Regression (linier dan gaussian non linier) di tujukan Tabel 4.

Tabel 4 Perbandingan Prediksi Temperatur rata-rata *Linear Regression* dan *Support Vector Regression* (linier dan gaussian non linier)

Data Aktual	Data Prediksi <i>Linear Regression</i>	Data Prediksi <i>Linear SVR</i>	Data Prediksi <i>Gaussian SVR (non linier)</i>
25.3	25.4	25.4	25.4
26.7	26.7	26.5	26.6
26.6	26.7	26.5	26.7
24.5	24.4	24.4	24.5
24.2	24.0	24.1	24.2
24.2	24.0	24.1	24.2
25.7	25.7	25.7	25.5
25.2	25.3	25.3	25.2
25.5	25.6	25.6	25.6
25.3	25.2	25.2	25.2
25.3	25.4	25.4	25.4
25.8	25.9	25.7	25.8
24.6	24.6	24.6	24.6
25.6	25.4	25.2	25.6
26.2	26.2	26.2	26.2
23.2	23.3	23.3	23.2
24.6	24.6	24.6	24.6
23.7	23.6	23.7	23.8
25.7	25.7	25.6	25.7
25.9	25.9	25.8	25.8

#### 4. KESIMPULAN

Penelitian untuk prediksi data temperatur udara rata-rata di Malang periode data tahun 2019-2023 menggunakan dataset dari BMKG dengan parameter cuaca temperatur rata-rata (TAV), kelembaban udara (HAV), kecepatan angin (WAV), curah hujan (RR), tekanan udara (PPP), penyinaran matahari (SUN) dan titik embun (DEW\_POINT). Kedua Algoritma *machine learning* dapat di terapkan untuk prediksi temperatur udara rata-rata di Malang dengan hasil prediksi yang baik yaitu mencapai lebih dari 90%. Kinerja model dievaluasi menggunakan pengukuran metrik MSE, RMSE, MAE, MAPE dan R<sup>2</sup>. Performa model *machine learning* dari algoritma *linear regression* menunjukkan hasil MSE bernilai  $0,0204 \pm 0,0039$ , RMSE bernilai  $0,1425 \pm 0,0139$ , MAE bernilai  $0,0972 \pm 0,005$ , MAPE bernilai  $0,0039 \pm 0,0002$  dan R<sup>2</sup> bernilai  $0,9810 \pm 0,0033$  sedangkan dari algoritma *linear support vector Regression* menghasilkan nilai MSE  $0,0257 \pm 0,0027$ , RMSE bernilai  $0,1603 \pm 0,0086$ , MAE bernilai  $0,0962 \pm 0,0023$ , MAPE bernilai  $0,0038 \pm 0,0001$  dan R<sup>2</sup> bernilai  $0,9761 \pm 0,0022$ . Adapun dari Algoritma Gaussian *support vector Regression* (non linier SVR) menunjukan hasil MSE bernilai  $0,0117 \pm 0,0014$ , RMSE bernilai  $0,1082 \pm 0,0064$ , MAE bernilai  $0,0714 \pm 0,0006$ , MAPE bernilai  $0,0029 \pm 0,0000$  dan R<sup>2</sup> bernilai  $0,9891 \pm 0,0011$ . Algoritma Gaussian *support vector Regression* (non linier SVR) lebih akurat dibanding dengan *linear support vector Regression* dan algoritma *linear regression* berdasarkan nilai dari korelasinya yang lebih mendekati 1 (satu) dan tingkat error nya lebih mendekati 0 (nol) dan nilai error yang lebih rendah pada semua metrik pengukuran.

#### 5. SARAN

Penelitian ini memerlukan pengembangan lebih lanjut untuk memperoleh hasil yang paling akurat dan kemungkinan dapat di terapkan untuk area cakupan yang lebih luas di Propinsi Jawa Timur.

DAFTAR PUSTAKA

- [1] S. Prasetyo, Ulil Hidayat, Yosafat Donni Haryanto, and Nelly Florida Riama, “Karakteristik Suhu Udara di Pulau Jawa Kaitannya Dengan Kelembapan Udara, Curah Hujan, SOI, dan DMI,” *J. Geogr. Edukasi dan Lingkung.*, vol. 5, no. 1, pp. 15–26, 2021, doi: 10.22236/jgel.v5i1.5971.
- [2] S. Rahmabudhi, “Analisis Hubungan Suhu Udara di Provinsi Banten Terhadap Parameter Kelembapan Udara, Curah Hujan, ENSO, SOI, dan IOD,” *Bul. Meteorol. Klimatologi dan Geofis.*, vol. 5, no. 1, pp. 37–47, 2024.
- [3] A. Sutaryani, S. Sunarno, and D. Djuniadi, “Perbandingan Performa Model Machine Learning Dalam Prediksi Suhu Di Semarang,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 12, no. 3, pp. 2270–2275, 2024.
- [4] R. Yulistiani, Ramadhan, Ilham, Q. Said, and M. Mega Santoni, “Metode Regresi Linier Berganda dan SVR dalam Menentukan Tingkat Pengaruh Cuaca Terhadap Produktivitas Padi di Indonesia,” *Semin. Nas. Inform. Sist. Inf. Dan Keamanan Siber*, no. 2017, pp. 29–34, 2018.
- [5] R. Limehuwey, W. Multi, and S. H. Kotarumalos, “Prediksi Suhu Udara Rata-rata Harian Dengan Metode Regresi Linier di Kota Ambon,” *Tanah Goyang*, vol. 1, no. 1, pp. 18–25, 2023.
- [6] N. A. Siregar, “Peramalan Curah Hujan Di Kota Medan menggunakan Metode Support Vector Regression,” *J. Informatics Data Sci.*, vol. 1, no. 1, pp. 7–9, 2022, doi: 10.24114/j-ids.v1i1.38544.
- [7] G. Namlea Lesnusa, D. Shinta Angreni, and R. Ardiansyah, “Indonesian Journal of Computer Science,” *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- [8] K. A. Zahroh, D. Candra, R. Novitasari, and W. D. Utami, “Prediksi Kecepatan Angin di Kabupaten Kupang Menggunakan Metode Support Vector Regression ( SVR ) Wind Speed Prediction in Kupang Regency Using the Support Vector Regression ( SVR ) Method,” *TELKA*, vol. 10, no. 3, pp. 193–203, 2024.
- [9] B. Septian Cahya Putra, I. Tahyudin, B. Adhi Kusuma, and K. Nur Isnaini, “Efektivitas Algoritma Random Forest , XGBoost , dan Logistic Regression dalam Prediksi Penyakit Paru-paru,” vol. 23, no. 4, pp. 909–922, 2024.
- [10] D. A. Trianggana, “Peramalan Jumlah Siswa-Siswi Melalui Pendekatan Metode Regresi Linear,” *J. Media Infotama*, vol. 16, no. 2, pp. 115–120, 2020, doi: 10.37676/jmi.v16i2.1149.
- [11] M. Yusuf, A. Setyanto, and K. Aryasa, “Analisis Prediksi Curah Hujan Bulanan Wilayah Kota Sorong Menggunakan Metode Multiple Regression,” *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 405–417, 2022.
- [12] D. Faatihah Ramadhani Putri, I. Arianti, and J. Windu Nugroho, “Analysis of Using Support Vector Regression (SVR) Algorithm to Predict The Occurrence of Sea Tides in Tanjung Medang, Riau by Saas Storage in Cloud Computing,” *J. Mandiri IT*, vol. 11, no. 1, pp. 34–40, 2022.
- [13] M. El-Morshedy, Z. Almaspoor, G. Srinivasa Rao, M. Ilyas, and A. Al-Bossly, “On Predictive Modeling for the AI2O3Data Using a New Statistical Model and Machine Learning Approach,” *Adv. Civ. Eng.*, vol. 2022, no. ii, 2022, doi: 10.1155/2022/9348980.
- [14] R. Bárcenas, R. Fuentes-García, and L. Naranjo, “Mixed kernel SVR addressing Parkinson’s progression from voice features,” *PLoS One*, vol. 17, no. 10 October, pp. 1–23, 2022, doi: 10.1371/journal.pone.0275721.
- [15] H. Muthiah, U. Sa, and A. Efendi, “Support Vector Regression (SVR) Model for Seasonal Time Series Data,” *Proc. Second Asia Pacific Int. Conf. Ind. Eng. Oper. Manag.*, no. September 14-16, 2021, pp. 3191–3200, 2021.
- [16] D. C. R. Novitasari *et al.*, “Weather Parameters Forecasting as Variables for Rainfall Prediction using Adaptive Neuro Fuzzy Inference System (ANFIS) and Support Vector

- Regression (SVR),” *J. Phys. Conf. Ser.*, vol. 1501, no. 1, 2020, doi: 10.1088/1742-6596/1501/1/012012.
- [17] Y. Yang *et al.*, “A support vector regression model to predict nitrate-nitrogen isotopic composition using hydro-chemical variables,” *J. Environ. Manage.*, vol. 290, no. November 2020, p. 112674, 2021, doi: 10.1016/j.jenvman.2021.112674.
- [18] A. Rafati, M. Joorabian, E. Mashhour, and H. R. Shaker, “High dimensional very short-term solar power forecasting based on a data-driven heuristic method,” *Energy*, vol. 219, p. 119647, 2021, doi: 10.1016/j.energy.2020.119647.
- [19] T. M. N. Utami, D. C. R. Novitasari, F. Setiawan, N. Ulinuha, and Y. Farida, “Tide Prediction in Prigi Beach using Support Vector Regression (SVR) Method,” *Sci. J. Informatics*, vol. 8, no. 2, pp. 194–201, 2021, doi: 10.15294/sji.v8i2.28906.
- [20] I. D. Sulistyowati, S. Sunarno, and D. Djuniadi, “Penerapan Machine Learning Dengan Algoritma Support Vector Machine Untuk Prediksi Kelembapan Udara Rata - Rata,” *Just IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 15, no. 1, pp. 284–290, 2024.
- [21] I. Colanus, R. Drajana, and F. Selection, “258730-Metode-Support-Vector-Machine-Dan-Forwar-C588Ca21,” vol. 9, pp. 116–123, 2017.
- [22] B. S. D. Pratama W, “Prediksi Pemeliharaan Transformator Distribusi Berbasis Artificial Neural Network,” *Snestik*, pp. 112–120, 2023.
- [23] A. A. Suryanto, “Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi,” *Saintekbu*, vol. 11, no. 1, pp. 78–83, 2019, doi: 10.32764/saintekbu.v11i1.298.
- [24] P. Nehra and A. Nagaraju, “Host utilization prediction using hybrid kernel based support vector regression in cloud data centers,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6481–6490, 2022, doi: 10.1016/j.jksuci.2021.04.011.
- [25] F. Novianti, N. Ulinuha, M. Hafiyusholeh, and A. Arianto, “Prediksi Penggunaan Bahan Bakar pada PLTGU menggunakan Metode Support Vector Regression (SVR),” *Techno.Com*, vol. 21, no. 2, pp. 249–255, 2022, doi: 10.33633/tc.v21i2.5712.