VARIANSI: Journal of Statistics and Its Application on Teaching and Research **ISSN 2684-7590 (Online)**
Vol. 7 No. 2 (2025), 115-168

DOI: 10.35580/variansiunm400

Penerapan Metode *Support Vector Regression* (SVR) Dalam Memprediksi Indeks Standar Pencemar Udara (ISPU) Di Kota Makassar

Sudarmin, Ruliana*, Sasa Arisa

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Keywords: Air Quality, ISPU*, Support Vector Regression* (SVR)*, Grid Search Optimization, Root Mean Square Error* (RMSE).

Abstract:

Air quality is a comprehensive indicator that reflects air pollution. The air quality index is defined as a description or value of the transformation of individual parameters of interrelated air pollution, such as PM_{10}, SO_2, CO, O_3, NO_2 into one value or a set of values so that it is easy to understand for the general public. Therefore, predictions of the Air Pollutant Standard Index in the future are needed for future policy decision making. SVR is a development of Support Vector Machine (SVM) for regression cases. The aim of this study is to predict the Air Pollutant Standard Index (ISPU) in the future using the SVR method. In the SVR method, the best kernel is used as an aid in solving non-linear problems, the Min-Max Normalization method for data normalization, division of training and testing data, selection of the best model with Grid Search Optimization. The best prediction results were obtained using a radial kernel with values $k = 2$ with parameters $\varepsilon = 0.1$, $C = 10$, and $\gamma = 1$ with the smallest error of 0.0086, with an RMSE of 0.0894. The RMSE value indicates that the model's ability to follow data patterns well. From the model on the radial kernel, the predicted results of the Air Pollution Standard Index from January 1, 2025 to July 31, 2025 were obtained which were not constant or fluctuating with an interval range of 41,14 – 58,69 and based on the Air Pollution Standard Index category, it was in the fairly good category index.

1. Pendahuluan

Prediksi digunakan untuk mendapatkan informasi yang akan terjadi dalam kurun waktu tertentu di masa mendatang dengan menggunakan data masa lalu (historis) agar dapat membuat keputusan dan perencanaan yang efektif dan efisien. Salah satu metode yang dapat dilakukan untuk melakukan prediksi adalah *Support Vector Machine* (SVM). Vladimir N. Vapnik tahun 1999 memperkenalkan konsep penting dalam *Support Vector Machine* dan bagaimana pendekatan ini diperluas ke *Support Vector Regression* dengan menggunakan konsep *ε-insentive loss function*. SVM untuk kasus klasifikasi bisa digeneralisasi untuk melakukan pendekatan fungsi regresi (Antony et al., 2021). Berbeda dengan SVM yang membagi *dataset* (klasifikasi) ke dalam dua zona, SVR berusaha agar semua *dataset* masuk kedalam satu zona, dengan tetap meminimalisasi nilai *error* (ε).

Support Vector Regression (SVR) adalah penerapan (SVM) yang diperkenalkan untuk kasus regresi. Konsep SVR didasarkan pada *risk minimization*, yaitu untuk mengestimasi suatu fungsi $\boldsymbol{f(x)}$ dengan cara meminimalkan batas atas dari *generalization error*, sehingga SVR mampu mengatasi *overfitting* (Putu et al., 2019). *Output* dari SVR adalah berupa bilangan riil dan kontinu. Tujuan algoritma SVR adalah menemukan garis pemisah atau bisa disebut dengan *hyperplane* terbaik. *Hyperplane* terbaik dapat ditemukan dengan cara mengukur magin dengan *hyperplan*e tersebut.

* Corresponding author.
E-mail address: ruliana.t@unm.ac.id

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Margin sendiri adalah jarak dari *hyperplane* dengan data yang terdekat. Data yang paling dekat dari margin disebut dengan *support vector* (Dini Maulana et al., 2019).

Pengamatan yang akan dilakukan menggunakan data Indeks Standar Pencemar Udara (ISPU) di Kota Makassar yang merupakan data nonlinear sehingga cocok untuk menggunakan metode SVR. Indeks Standar Pencemar Udara (ISPU) merupakan kondisi kualitas udara ambien pada suatu wilayah sebagai dasar dampak pada kesehatan makhluk hidup (Khumaidi et al., 2020). Di Indonesia, ISPU berfungsi sebagai indikator tingkat kualitas udara. Kualitas udara merupakan indikator komprehensif yang mencerminkan polusi terhadap udara. Polusi udara menggambarkan suatu kondisi di mana kualitas udara menjadi rusak dan terkontaminasi oleh zat pencemar.

Salah satu tantangan utama yang sering terjadi di kota besar dengan aktivitas tinggi adalah pencemaran udara. Masalah ini disebabkan oleh tingginya jumlah kendaraan bermotor, pembangunan infrastruktur, dan aktivitas industri. Sebagai salah satu kota besar, kota Makassar juga mengalami hal yang sama. Berdasarkan data dari Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia diperoleh bahwa di Kota Makassar sendiri memiliki skor ISPU sebesar 42 yang berarti berada pada tingkat polusi yang relatif baik (MENLHK, 2023). Namun, menurut data BPS Kota Makassar menunjukkan terjadi peningkatan jumlah kepadatan penduduk di Kota Makassar dan jumlah kendaraan bermotor di Kota Makassar tiap tahunnya (Kota Makassar Dalam Angka 2024, 2024).

Berdasarkan penjelasan diatas metode yang akan digunakan pada penelitian ini adalah SVR. Ada beberapa penelitian yang telah membahas mengenai SVR, diantaranya : (1) Penelitian yang dilakukan oleh (Dini Maulana et al., 2019) yang membahas mengenai implementasi metode SVR dalam peralaman penjualan roti dan diperoleh nilai RMSE sebesar 0,0017 dengan hasil yang terbilang sangat baik. (2) Penelitian yang dilakukan (Mawan P & Papilaya, 2023) yang membahas mengenai Analisa prediksi harga emas menggunakan metode SVR dan diperoleh nilai MAPE sebesar 4,8% yang dapat dikategorikan sebagai nilai yang sangat baik. (3) Penelitian oleh (Ruliana et al., 2024) yang membahas penerapan metode Support Vector Regression (SVR) dalam memprediksi inflasi di Kota Makassar dan diperoleh nilai RMSE sebesar 0,029 yang artinya kemampuan model dapat mengikuti pola data dengan baik. (4) Penelitian yang dilakukan oleh (Safira et al., 2023) membahas mengenai prediksi jumlah kasus terkonfirmasi Covid-19 di Indonesia menggunakan metode SVR dan diperoleh nilai MAPE sebesar 0,49% yang artinya akurasi prediksi sangat baik atau model mempunyai kemampuan prediksi yang sangat baik. Dari hasil penelitian yang telah dilakukan sebelumnya, metode SVR yang diterapkan pada penelitian diatas memberikan kesimpulan bahwa metode SVR sudah cukup baik dalam hasil prediksinya.

2. Tinjauan Pustaka

2.1 *Support Vector Regression (SVR)*

Menurut Smola & Scholkopt (2003), *Support Vector Regression* (SVR) bertujuan untuk menemukan sebuah fungsi $f(x)$ sebagai suatu *hyperplane* (garis pemisah) berupa fungsi regresi yang mana sesuai dengan semua *input* data dengan membuat galat sekecil mungkin.

Ide dasar penggunaan metode SVR adalah dengan memisalkan terdapat n set data *training*, (x_i, y_i), dengan $i = 1,2, \dots, n$. Sedangkan $x_i = \{x_1, x_2, \dots, x_p\} \in R^n$ merupakan vector dari ruang input dan $y_i = \{y_1, y_2, \dots, y_n\} \in R$ merupakan nilai *output* berdasarkan x_i yang bersesuaian. Persamaan fungsi regresi secara umum dapat ditulis sebagai berikut (Smola & Schölkopf 2003) :

$$f(x) = \langle w.x_i \rangle + b \tag{2.1}$$

keterangan :
$f(x)$ = fungsi regresi;
x = vektor input;
w = vektor pembobot;
b = koefisien bias.

2.2 Fungsi Kernel

Untuk membantu mengatasi permasalahan nonlinear pada dimensi tinggi yang harus dilakukan yaitu mengganti inner product (x_i dan x_j) dengan fungsi kernel. Fungsi kernel digunakan dalam berbagai algoritma machine learning seperti *Support Vector Machine* (SVM) dan *Support Vector Regression* (SVR). Pemilihan fungsi kernel yang sesuai penting untuk diperhatikan karena fungsi kernel menentukan fitur baru di lokasi *hyperplane* yang nantinya akan dicari (Laverda Subiyanto et al., 2023) Persamaan kernel yang digunakan pada metode *Support Vector Regression* (SVR) dapat dilihat sebagai berikut :

1) Linear : $K(x_i, x_j) = K({x_i}^T x_j)$ (2.2)
2) Polynomial : $K(x_i, x_j) = (x_i\, x_j + 1)^P$, (2.3)
   $P = 1,2,3, \dots, n$ dengan P adalah derajat polinomial
3) Radial Basis Function (RBF) : $K(x_i, x_j) = \exp\left(-\gamma(x_i - x_j)^2\right)$ (2.4)
4) Sigmoid : $K(x_i, x_j) = \tan(\gamma(x_i - x_j) + c)$ (2.5)

2.3 Normalisasi Data

Sebelum melakukan pelatihan, perlu dilakukan normalisasi pada data *training* dan data *testing*. Normalisasi adalah sebuah teknik yang digunakan dalam penskalaan atau sebuah tahap dalam *preprocessing*. Tujuan normalisasi pada data adalah menemukan bobot yang sama dari semua atribut data dan tidak bervariasi (Malem Ginting et al., 2021). Dalam melakukan normalisasi data, terdapat beberapa metode yang dapat digunakan, diantaranya normalisasi data min max. Normalisasi data min max adalah metode normalisasi dengan mentransformasi linier data asli sehingga menghasilkan nilai perbandingan yang berkolerasi antar data sebelum dan sesudah proses. Data ini umumnya berskala 0 hingga 1 atau -1 hingga 1 (Mukhametzyanov, 2023). Berikut merupakan formulasi perhitungan normalisasi data :

$$N = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (2.6)$$

keterangan :
$N =$ nilai hasil normalisasi
$x =$ atribut data
$x_{min} =$ nilai minimum data
$x_{max} =$ nilai maksimum data

2.4 *Grid Search Optimization*

Permasalahan yang sering terjadi ketika menggunakan metode SVR adalah pada saat menentukan parameter model yang optimal. Salah satu cara optimasi yang dapat digunakan untuk menentukan parameter terbaik pada metode SVR adalah optimasi *Grid Search* atau *Grid Search Optimization.* Dalam aplikasinya, optimasi *grid search* biasanya diukur dengan *cross-validation* (*Grid Search* CV) pada data *training*.

Cross-validation adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data *training* dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan (Yasin et al., 2014).

Salah satu metode *cross validation* yang umum digunakan adalah *k-fold validation*. Menurut (Han et al., 2011), prosedur dari metode *cross validation* adalah sebagai berikut:

1. Membagi data menjadi *k* bagian dengan ukuran yang sama,
2. *k*-1 bagian dijadikan data *training* dan satu bagian dijadikan data *testing*.
3. Proses ini dilakukan sebanyak k pengulangan pada setiap kombinasi data *training* dan data *testing*.

2.5 *Root Mean Square Error* (RMSE)

Salah satu ukuran untuk mengukur tingkat akurasi peramalan adalah dengan menggunakan *Root Mean Square Error* (RMSE). RMSE mengukur seberapa tersebar dari residual. Dengan kata lain, RMSE menyatakan seberapa terkonsentrasi data di sekitar garis prediksi (Laverda Subiyanto et al., 2023). Keakuratan estimasi kesalahan (galat) ini ditunjukan dengan kecilnya nilai RMSE yang diperoleh. Perhitungan galat dengan menggunakan RMSE (Priliani et al., 2018), dirumuskan sebagai berikut :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (2.7)$$

Dengan:
y_i = nilai aktual periode ke-i
$\hat{y}_i$ = nilai prediksi periode ke-i
n = jumlah data

2.6 Indeks Standar Pencemar Udara (ISPU)

Indeks Standar Pencemar Udara (ISPU) berfungsi sebagai tolok ukur utama untuk mengkategorikan dan menjelaskan kualitas udara yang didasarkan pada dampak besar terhadap kesejahteraan individu dan vitalitas semua makhluk hidup. Terdapat lima parameter dalam Indeks Standar Pencemar Udara sebagai pengamatan kualitas udara seperti : Tingkat Partikulat (PM_{10}), Oksida Nitrogen (NO_2), Sulfur Dioksida (SO_2), Karbon Monoksida (CO), dan ozon permukaan (O_2). Jika senyawa-senyawa tersebut melampaui ambang batas yang dapat diterima, maka berpotensi membahayakan kesejahteraan seseorang, bahkan menyebabkan kematian, terutama pada sistem pernafasan (Sajiwo et al., 2024).

Berdasarkan Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor 14 Tahun 2020, Indeks Standar Pencemar Udara (ISPU) dibagi menjadi lima kategori berbeda, yaitu :

Tabel 1. Kategori Indeks Standar Pencemar Udara

Rentang	Kategori	Keterangan
$0 - 50$ (Hijau)	Baik	Tingkat kualitas udara yang tidak memberikan efek bagi Kesehatan manusia atau hewan dan tidak berpengaruh pada tumbuhan, bangunan ataupun nilai estetika.
$51 - 100$ (Biru)	Sedang	Tingkat kualitas udara yang tidak memberikan efek bagi Kesehatan manusia atau hewan, tetapi berpengaruh pada tumbuhan yang sensitif dan nilai estetika.
$101 - 199$ (Kuning)	Tidak Sehat	Tingkat kualitas udara yang bersifat merugikan pada manusia ataupun kelompok hewan yang sensitive atau bisa menimbulkan kerusakan pada tumbuhan ataupun nilai estetika.
$200 - 299$ (Merah)	Sangat Tidak Sehat	Tingkat kualitas udara yang dapat merugikan Kesehatan pada sejumlah segmen populasi yang terpapar.
> 300 (Hitam)	Berbahaya	Tingkat kualitas udara berbahaya yang secara umum dapat merugikan Kesehatan yang serius pada populasi.

Rumus Linear Interpolasi ISPU yang digunakan untuk mengubah konsentrasi polutan (PM_{10}, SO_2, CO, O_3, NO_2) ke dalam bentuk angka indeks kualitas udara sebagai berikut

$$I = \frac{I_a - I_b}{X_a - X_b}(X_x - X_b) + I_b \ \ldots \quad (2.8)$$

Dimana :
I = ISPU terhitung
I_a = ISPU batas atas
I_b = ISPU batas bawah
X_a = Ambien batas atas ($\mu g/m^3$)
X_b = Ambien batas bawah ($\mu g/m^3$)
X_x = Kadar Ambien nyata hasil pengukuran ($\mu g/m^3$)

Adapun ISPU gabungan adalah nilai ISPU yang mewakili kualitas udara pada suatu waktu dan lokasi tertentu, dihitung berdasarkan nilai ISPU tertinggi dari masing-masing polutan yang diukur ((PM_{10}, SO_2, CO, O_3, NO_2). Berikut perhitungan ISPU gabungan :

$$ISPU\ Gabungan = \max\left(ISPU_{PM_{10}}, ISPU_{SO_2}, ISPU_{CO}, ISPU_{O_3}, ISPU_{NO_2}\right)$$

Artinya :
- Setiap polutan dihitung ISPU-nya masing-masing berdasarkan konsentrasinya.
- ISPU Gabungan adalah nilai tertinggi diantara semua ISPU tersebut.
- Polutan dengan nilai tertinggi tersebut juga disebut kontributor utama pencemaran (parameter pencemar kritis) udara pada waktu itu.

3. Metode Penelitian

3.1 Jenis Penelitian
Jenis penelitian yang digunakan adalah penelitian kuantitatif. Penelitian kuantitatif adalah penelitian yang berfokus pada analisis data berupa angka yang diolah menggunakan metode statistika.

3.2 Sumber Data
Pada penelitian ini, menggunakan data sekunder yang bersumber dari Dinas Lingkungan Hidup Kota Makassar. Dalam penelitian ini data yang digunakan merupakan data harian dengan periode 1 Januari 2024 sampai 31 Desember 2024 sebanyak 366 data.

3.3 Definisi Operasional Variabel
Adapun definisi operasional variabel yang digunakan pada penelitian ini mencakup nilai Indeks Standar Pencemar Udara (ISPU) tertinggi berdasarkan konsentrasi lima polutan utama (PM_{10}, SO_2, CO, O_3, dan NO_2) di Kota Makassar setiap hari, yang terdiri dari 366 data mulai 1 Januari 2024 – 31 Desember 2024. Indeks Standar Pencemar Udara (ISPU) merupakan indeks kualitas udara yang menggambarkan tingkat pencemaran udara dan dampaknya terhadap Kesehatan manusia berdasarkan beberapa polutan utama. Nilai ISPU harian yang dihitung berdasarkan konsentrasi lima polutan utama (PM_{10}, SO_2, CO, O_3, dan NO_2) menggunakan metode interpolasi linear sesuai pedoman Permen LHK. Nilai ISPU harian gabungan, ditentukan berdasarkan nilai ISPU tertinggi dari kelima polutan.

3.4 Prosedur Penelitian
Adapun tahapan-tahapan yang dilakukan berdasarkan pada tujuan penelitian adalah sebagai berikut :
1. Melakukan pengambilan data dari Dinas Lingkungan Hidup Kota Makassar.
2. Melakukan normalisasi data.
3. Membuat model regresi dengan metode *Support Vector Regression* (SVR).
4. Melakukan prediksi dengan metode *Support Vector Regression* (SVR).
5. Membuat interpretasi dan Kesimpulan berdasarkan hasil pengolahan data.
6. Menyusun laporan hasil penelitian.

3.5 Teknik Analisis Data

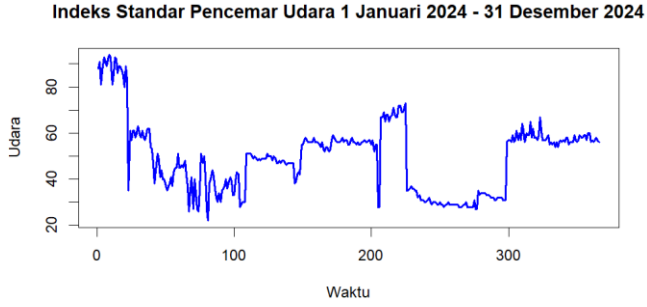
Adapun Langkah-langkah teknik analisis yang dilakukan adalah :

1. Melakukan pengumpulan data sekunder berupa data harian Indeks Standar Pencemar Udara Kota Makassar yang diperoleh dari Dinas Lingkungan Hidup Kota Makassar dengan periode data 1 Januari 2024 sampai 31 Desember 2024
2. Menormalisasikan data
3. Menentukan data *training* dan data *testing*.
4. Penentuan nilai parameter pada data *training*. Penentuan nilai parameter C (cost), ε (Epsilon), dan γ (Gamma) dengan menerapkan *Grid Search Optimization*
5. Penentuan model terbaik dengan melihat akurasi hasil ketepatan prediksi dengan perhitungan RMSE
6. Mengimplementasikan model terbaik pada data *testing*
7. Melakukan prediksi tiga bulan kedepan
8. Melihat akurasi hasil ketepatan prediksi dengan perhitungan RMSE
9. Membuat Interpretasi dan Kesimpulan berdasarkan hasil pengolahan data

4. Hasil dan Pembahasan

4.1 Analisis Deskriptif

Data yang digunakan dalam penelitian ini adalah data harian Indeks Standar Kualitas Pencemar Udara (ISPU) di Kota Makassar dari tanggal 1 Januari 2024 sampai 31 Desember 2024 dengan total 366 data yang bersumber dari Dinas Lingkungan Hidup Kota Makassar. Berikut grafik dari data yang diteliti:



Gambar 1. Plot Time Series Indeks Kualitas Udara di Kota Makassar

Dari gambar 1 dapat dilihat bahwa Indeks Standar Pencemar Udara di Kota Makassar yang dilakukan dalam kurun satu tahun (366 hari), Indeks Standar Pencemar Udara jika dilihat dari *plot* datanya tidak tetap atau tidak membentuk garis lurus yang artinya data yang diteliti adalah data *non-linear*.

Tabel 2. Statistik Deskriptif

Statistika Deskriptif	ISPU
Mean	49,62
Median	51
Standar Deviasi	15,54
Min	22
Max	94

Pada Tabel 2, dapat dilihat bahwa nilai maksimum sebesar 94 yang terdapat pada bulan Januari 2024 dan nilai minimum dari data ISPU sebesar 22 pada bulan Maret 2024. Median dari data ISPU sebesar 51, mean sebesar 49,62, dan standar deviasi atau jarak antar data terhadap nilai rata-ratanya sebesar 15,54.

Langkah selanjutnya yaitu menormalisasikan data menggunakan metode normalisasi data *min max*. Setelah melakukan normalisasi data, dilanjutkan dengan membagi data *training* dan data *testing*.

4.2 Normalisasi Data

Normalisasi data merupakan proses penyusunan struktur basis data guna mengurangi atau menghilangkan sebagian besar ambiguity. Dalam proses peramalan, normalisasi data adalah salah satu tahapan penting untuk meminimalkan variasi dari data dan nilai error prediksi. Normalisasi min max merupakan metode normalisasi yang menerapkan transformasi linear pada data asli sehingga menghasilkan nilai perbandingan yang berkorelasi antara nilai sebelum dan sesudah proses. Hasil dari normalisasi data dapat dilihat pada Tabel 3.

Tabel 3. Data Hasil Normalisasi

Hari	ISPU
1	0,916
2	0,958
3	0,819
4	0,902
5	0,986
6	0,958
7	0,930
8	0,972
9	1
10	0,986
.	.
.	.
.	.
366	0,472

4.3 Pembagian Data *Training* dan Data *Testing*

Sebelum data dianalisis menggunakan metode *Support Vector Regression* (SVR) terlebih dahulu data dibagi menjadi data training dan data testing. Data *training* merupakan data yang digunakan untuk membangun model dengan metode SVR. Sedangkan data *testing* adalah data yang digunakan untuk memprediksi dengan metode SVR berdasarkan model yang telah didapatkan sebelumnya. Pada penelitian ini dilakukan pembagian data *training* dan data *testing* dengan perbandingan 80 : 20 (Gholamy et al., 2018) dimana jumlah data *training* memiliki presentase lebih besar agar model bisa belajar dengan baik dalam memahami dan mengenali pola umum sehingga model yang telatih dapat lebih optimal dalam peramalan pada data *testing*. Pembagian data *training* dan data *testing* dilakukan secara acak untuk menghindari bias, meningkatkan generalisasi model dan memastikan pembagian data tetap representatif. Hasil pembagian data *training* dan data *testing* dapat dilihat pada 4 dan Tabel 5.

Tabel 4. Data *Training*

Hari	ISPU
1	0,917
2	0,958
3	0,819
4	0,903
5	0,986
.	.
.	.
295	0,472

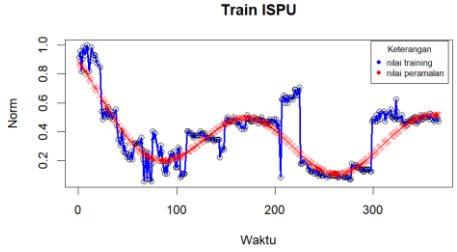
Tabel 5. Data *Testing*

No	ISPU
1	0,986
2	0,972
3	0,889
4	0,806
5	0,931
.	.
.	.
71	0,472

4.4 Penerapan *Metode Support Vector Regression*

4.4.1 Membangun Model *Support Vector Regression*

Model SVR yang akan dibangun pada *software* R dibutuhkan *package* tambahan yaitu *package* e1071. Setelah menginstal *package* e1071 selanjutnya membuat model SVR pada data *training*. Setelah model didapatkan maka dilakukan prediksi pada data training tanpa menggunakan kernel.



Gambar 2. Prediksi SVR tanpa menggunakan kernel

Titik berwarna merah pada Gambar 2 merupakan hasil prediksi dengan parameter $\varepsilon = 0{,}1, C = 1$, dan $\gamma = 1$, sedangkan titik biru merupakan data *training*. Untuk mendapatkan hasil yang maksimal dengan melihat parameter yang terbaik serta *error* paling minim maka di perlukan pengoptimalan model.

4.4.2. Pemilihan Model Terbaik dengan *Grid Search Optimization*

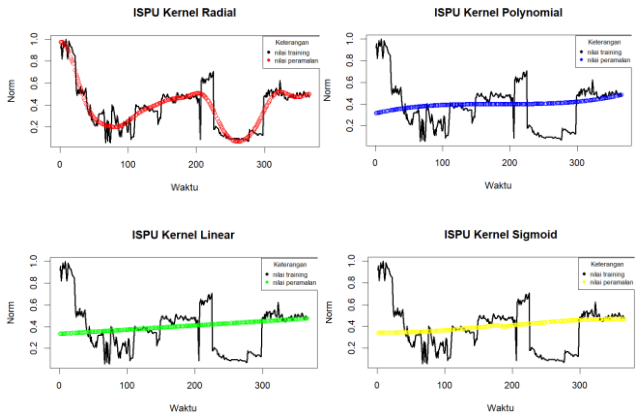
Grid Seacrh digunakan untuk menemukan parameter optimal dalam suatu model, sehingga model yang digunakan dapat memprediksi secara akurat data yang digunakan. Pengoptimalan model SVR menggunakan *Grid Search* membutuhkan parameter sesuai dengan kernelnya. *Grid Search* dijalankan dengan Teknik *k-fold cross validation.*

Tabel 6. Parameter Hasil Grid Search Optimization

Kernel	Jumlah k	ε Terbaik	C Terbaik	γ Terbaik	*Error* terkecil
Radial	**2**	**0,1**	**10**	**5**	**0,0086**
	5	0,1	10	5	0,0091
	10	0,1	10	5	0,0092
Polynomial	**2**	**0,1**	**0,05**	**1**	**0,0502**
	5	0,1	0,05	1	0,0506
	10	0,1	0,05	1	0,0506
Linear	**2**	**0,1**	**0,05**	**1**	**0,0504**
	5	0,1	0,05	1	0,0509
	10	0,1	0,05	1	0,0512
Sigmoid	2	0,1	0,05	1	0,0504
	5	0,1	0,05	1	0,0505
	10	**0,1**	**0,05**	**5**	**0,0496**

Dari Tabel 6 diatas, didapatkan nilai parameter yang berbeda-beda pada setiap kernel dan jumlah *k.* Terlihat bahwa kernel Radial error terkecilnya saat nilai $k = 2, \varepsilon = 0{,}1,\ C = 10$, dan $\gamma = 5$ sebesar 0,0086. Pada kernel polynomial *error* terkecilnya saat nilai $k = 2, \varepsilon = 0{,}1,\ C = 0{,}05$, dan $\gamma = 1$ sebesar 0,0502. Pada kernel linear *error* terkecilnya saat nilai $k = 2, \varepsilon = 0{,}1,\ C = 0{,}05$, dan $\gamma = 1$ sebesar 0,0504. Pada kernel sigmoid *error* terkecilnya saat nilai $k = 10, \varepsilon = 0{,}1,\ C = 0{,}05$, dan $\gamma = 5$ sebesar 0,0496.

Berdasarkan hasil *grid search* diketahui parameter dari model terbaik yang akan digunakan adalah parameter dari kernel *radial basis function* dengan melihat *error* terkecil yaitu 0,0086. Kernel yang terpilih selanjutnya digunakan untuk memprediksi data *training.* Prediksi data *training* dilakukan untuk mengukur tingkat keakuratan atau performa model pada data yang sudah diketahui hasilnya. Dengan melakukan prediksi pada data *training,* kita dapat membandingkan prediksi model dengan nilai sebenarnya pada data tersebut. Prediksi data *training* terlihat pada Gambar 3



Gambar 3. Prediksi Data Training dengan kernel

Dari hasil prediksi dengan menggunakan parameter terbaik pada setiap kernel, dapat dilihat bahwa kernel Radial yang mempunyai hasil prediksi pada titik warna merah yang mendekati data aktualnya. Untuk lebih akuratnya dapat dilihat pada akurasi data *training* nya menggunakan RMSE.

4.4.3. Hasil Akurasi Data *Training* pada setiap Kernel

Salah satu metode yang dapat digunakan untuk menghitung hasil evaluasi suatu model adalah dengan menggunakan metode *Root Mean Square Error* (RMSE), dimana semakin kecil (mendekati nol) nilai RMSE maka hasil prediksi akan semakin akurat. Hasil RMSE pada setiap kernel dapat dilihat pada Tabel 7.

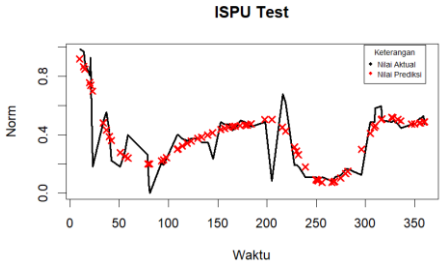
Tabel 7. RMSE tiap kernel

Kernel	RMSE
Radial	0,0894
Polynomial	0,2237
Linear	0,2255
Sigmoid	0,2260

Dari hasil Tabel 7 nilai RMSE terkecil pada kernel radial sebesar 0,0894, pada kernel polynomial 0,2237, pada kernel linear sebesar 0,2255, dan pada kernel sigmoid sebesar 0,2260. Dari hasil tersebut dapat disimpulkan bahwa kernel radial yang mempunya nilai RMSE terkecil, sehingga model pada kernel radial yang paling cocok digunakan untuk melakukan prediksi pada data *testing*.

4.4.4 Implementasi pada Data Testing

Implementasi dilakukan setelah mendapatkan model yang cocok pada data *training.* Selanjutnya, model yang telah didapat dilakukan analisis dengan memasukkan model tersebut beserta parameternya pada data *testing*. Tujuan dari data *testing* sendiri untuk menguji performa model yang telah dilatih pada data yang belum pernah dilihat sebelumnya. Data *testing* digunakan untuk mengukur sejauh mana model mampu menggeneralisasi dan melakukan prediksi atau peramalan yang akurat pada data baru. Hasil prediksinya dapat dilihat pada Gambar 4.



Gambar 4. Plot Nilai Prediksi pada Data Testing

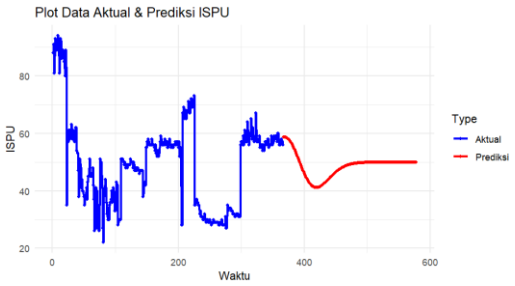
Pada Gambar 4 terlihat hasil prediksi pada data testing berupa garis berwarna merah sedangkan data *testing* ditunjukkan dengan garis berwarna hitam. Hasil nilai prediksi diketahui mendekati nilai pada data aktual. Ketepatan hasil prediksi pada data *testing* dapat dilihat menggunakan RMSE. Dengan bantuan *software* R hasil perhitungan RMSE sebesar 0,1131 yang artinya kemampuan model sangat baik dalam mengikuti pola data.

4.4.5. Hasil Peramalan Indeks Kualitas Udara di Kota Makassar

Hasil prediksi Indeks Kualitas Udara di Kota Makassar dari tanggal 1 Januari 2024 – 31 Desember 2024 dengan model terbaik dapat dilihat pada Tabel 3.9 terlihat Indeks Kualitas Udara pada tanggal 1 Januari 2025 – 31 Juli 2025 mengalami kenaikan dan penurunan di hari tertentu.

Tabel 8. Hasil Peramalan Indeks Standar Pencemar Udara

Hari	ISPU
1	59,165
2	59,493
3	59,835
4	60,189
5	60,554
6	60,929
7	61,312
8	61,702
9	62,098
10	62,498
.	.
.	.
212	56,441



Gambar 5. Hasil Peramalan ISPU

Plot grafik hasil peramalan Indeks Standar Pencemar Udara di Kota Makassar mengalami penurunan hingga 41,14 yaitu pada pertengahan Februari 2025. Indeks standar pencemar udara memiliki interval 41,14 – 58,69 dan berdasarkan kategori Indeks Standar Pencemaran Udara berada pada indeks kategori Sedang. Plot hasil peramalan ISPU dapat dilihat pada Gambar 5.

5. Pembahasan

Berdasarkan data yang diperoleh dari Dinas Lingkungan Hidup Kota Makassar, menunjukkan bahwa Indeks Standar Pencemar Udara selama tahun 2024 mengalami fluktuasi. Pada penelitian ini dilakukan pembagian data *training* dan data *testing* dengan komposisi 80 : 20, hal ini sejalan dengan penelitian yang dilakukan (Gholamy et al., 2018) yang menyatakan bahwa empiris pembagian 80 : 20 merupakan pembagian terbaik pada *set training* dan *testing.*

Pada hasil pengujian model data *training* diperoleh model terbaik setiap kernel yaitu pada kernel radial saat nilai $k = 2, \varepsilon = 0{,}1,\ C = 10$, dan $\gamma = 5$ dengan akurasi sebesar 0,0086, model terbaik pada kernel polynomial saat nilai $k = 2, \varepsilon = 0{,}1,\ C = 0{,}05$, dan $\gamma = 1$ dengan akurasi sebesar 0,0502, model terbaik pada kernel linear saat nilai $k = 2, \varepsilon = 0{,}1,\ C = 0{,}05$, dan $\gamma = 1$ dengan akurasi sebesar 0,0504, pada kernel sigmoid nilai $k = 10, \varepsilon = 0{,}1,\ C = 0{,}05$, dan $\gamma = 5$ dengan akurasi sebesar 0,0496, sehingga dari hasil akurasi pada setiap kernel dapat disimpulkan bahwa hasil parameter pada kernel radial yang paling tepat digunakan untuk prediksi pada data *testing*. Hal ini sejalan dengan penelitian yang dilakukan oleh (Yohannnes et al., 2023)dari penelitiannya, kernel RBF memberikan akurasi yang paling baik dibandingkan kernel linear atau polynomial.

Pada hasil pengujian model terbaik pada data *training* selanjutnya diterapkan pada data *testing* dengan hasil akurasi yang sangat baik dalam mengikuti pola data yaitu sebesar 0,1131, dapat dilihat juga pada Gambar 2.7 dimana hasil nilai prediksinya mendekati pola nilai data aktual. Sehingga, dari model yang diterapkan pada data *testing* diperoleh hasil prediksi Indeks Standar Pencemar Udara di Kota Makassar dari 1 Januari 2025 sampai 31 Juli 2025 pada Gambar 2.8 , dapat dilihat bahwa hasil prediksi 6 bulan kedepan mengalami perubahan yang tidak tetap atau fluktuatif. Prediksi ISPU memiliki interval 41,14 – 58,69 dan berdasarkan kategori Indeks Standar Pencemaran Udara berada pada indeks kategori sedang.

Interval ini menunjukkan bahwa Indeks Stadar Pencemar Udara masih berada dalam kategori sedang atau cukup baik menurut standar yang ditetapkan oleh Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor 14 Tahun 2020 tentang Indeks Standar Pencemar Udara. Kategori sedang mencerminkan bahwa kualitas udara masih dapat diterima untuk kegiatan masyarakat secara umum, namun ada kemungkinan memberikan dampak kesehatan ringan bagi kelompok yang sensitif seperti anak-anak, lansia, dan penderita gangguan pernapasan.

Hasil prediksi ini mengindikasikan bahwa meskipun kualitas udara tidak tergolong buruk, pemantauan secara rutin tetap diperlukan untuk mencegah perburukan kualitas udara. Hal ini penting terutama dalam konteks perkotaan seperti Makassar, yang memiliki aktivitas industry dan transportasi yang cukup padat.

6. Kesimpulan dan Saran

6.1 Kesimpulan

Berdasarkan penelitian dan pembahasan yang telah dilakukan untuk meramalkan Indeks Standar Pencemar Udara di Kota Makassar dengan Metode SVR diperoleh kesimpulan bahwa model terbaik yang digunakan adalah menggunakan kernel radial saat nilai $k = 2$ dengan parameter $\varepsilon = 0{,}1$, $C = 10$, dan $\gamma = 5$ dengan hasil akurasi sebesar 0,0086 dan nilai RMSE sebesar 0,0894. Adapun hasil prediksi dari Indeks Standar Pencemar Udara di Kota Makassar cenderung fluktuatif sampai akhir Juli 2025 dan berdasarkan hasil prediksi Indeks Standar Pencemar Udara di Kota Makassar dengan rentang nilai 41,14 – 58,69 kualitas udara tergolong sedang atau cukup baik

6.2 Saran

Dalam penelitian ini tentunya masih ada kekurangan, sehingga saran untuk penelitian selanjutnya dapat menggunakan parameter terbaik pada kernel polynomial, linear, dan sigmoid. Dapat menggunakan metode normalisasi lainnya seperti *Z-Score* atau *Decimal Scalling*.

References

Antony, E., Sreekanth, N. S., Sunil Kumar, R. K., & Nishanth, T. (2021). *Data preprocessing techniques for handling time series data for environmental science studies*. *International Journal of Engineering Trends and Technology*, *69*(5), 196–207. https://doi.org/10.14445/22315381/IJETT-V69I5P227

Badan Pusat Statistik. (2024). Kota Makassar Dalam Angka *2024*.

Dini Maulana, N., Darma Setiawan, B., & Dewi, C. (2019). Implementasi Metode *Support Vector Regression (SVR)* Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery) (Vol. 3, Issue 3). http://j-ptiik.ub.ac.id

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). *A Pedagogical Explanation A Pedagogical Explanation Part of the Computer Sciences Commons*. https://scholarworks.utep.edu/cs_techrep.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.

Khumaidi, A., Raafi, R., Permana Solihin, I., & Rs Fatmawati, J. (2020). Pengujian Algoritma *Long Short Term Memory* untuk Prediksi Kualitas Udara dan Suhu Kota Bandung. Jurnal Telematika, *15*(1).

Kurniawan, A. (2018). Pengukuran Parameter Kualitas Udara (CO, NO_2, SO_2, O_3 Dan PM_{10}) Di Bukit Kototabang Berbasis ISPU. *Jurnal Teknosains*, *7*(1), 1. https://doi.org/10.22146/teknosains.34658

Laverda Subiyanto, M., Amanda, Y., Nadhil Fachrian, M., Yazid Busthomi Rohim, A., & Chamidah, N. (2023). Peramalan Kasus Harian *Monkeypox* Dunia Dengan Pendekatan *Support Vector Regression*.

Malem Ginting, L., MTSigiro, M., Delima Manurung, E., & Jasa Putra Sinurat, J. (2021). Perbandingan Metode Algoritma *Support Vector Regression* dan *Multiple Linear Regression* Untuk Memprediksi Stok Obat. *In Journal of Applied Technology and Informatics* (Vol. 1, Issue 2). http://journal-jati.del.ac.id/

Mawan P, F. N., & Papilaya, F. S. (2023). Analisa Prediksi Harga Emas Dengan Kemungkinan Terjadinya Resesi Menggunakan Metode SVR. *Science and Information Technology*. https://doi.org/10.31598

Mukhametzyanov, I. Z. (2023). *On The Conformity of Scales of Multidimensional Normalization: An Application For The Problems of Decision Making*. *Decision Making: Applications in Management and Engineering*, *6*(1), 399–431. https://doi.org/10.31181/dmame05012023i

Priliani, E. M., Putra, A. T., & Muslim, M. A. (2018). *Forecasting Inflation Rate Using Support Vector Regression* (SVR) *Based Weight Attribute Particle Swarm Optimization* (WAPSO). *Scientific Journal of Informatics*, *5*(2), 118–127. https://doi.org/10.15294/sji.v5i2.14613

Putu, N., Hendayanti, N., Ketut, I., Suniantara, P., Nurhidayati, M., Teknologi, I., Bisnis, D., Bali, S., Islam, I. A., & Ponorogo, N. (2019). Penerapan *Support Vector Regression* (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik ke Bali (Vol. 3, Issue 1).

Ruliana, Rais, Z., Marni, & Saleh Ahmar, A. (2024). *Implementation of the Support Vector Regression* (SVR) *Method in Inflation Prediction in Makassar City. ARRUS Journal of Mathematics and Applied Science*, *4*(1). https://doi.org/10.35877/mathscience2608

Safira, A. N., Warsito, B., & Rusgiyono, A. (2023). Analisis *Support Vector Regression* (SVR) Dengan Algoritma *Grid Search Time Series Cross Validation* Untuk Prediksi Jumlah Kasus Terkonfirmasi Covid-19 di Indonesia. *Jurnal Gaussian*, *11*(4), 512–521. https://doi.org/10.14710/j.gauss.11.4.512-521

Sajiwo, A. F. B., Rahmat, B., & Junaidi, A. (2024). Klasifikasi Indeks Standar Pencemaran Udara (ISPU) Menggunakan Algoritma *XGBoost* Dengan Teknik Imbalaced Data (SMOTE). Jurnal Informatika Dan Teknik Elektro Terapan, *12*(3). https://doi.org/10.23960/jitet.v12i3.4699

Smola, A. J., Schölkopf, B., & Schölkopf, S. (2003). *A Tutorial on Support Vector Regression* *.

Yasin, H., Prahutama, A., Utami, T. W., Jurusan, D., & Undip, S. (2014). PrediksiI Harga Saham Menggunakan *Support Vector Regression* Dengan Algoritma *Grid Search*.

Yohannnes, Al Rivan Muhammad Ezar, Devella Siska, & Meiriyama. (2023). Klasifikasi Motif Songket Palembang Menggunakan *Support Vector Machine* Berdasarkan *Histogram of Oriented Greadienst*. *9*, 143–149.