

ANALISIS *SUPPORT VECTOR REGRESSION* (SVR) UNTUK MERAMALKAN INDEKS KUALITAS UDARA DI KOTA MAKASSAR

Rahmat Wahyudi¹, Suwardi Annas², Zulkifli Rais³

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

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

Abstract:

Air pollution is one of the unresolved problems to date, especially in big cities in Indonesia. This condition is certainly very worrying considering the pollutants released by motorized vehicles such as carbon monoxide (CO), particulate matter (PM), nitrogen oxides (NO₂), sulfur dioxide (SO₂), and carbon dioxide (CO₂) are very dangerous to human health. Therefore it is necessary to do research to determine the forecasting of the air quality index in the future. So in this study the SVR method was used to predict the air quality index in Makassar City. SVR is a Support Vector Machine (SVM) development for regression cases. In this study the SVR method was used with the best kernel as an aid in solving non-linear problems, the Min - Max Normalization method for data normalization, the division of training data and data testing used was 80%:20%, the selection of the best model was with Grid Search Optimization. Forecasting results showed that the five air quality index variables in the city of Makassar were classified as good with RMSE values, namely Particulate (PM10) 0.12352, Sulfur Dioxide (SO₂) 0.11502, Ozone (O₃) 0.13561, Nitrogen dioxide (NO₂) 0.11380, Carbon Monoxide (CO) 0.00699 means the model's ability to follow data patterns well.

1. Pendahuluan

Peramalan adalah teknik untuk memperkirakan nilai di masa depan dengan memperhatikan data masa lalu dan sekarang. Peramalan adalah alat penting untuk perencanaan yang efektif dan efisien. Peramalan membutuhkan metode, model, atau pendekatan yang akurasi harus diuji. Semakin akurat suatu model, semakin baik untuk prediksi. Salah satu metode peramalan yang dapat digunakan adalah metode peramalan analisis deret waktu.

Analisis deret waktu adalah teknik statistik yang digunakan untuk memprediksi struktur probabilistik kondisi yang akan terjadi di masa depan dalam rangka pengambilan keputusan (Aswi & Sukarna, 2006). Metode yang umum digunakan untuk pengklasifikasian atau peramalan adalah: *Support Vector Machine*, *Backpropagation*, *Radial Basis Function*, *K-Nearest Network*, Analisis Diskriminan, *Simple Logistic Classifier* (SLC), *Fuzzy* dan masih banyak lagi. Pada umumnya, data yang tersedia di kehidupan nyata jarang linier, sebagian besar non-linier. Salah satu metode yang dapat digunakan untuk melakukan peramalan dengan kasus data nonlinier adalah SVR.

Metode SVR (*Support Vector Regression*) merupakan pengembangan SVM untuk kasus regresi yang bertujuan untuk menemukan sebuah fungsi sebagai suatu *hyperplane* (garis pemisah) berupa fungsi regresi yang mana sesuai dengan semua input data dengan error sekecil mungkin (Scholkopf & Smola, 2004). Keunggulan SVR adalah kemampuan untuk mengatasi masalah data *non-linear* dengan menggunakan fungsi kernel sehingga dapat mengatasi

* Corresponding author.

E-mail address: rahmatwahyudi13@gmail.com



masalah *overfitting*, sehingga konsep algoritma SVR dapat menghasilkan nilai prediksi yang baik (Furi dkk., 2015). Data Indeks Kualitas Udara Makassar merupakan data *non-linear* sehingga cocok untuk menggunakan metode SVR.

Makassar merupakan kota dengan tingkat kemajuan yang cukup pesat dalam berbagai bidang, salah satunya di bidang transportasi. Dampak dari majunya industri transportasi sendiri adalah jumlah kendaraan yang terus bertambah setiap tahunnya, yang menyebabkan padatnya kendaraan dan juga bertambahnya polusi udara. Berdasarkan data dari Samsat Kota Makassar, pada 2018 tercatat 1.425.151 unit atau bertambah 87.009 unit dibandingkan 2017. Adapun, pada 2016 jumlah kendaraan bermotor di Kota Makassar berkisar 1.252.755 unit. Artinya, dalam dua tahun tercatat pertambahan 172.395 unit. Polusi udara merupakan salah satu permasalahan yang belum terselesaikan sampai saat ini terutama di kota besar di Indonesia (Prahardis dkk., 2018) Kondisi ini tentu sangat mengkhawatirkan mengingat polutan yang dikeluarkan oleh kendaraan bermotor seperti karbon monoksida (CO), nitrogen oksida (NOx), sulfur dioksida (SO₂), dan karbon dioksida (CO₂) sangat berbahaya bagi kesehatan manusia. Selain itu, informasi tentang kualitas udara ini bisa menjadi acuan bagi pemerintah untuk mengkaji kembali kebijakan yang berkaitan dengan pengaturan dan pembatasan jumlah kendaraan bermotor (Aini dkk., 2019)

Indeks kualitas udara didefinisikan sebagai gambaran atau nilai hasil transformasi parameter-parameter individual polusi udara yang saling berhubungan, seperti SO₂, NOx, SPM, Ox, CO menjadi satu nilai atau satu set nilai sehingga mudah dimengerti bagi masyarakat awam (Budiyono dkk. 2010). Tujuannya agar memberikan kemudahan dari keseragaman informasi mutu udara kepada masyarakat di lokasi dan waktu tertentu serta sebagai bahan pertimbangan dalam melakukan upaya-upaya pengendalian pencemaran udara baik bagi pemerintah pusat maupun pemerintah daerah. Pencemaran udara disebabkan oleh adanya gas buang kendaraan bermotor, kegiatan industri, pemukiman penduduk dan karena sampah. Karena data indeks kualitas udara merupakan data *non-linear*, maka peneliti memilih menggunakan metode SVR untuk menganalisis data tersebut. Metode SVR diterapkan dengan kernel *Radial Basis Function* (RBF), dengan demikian metode di atas sangat cocok untuk menangani masalah hubungan *non-linear* sesuai dengan data yang akan digunakan (Caraka dkk., 2017)

Beberapa penelitian sebelumnya yang menggunakan metode SVR (Haqee dkk., 2022) “Prediksi Indeks Standar Pencemar Udara Menggunakan Metode Support Vector Regression Berbasis Web (Studi Kasus Pada Dki Jakarta)”, dengan nilai RMSE dari 0.052748 sampai 0.117431 dan R² dari 0.393280 sampai 0.91676. Penelitian lainnya yaitu (Siregar, 2022). “Peramalan Curah Hujan Di Kota Medan menggunakan Metode *Support Vector Regression*” diperoleh hasil peramalan yang akurat dengan nilai RMSE 0.0388. Penelitian yang dilakukan oleh (Isnaeni dkk., 2022) “Analisis *Support Vector Regression* (SVR) dengan kernel *Radial Basis Function* (RBF) untuk memprediksi laju Inflasi di Indonesia” diperoleh nilai RMSE 0,0020 yang berarti kemampuan model mengikuti pola data dengan baik.

Berdasarkan beberapa penelitian sebelumnya, metode SVR yang diterapkan pada penelitian di atas memberikan kesimpulan bahwa metode SVR sudah cukup baik dalam hasil prediksinya. Kesalahan yang dihasilkan sudah cukup minim dengan bantuan kernel dan pemilihan parameter yang tepat. Berdasarkan hal tersebut penulis mengambil penelitian skripsi dengan judul Analisis *Support Vector Regression* untuk meramalkan Indeks Kualitas Udara di Kota Makassar.

2. Tinjauan Pustaka

2.1 Peramalan

Peramalan atau bisa disebut dengan *forecasting* adalah suatu kegiatan untuk meramalkan atau memprediksi keadaan di masa mendatang melalui pengujian di masa lalu (Fajarita & Hati, 2018). Peramalan merupakan proses meramalkan nilai di masa depan yang mencakup permintaan dalam hal kuantitas, kualitas, waktu dan lokasi untuk memenuhi permintaan barang, atau jasa. Beberapa teknik peramalan mencoba memproyeksikan pengalaman sejarah dalam bentuk deret waktu ke masa depan.

2.2 *Support Vector Regression* (SVR)

Support Vector Regression (SVR) merupakan pengembangan SVM untuk kasus regresi. Tujuan dari SVR adalah untuk menemukan sebuah fungsi sebagai suatu *hyperplane* (garis pemisah) berupa fungsi regresi yang mana sesuai dengan semua *input* data dengan sebuah *error* dan membuatnya sekecil mungkin (Scholkopf & Smola, 2004). Jika nilai sama dengan 0 maka diperoleh suatu persamaan regresi yang sempurna. SVR mempunyai kesamaan dengan pengklasifikasian SVM yaitu dapat memaksimalkan margin dan trik kernel untuk memetakan data yang *non-linear*.

Dalam SVR, tujuan yang akan dicapai adalah untuk menentukan fungsi $f(x)$ yang mempunyai nilai epsilon (ϵ) paling besar dari target aktual y_i , untuk keseluruhan data *training* dan pada saat yang sama juga dicari fungsi yang serata mungkin. Dengan demikian, semua kesalahan (selisih antara *output* fungsi dengan target aktual) yang nilainya kurang dari ϵ akan diabaikan, tetapi tidak akan menerima semua kesalahan yang lebih besar dari ϵ . Misalkan terdapat l

data *training*, (x_i, y_i) , $i = 1, \dots, l$ dengan data input $x = \{x_1, \dots, x_l\} \subseteq R^N$ dan $y = \{y_1, \dots, y_l\} \subseteq R$ dan l adalah banyaknya data *training*. Fungsi regresi dari metode SVR adalah sebagai berikut:

$$f(x) = w(x) + b \tag{2.1}$$

keterangan :

- w : vektor pembobot
- (x) : fungsi yang memetakan x dalam suatu dimensi
- b : bias

2.3 Fungsi Kernel

Pada kenyataannya tidak semua data itu bersifat *linier* sehingga mencari bidang pemisah secara *linier* itu menjadi sulit. Maka dari itu pada kasus *non-linier* menggunakan fungsi kernel. Fungsi ini dapat melakukan mapping pada ruang *input* dari data *training* ke *feature space* yang lebih tinggi. Ada beberapa fungsi kernel yang sering digunakan dalam *literature* SVM diantaranya kernel *Linear*, kernel *Polynomial*, kernel *Radial Basic Function* (RBF) atau biasa disebut dengan kernel *Gaussian* dan kernel *Sigmoid*. Rumus fungsi kernel tersebut terdapat pada **Tabel 1** (Septeningrum dkk., 2015)

Tabel 1 Jenis-Jenis Kernel

Nama Kernel	Fungsi
Gaussian RBF	$K(x_i, x_j) = \exp(\gamma(x_i, x_j)^2)$
<i>Linear</i>	$K(x_i, x_j) = x_i, x_j$
<i>Polynomial</i>	$K(x_i, x_j) = (x_i, x_j + c)^d$
<i>Sigmoid</i>	$K(x_i, x_j) = \tan(\sigma(x_i, x_j) + c)$

Cara kerja SVR ditentukan oleh jenis fungsi kernel yang akan dipakai dan pengaturan parameter kernel. Maka pada penelitian ini peneliti akan menggunakan RBF sebagai fungsi kernel SVR.

2.4 Normalisasi Data

Normalisasi data dilakukan bertujuan agar data berada pada rentang yang sama sehingga kompleksitas perhitungan berkurang. Kemudian normalisasi data bertujuan untuk mengurangi tingkatan error dalam komputasi. Normalisasi merupakan proses penskalaan nilai atribut dari suatu data tertentu menjadi nilai dalam rentang tertentu. Normalisasi dilakukan bertujuan untuk mengurangi adanya kesalahan pada proses data mining. Pada umumnya teknik normalisasi ini dapat dibagi kedalam 5 jenis yaitu : 1) *Min-Max*, 2) *Z-Score*, 3) *Decimal Scaling*, 4) *Sigmoidal*, 5) *Softmax* (Wirawan I Nyoman Trisna & Eksistiyanto Ivan, 2015)

Metode *Min-Max* merupakan metode paling sederhana dengan melakukan transformasi *linier* terhadap data asli, kelebihan dari metode ini adalah keseimbangan nilai perbandingan antara nilai sebelum melewati proses normalisasi dan nilai setelah melewati proses normalisasi. Pada penelitian ini menggunakan normalisasi data *min-max* dengan rentang *min-max* 0,1 sampai 1,0 (Luqyana dkk., 2018). Berikut merupakan formulasi perhitungan normalisasi data.

$$N = \frac{y - \min(y)}{\max(y) - \min(y)} \tag{2.2}$$

Keterangan:

- N = Nilai hasil *Normalisasi*
- y = Nilai Indeks Kualitas Udara
- $\min(y)$ = Nilai minimum Indeks Kualitas Udara
- $\max(y)$ = Nilai maksimum Indeks Kualitas Udara

2.5 Data Runtun Waktu (Time Series)

Data runtun waktu (*time series*) merupakan data yang disusun dan dikumpulkan, sepanjang waktu secara berurutan. Periode waktunya dapat berupa tahunan, kuartal, bulanan, mingguan dan harian maupun perjam. Model runtun waktu berusaha untuk memprediksi masa depan menggunakan data historis, dengan kata lain model runtun waktu mencoba melihat apa yang terjadi pada suatu kurun waktu tertentu dan menggunakan data masa lalu untuk memprediksi masa yang akan datang (Beaumont dkk., 1984)

Jika pengamatan dalam data runtun waktu dapat diprediksi dengan pasti dan tidak memerlukan lebih lanjut penyelidikan, itu disebut *time series* deterministik dan jika pengamatan hanya dapat menunjukkan struktur dari keadaan probabilistik, maka itu disebut stokastik. Dalam pemodelan analisis *time series* diasumsikan bahwa data stasioner. Data yang stasioner dikatakan jika tidak ada perubahan tren rata-rata dan perubahan varians (Ahmar dkk., 2018)

2.6 Grid Search Optimization

Grid Search Optimization digunakan untuk penentuan nilai parameter. Dalam penggunaannya, *Grid Search Optimization* biasanya diukur dengan *cross validation* pada data *training*. Oleh karena itu disarankan untuk mencoba beberapa variasi pasangan parameter pada *hyperplane*. *Cross validation* adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data *Training* dibagi secara random kedalam 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 (Han dkk., 2011)

2.7 Root Mean Square Error (RMSE)

Salah satu metode untuk menghitung hasil evaluasi suatu model kita dapat menggunakan perhitungan *Root Mean Squared Error RMSE* digunakan untuk mengukur tingkat kesalahan suatu model dalam memprediksi seperangkat nilai, dimana semakin kecil (mendekati nol) nilai *RMSE* maka hasil prediksi akan semakin akurat. Nilai kesalahan (galat) digunakan untuk mengetahui besarnya simpangan nilai dugaan terhadap nilai aktual. Perhitungan galat dengan menggunakan *Root Mean Square Error (RMSE)* dirumuskan oleh sebagai berikut:

$$R_{mse} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \tag{2.3}$$

Dimana :

y_i : Data *actual* periode t

\hat{y}_i : Data prediksi periode t

n : Jumlah data

2.8 Indeks Kualitas Udara

Indeks Kualitas Udara atau Indeks Standar Pencemaran Udara (ISPU) merupakan ukuran yang digunakan untuk menilai pencemaran udara di suatu daerah. Salah satu negara yang menggunakan ISPU sebagai satuan ukur kualitas udara adalah Indonesia. Indeks kualitas udara didefinisikan sebagai gambaran atau nilai hasil transformasi parameter-parameter individual polusi udara yang saling berhubungan, seperti SO₂, NO_x, SPM, Ox, CO menjadi satu nilai atau satu set nilai sehingga mudah dimengerti bagi masyarakat awam (Budiyono dkk. 2010).

Pencemaran udara disebabkan oleh adanya gas buang kendaraan bermotor, kegiatan industri, pemukiman penduduk dan karena sampah. Pemerintah Indonesia telah mengeluarkan beberapa kebijakan dalam upaya pengendalian pencemaran udara yang dimaksudkan untuk menekan beban pencemaran udara baik dari sumber bergerak, tidak bergerak maupun akibat asap dari kebakaran hutan dan lahan. Peraturan Pemerintah No. 14 Tahun 2020 tentang pengendalian pencemaran udara mengatur bahan pencemar yang perlu dipantau yaitu sulfur dioksida (SO₂), karbon monoksida (CO), nitrogen dioksida (NO₂), partikulat matter 2.5, partikulat matter 10 (PM10), hidrokarbon (HC), dan ozon (O₃).

Di dalam Peraturan Menteri Lingkungan Hidup dan kehutanan nomor 14 tahun 2020 pasal 1, Indeks Kualitas Udara atau Indeks Standar Pencemar Udara yang selanjutnya disingkat ISPU adalah angka yang tidak mempunyai satuan yang menggambarkan kondisi mutu udara ambien di lokasi tertentu, yang didasarkan kepada dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup lainnya. Nilai ISPU tidak mempunyai satuan. Tujuan disusunnya ISPU agar memberikan kemudahan dari keseragaman informasi mutu udara ambien kepada masyarakat di lokasi dan waktu tertentu serta sebagai bahan pertimbangan dalam melakukan upaya-upaya pengendalian pencemaran udara baik bagi pemerintah pusat maupun pemerintah daerah.

Parameter yang termasuk dalam ISPU yakni PM10, PM25, CO, SO₂, NO₂, O₃ dan HC. Kategori ISPU sendiri terbagi atas lima kategori dengan awal berstatus baik hingga terberat yaitu berbahaya dengan rentang masing-masing. Kategori ISPU dapat dilihat pada **Tabel 2**.

Tabel 2 Kategori Indeks Standar Pencemar Udara

Rentang	Kategori	Penjelasan
1-50	Baik	Tingkat mutu udara yang sangat baik, tidak memberikan efek negatif terhadap manusia, hewan dan tumbuhan.
51-100	Sedang	Tingkat mutu udara masih dapat diterima pada Kesehatan manusia, hewan dan tumbuhan.
101-200	Tidak Sehat	Tingkat mutu udara yang bersifat merugikan pada manusia, hewan dan tumbuhan.
201-300	Sangat Tidak Sehat	Tingkat mutu udara yang dapat meningkatkan resiko Kesehatan pada sejumlah segmen populasi yang terpapar.

>300

Berbahaya

Tingkat mutu udara yang dapat merugikan Kesehatan serius pada populasi dan perlu penanganan cepat.

3. Metode Penelitian

3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Dinas Lingkungan Hidup Provinsi Sulawesi Selatan. Dalam penelitian ini data yang digunakan merupakan data harian dengan periode Januari 2022 sampai Desember 2022 sebanyak 365 data.

3.2 Defenisi Operasional Peubah

1. Partikulat (PM₁₀) (X₁)

Partikel adalah pencemar udara yang dapat bercampur dengan bahan atau bentuk pencemar lainnya. Partikel secara sempit dapat diartikan sebagai bahan pencemar udara yang berbentuk padatan. Pencemar partikel dapat meliputi berbagai macam bentuk, mulai dari bentuk yang sederhana sampai dengan bentuk yang rumit atau kompleks yang kesemuanya merupakan bentuk pencemaran udara

2. Sulfur dioksida (SO₂) (X₂)

Sulfur dioksida mempunyai karakteristik bau yang tajam dan tidak terbakar di udara. Konsentrasi gas SO₂ di udara akan mulai terdeteksi oleh indera manusia jika konsentrasinya berkisar antara 0,3 – 1 ppm.

3. Karbon monoksida (CO) (X₃)

Karbon monoksida (CO) adalah suatu gas yang tidak bewarna, tidak berbau dan juga tidak memiliki rasa. Gas karbon monoksida sebagian besar berasal dari pembakaran bahan fosil dengan udara. Di daerah perkotaan yang padat lalu lintas menghasilkan gas CO yang paling tinggi dihasilkan dari gas buang kendaraan bermotor, emisi dari pabrik atau industri, pembangkit listrik. Di dalam rumah pencemaran udara oleh CO disebabkan gas untuk memasak, untuk pemanas air dan pemanas ruangan. Asap rokok juga merupakan sumber utama dari pencemaran CO ini.

4. Ozon (O₃) (X₄)

Ozon (O₃) merupakan suatu zat gas yang terkandung pada udara yang kita hirup. Ozon dapat bersifat baik atau buruk tergantung kondisi zat tersebut terbentuk. Ozon yang buruk terbentuk di dekat permukaan tanah ketika polutan (yang disebabkan oleh mobil, pembangkit listrik, industri, kilang minyak, dan pabrik kimia) bereaksi secara kimia dengan cahaya matahari. Polusi ozon lebih sering terjadi ketika bulan-bulan dengan cuaca yang hangat

3.3 Teknik Analisis Data

Adapun langkah-langkah teknik analisis yang akan dilakukan dalam penelitian ini adalah:

1. Pengumpulan data sekunder yakni data harian Indeks Kualitas Udara Kota Makassar yang diperoleh dari Dinas Lingkungan Hidup Provinsi Sulawesi Selatan dengan periode data Januari 2022 sampai Desember 2022
2. Menormalisasikan data
3. Menentukan data *training* dan data *testing*. Dalam penelitian ini jumlah data *training* dan data *testing* masing-masing sebanyak 80% - 20% (Gholamy dkk., 2018)
4. Pemodelan dengan *Support Vector Regression* pada data *Training*.
5. Penentuan nilai parameter pada data *training*. Penentuan nilai parameter C (cost), ϵ (Epsilon), dan γ (Gamma) dengan menerapkan *Grid Search Optimization*.
6. Penentuan model terbaik dengan melihat akurasi hasil ketepatan prediksi dengan perhitungan *RMSE*
7. Mengimplementasikan model terbaik pada data *test*.
8. Melakukan prediksi satu bulan kedepan.
9. Menarik kesimpulan

4. Hasil dan Pembahasan

4.1 Analisis Deskriptif

Analisis deskriptif digunakan untuk melihat gambaran umum dari data penelitian. Data yang digunakan dalam penelitian ini adalah data harian Indeks Kualitas Udara Kota Makassar dari tanggal 1 Januari 2022 sampai 31 Desember 2022 dengan total 365 data yang bersumber dari Dinas Lingkungan Hidup Provinsi Sulawesi Selatan.

Tabel 3 Analisis Deskriptif

	PM10	SO ₂	O ₃	NO ₂	CO
Rata-rata	9,91	15,34	11,75	5,26	21,02
Simpangan Baku	4,15	7,26	6,89	5,49	10,22
Minimum	2,50	2,00	0,75	0,46	0,33
Maksimum	24,46	51,00	38,83	41,92	56

4.2 Normalisasi Data

Normalisasi data adalah proses pembentukan struktur basis data sehingga sebagian besar *ambiguity* bisa dihilangkan. *Min-Max normalization* merupakan metode normalisasi dengan melakukan transformasi *linier* terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses (Nasution Darnisa dkk., 2019) Hasil dari normalisasi data dapat dilihat pada **Tabel 4**.

Tabel 4 Normalisasi Data

No	PM10	SO ₂	O ₃	NO ₂	CO
1	0.52	0.05	0.12	0.00	0
2	0.18	0.23	0.06	0.10	0.37
3	0.28	0.23	0.25	0.10	0.37
4	0.28	0.23	0.25	0.10	0.37
5	0.28	0.23	0.25	0.10	0.37
⋮	⋮	⋮	⋮	⋮	⋮
361	0.14	0.11	0.20	0.21	0.04
362	0.24	0.16	0.22	0.16	0.07
363	0.32	0.15	0.25	0.16	0.08
364	0.33	0.14	0.22	0.15	0.10
365	0.29	0.16	0.22	0.15	0.09

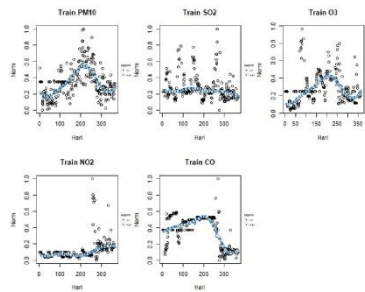
4.3 Pembagian Data *Training* dan Data *Testing*

Pada penelitian ini data yang digunakan adalah data harian Indeks Kualitas Udara di Kota Makassar yaitu dari tanggal 1 Januari 2022 – 31 Desember 2022. Sebelum data dianalisis menggunakan metode 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. Hasil pembagian data *training* dan data *testing* dapat dilihat pada lampiran. Untuk penelitian ini digunakan pembagian 80% - 20% (Gholamy dkk., 2018)

4.4 Pembahasan

1. Penerapan Metode *Support Vector Regression*

Untuk membangun model SVR pada *software* R dibutuhkan *package* tambahan yaitu e1071. Setelah menginstal *package* e1071 selanjutnya membuat model SVR pada data *train*. Setelah model didapatkan maka dilakukan prediksi pada data *train*.



Gambar 1 Plot data train dengan prediksi SVR tanpa kernel

Titik berwarna biru pada Gambar 4.1 merupakan hasil prediksi dengan parameter *default* yaitu $\epsilon = 0.1$, $C = 1$, dan $\gamma = 1$, sedangkan titik hitam merupakan data *train*. Untuk mendapatkan hasil yang maksimal diperlukan parameter terbaik dengan *error* paling minim maka diperlukan pengoptimalan model dengan pemilihan parameter yang tepat.

2. Pemodelan Model Terbaik dengan Grid Search Optimization

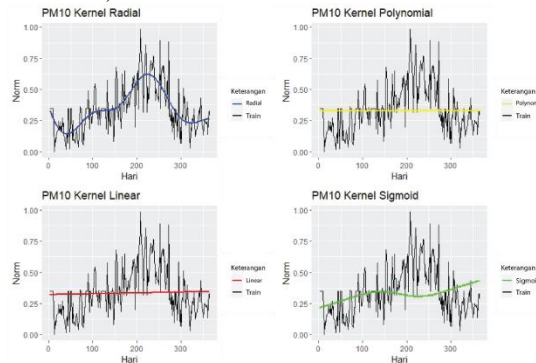
Pengoptimalan model yang akan digunakan adalah *Grid Search Optimization* dengan beberapa tipe kernel diantaranya *Radial Basis Function*, *Polynomial*, *Linear*, dan *Sigmoid*. Pengoptimalan model SVR menggunakan *Grid Search Optimization* membutuhkan parameter sesuai dengan kernelnya. *Grid Search* dijalankan dengan teknik *k-fold cross validation*.

a. Partikulat (PM10)

Tabel 5 Kernel dan Parameter variabel PM10

Kernel	Jumlah k	ϵ Terbaik	C terbaik	γ terbaik	<i>Error</i> terkecil
<i>Radial</i>	2	0,1	3	2	0.13497
	5	0,1	3	10	0.12352
	10	0,1	3	10	0.12852
<i>Polynomial</i>	2	0,1	0.05	10	0.18922
	5	0,1	0.05	10	0.18931
	10	0,1	0.1	10	0.18922
<i>Linear</i>	2	0,1	0.1	0.05	0.18779
	5	0,1	0.05	1	0.18779
	10	0,1	0.05	1	0.18779
<i>Sigmoid</i>	2	0,1	0.05	1	0.19531
	5	0,1	0.05	1	0.19531
	10	0,1	0.05	1	0.19531

Peubah Partikulat (PM10) memiliki nilai parameter terbaik ketika menggunakan kernel *Radial* dengan nilai $k = 5$, $\epsilon = 0,1$, $C = 3$, dan $\gamma = 10$ sebesar 0,11851



Gambar 2 Plot Kernel variabel PM10

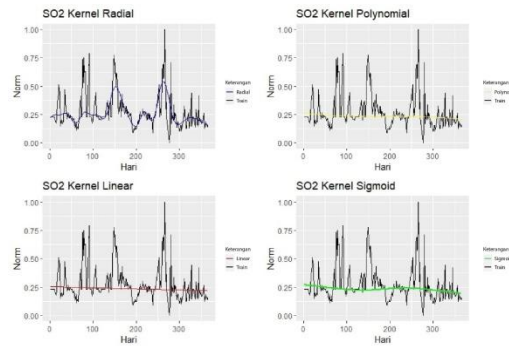
Model SVR dengan beberapa kernel pada peubah PM10 dibuat menggunakan data *training* PM10 memperoleh nilai *RMSE* terkecil pada kernel *Radial* sebesar 0.11851 yang artinya model memiliki tingkat akurasi yang baik. Visualisasi perbandingan nilai yang dihasilkan dari uji model SVR dengan data *training* PM10 dapat dilihat pada **Gambar 2**.

b. Sulfur dioksida (SO₂)

Tabel 6 Kernel dan Parameter variabel SO₂

Kernel	Jumlah <i>k</i>	ϵ Terbaik	<i>C</i> terbaik	γ terbaik	Error terkecil
<i>Radial</i>	2	0,1	1	10	0.12780
	5	0,1	100	10	0.11601
	10	0,1	100	10	0.11502
<i>Polynomial</i>	2	0,1	0.2	10	0.15469
	5	0,1	0.1	10	0.15469
	10	0,1	0.05	10	0.15468
<i>Linear</i>	2	0,1	0.05	1	0.15433
	5	0,1	0.2	1	0.15433
	10	0,1	0.2	1	0.15433
<i>Sigmoid</i>	2	0,1	0.1	1	0.18467
	5	0,1	0.05	1	0.15618
	10	0,1	0.05	1	0.15618

Sulfur dioksida (SO₂) memiliki nilai parameter terbaik ketika menggunakan kernel *Radial* dengan nilai *k* = 10, ϵ = 0,1, *C* = 100, dan γ = 10 sebesar 0.11502



Gambar 3 Plot Kernel variabel SO₂

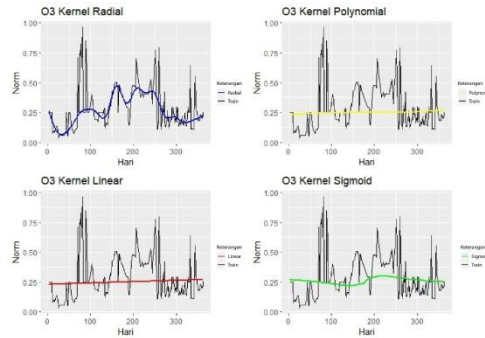
Model SVR dengan beberapa kernel pada peubah SO₂ dibuat menggunakan data *training* SO₂ memperoleh nilai *RMSE* terkecil pada kernel *Radial* sebesar 0.11502 yang artinya model memiliki tingkat akurasi yang baik. Visualisasi perbandingan nilai yang dihasilkan dari uji model SVR dengan data *training* SO₂ dapat dilihat pada **Gambar 3**

c. Ozon (O₃)

Tabel 7 Kernel dan Parameter variabel O₃

Kernel	Jumlah <i>k</i>	ϵ Terbaik	<i>C</i> terbaik	γ terbaik	Error terkecil
<i>Radial</i>	2	0,1	10	10	0.13561
	5	0,1	3	10	0.13852
	10	0,1	10	10	0.13581
<i>Polynomial</i>	2	0,1	0.15	10	0.18095
	5	0,1	0.05	10	0.18096
	10	0,1	0.05	10	0.18096
<i>Linear</i>	2	0,1	0.05	1	0.18059
	5	0,1	10	1	0.18059
	10	0,1	10	1	0.18097
<i>Sigmoid</i>	2	0,1	0.05	10	0.17859
	5	0,1	0.05	10	0.17859
	10	0,1	0.05	10	0.17895

Ozon (O₃) memiliki nilai parameter terbaik ketika menggunakan kernel *Radial* dengan nilai *k* = 2, ϵ = 0,1, *C* = 10, dan γ = 10 sebesar 0.13561



Gambar 4 Plot Kernel variabel O_3

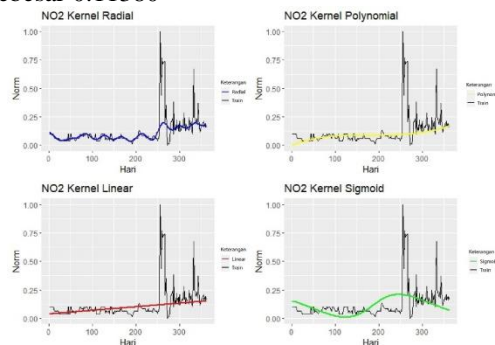
Model SVR dengan beberapa kernel pada peubah O_3 dibuat menggunakan data *training* O_3 memperoleh nilai *RMSE* terkecil pada kernel *Radial* sebesar 0,13561 yang artinya model memiliki tingkat akurasi yang baik. Visualisasi perbandingan nilai yang dihasilkan dari uji model SVR dengan data *training* O_3 dapat dilihat pada **Gambar 4**.

d. Nitrogen dioksida (NO_2)

Tabel 8 Kernel dan Parameter variabel NO_2

Kernel	Jumlah k	ϵ Terbaik	C terbaik	γ terbaik	<i>Error</i> terkecil
<i>Radial</i>	2	0,1	3	1	0.12560
	5	0,1	100	10	0.11383
	10	0,1	100	10	0.11380
<i>Polynomial</i>	2	0,1	0.2	10	0.13179
	5	0,1	0.2	10	0.13179
	10	0,1	0.1	10	0.13180
<i>Linear</i>	2	0,1	0.5	1	0.12742
	5	0,1	3	1	0.12742
	10	0,1	3	1	0.12742
<i>Sigmoid</i>	2	0,1	0.05	1	0.13210
	5	0,1	0.05	1	0.13210
	10	0,1	0.05	1	0.13210

Nitrogen dioksida (NO_2) memiliki nilai parameter terbaik ketika menggunakan kernel *Radial* dengan nilai $k = 10$, $\epsilon = 0,1$, $C = 100$, dan $\gamma = 10$ sebesar 0.11380



Gambar 4 Plot Kernel variabel NO_2

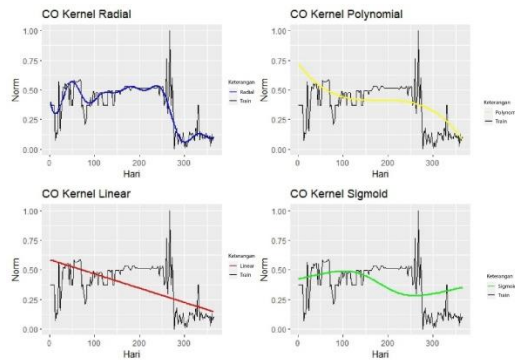
Model SVR dengan beberapa kernel pada peubah NO_2 dibuat menggunakan data *training* NO_2 memperoleh nilai *RMSE* terkecil pada kernel *Radial* sebesar 0,11380 yang artinya model memiliki tingkat akurasi yang baik. Visualisasi perbandingan nilai yang dihasilkan dari uji model SVR dengan data *training* NO_2 dapat dilihat pada **Gambar 5**.

e. Karbon Monoksida (CO)

Tabel 9 Kernel dan Parameter variabel CO

Kernel	Jumlah k	ϵ Terbaik	C terbaik	γ terbaik	Error terkecil
<i>Radial</i>	2	0,1	100	10	0.00783
	5	0,1	100	10	0.00699
	10	0,1	100	10	0.00741
<i>Polynomial</i>	2	0,1	0.2	10	0.03048
	5	0,1	0.2	10	0.03105
	10	0,1	0.15	10	0.03108
<i>Linear</i>	2	0,1	0.2	1	0.03088
	5	0,1	0.1	1	0.03133
	10	0,1	0.05	1	0.02901
<i>Sigmoid</i>	2	0,1	0.05	1	0.03165
	5	0,1	0.05	1	0.03412
	10	0,1	0.05	1	0.03348

Karbon Monoksida (CO) memiliki nilai parameter terbaik ketika menggunakan kernel *Radial* dengan nilai $k = 5$, $\epsilon = 0,1$, $C = 100$, dan $\gamma = 10$ sebesar 0.0069



Gambar 5 Plot Kernel variabel CO

Model SVR dengan beberapa kernel pada peubah CO dibuat menggunakan data *training* CO memperoleh nilai *RMSE* terkecil pada kernel *Radial* sebesar 0,00699 yang artinya model memiliki tingkat akurasi yang baik. Visualisasi perbandingan nilai yang dihasilkan dari uji model SVR dengan data *training* CO dapat dilihat pada **Gambar 6**.

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 10**.

Tabel 10 RMSE tiap kernel

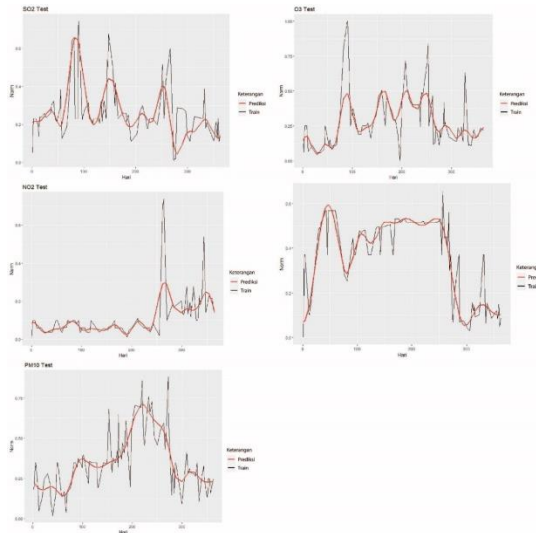
	PM10	SO ₂	O ₃	NO ₂	CO
<i>Radial</i>	0.11851	0.11502	0.13561	0.11380	0.00699
<i>Polynomial</i>	0.18922	0.15468	0.18095	0.13179	0.03048
<i>Linear</i>	0.18779	0.15433	0.18059	0.12742	0.02901
<i>Sigmoid</i>	0.19531	0.15618	0.17859	0.13210	0.03165

Dari hasil pada **Tabel 10** nilai *RMSE* terkecil pada setiap variabel terletak pada saat menggunakan kernel *Radial* masing-masing untuk variabel PM10 dengan nilai *RMSE* 0,11851, variabel SO₂ dengan nilai *RMSE* 0.11502, variabel O₃ dengan nilai *RMSE* 0,13561, variabel NO₂ dengan nilai *RMSE* 0,11380, dan variabel CO dengan nilai *RMSE* 0,11380

4. Implementasi pada Data Testing

Setelah mendapatkan model terbaik pada langkah sebelumnya. Selanjutnya, memasukkan model tersebut beserta parameternya pada data *testing*. Hasil prediksinya dapat dilihat pada gambar dibawah.

Pada **Gambar 4.7** terlihat hasil prediksi pada data *testing* berupa garis berwarna merah sedangkan data *testing* ditunjukkan dengan garis berwarna hitam. Dapat dilihat pada **Gambar 7** bahwa data hasil nilai prediksi mendekati nilai data aktual.



Gambar 6 Plot data test

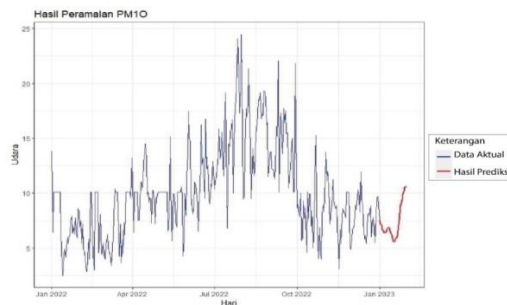
5. Hasil Prediksi Indeks Kualitas Udara di Kota Makassar

Hasil prediksi Indeks Kualitas Udara di Kota Makassar dari bulan tanggal 1 Januari 2023 – 31 Januari 2023 dengan model terbaik dapat dilihat pada **Tabel 11** terlihat indeks kualitas udara pada tanggal 1 Januari 2023 – 13 Maret 2023.

Tabel 11 Hasil Prediksi Indeks Kualitas Udara

Hari	PM10	SO ₂	O ₃	NO ₂	CO
1	7.526362	12.299127	4.782312	4.478085	1.346156
2	7.192237	12.15826	4.765154	4.345071	2.556828
3	7.089332	12.041041	4.360071	4.187222	9.088221
4	6.66742	12.191655	4.317654	4.0075	18.566296
5	6.60518	12.273768	3.679758	3.810933	23.459415
6	6.405105	12.373118	3.403996	3.604451	26.111957
7	6.408976	12.620231	3.726115	3.396502	29.989219
8	6.460192	12.923629	4.206734	3.196467	30.705151
9	6.7062	14.201282	4.944546	3.013933	31.877896
10	6.842125	14.698808	5.409989	2.857914	32.336823
⋮	⋮	⋮	⋮	⋮	⋮
29	10.539122	20.277775	18.320363	3.861033	24.969218
30	10.591695	18.565366	18.724555	3.738155	26.992035
31	10.670294	14.890655	18.871156	3.573235	28.442573

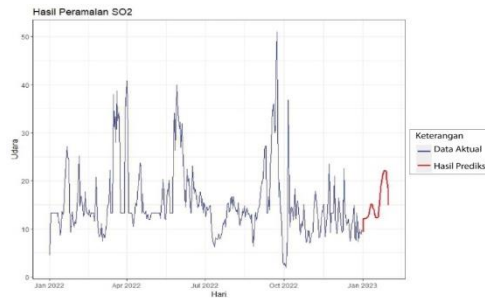
a. Partikulat (PM10)



Gambar 7 Hasil Peramalan PM10

Plot grafik hasil peramalan PM10 di Kota Makassar mengalami penurunan hingga 5,58 pada pertengahan Januari 2023 kemudian mengalami kenaikan sampai akhir Januari 2023 hingga 10,67. Konsentrasi PM10 memiliki interval 5,58 – 10,67 dan berdasarkan kategori Indeks Standar Pencemaran Udara berada pada indeks kategori baik. Plot hasil peramalan PM10 dapat dilihat pada **Gambar 8**.

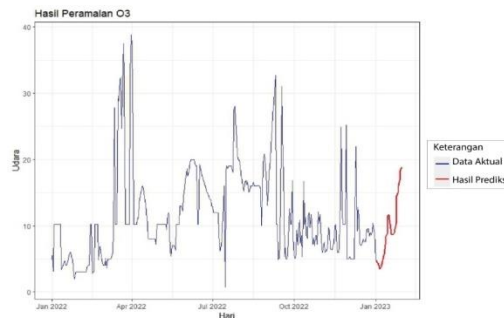
b. Sulfur dioksida (SO₂)



Gambar 8 Hasil Peramalan SO₂

Plot grafik hasil peramalan SO₂ di Kota Makassar mengalami kenaikan hingga 15,17 pada pertengahan Januari 2023 kemudian mengalami fluktuasi sampai akhir Januari 2023. Konsentrasi SO₂ memiliki interval 9,29 – 22,18 dan berdasarkan kategori ISPU berada pada indeks kategori baik. Plot hasil peramalan SO₂ dapat dilihat pada **Gambar 9**.

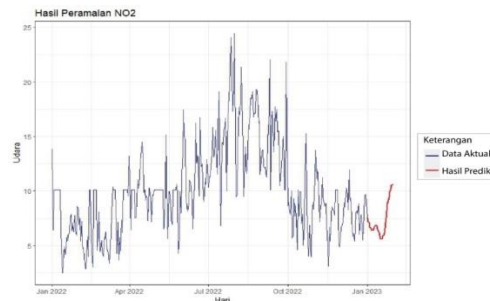
c. Ozon (O₃)



Gambar 9 Hasil Peramalan O₃

Plot grafik hasil peramalan O₃ di Kota Makassar mengalami kenaikan hingga 18,87 pada akhir Januari 2023. Konsentrasi ISPU memiliki interval 3,40 – 18,87 dan berdasarkan kategori ISPU berada pada indeks kategori baik. Plot hasil peramalan O₃ dapat dilihat pada **Gambar 10**.

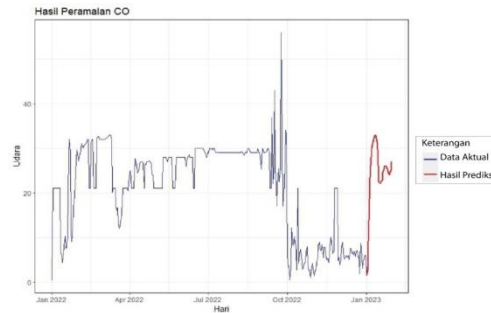
d. Nitrogen Dioksida (NO₂)



Gambar 10 Hasil Peramalan NO₂

Plot grafik hasil peramalan NO_2 di Kota Makassar mengalami penurunan hingga 15,17 pada pertengahan Januari 2023 kemudian mengalami fluktuasi sampai akhir Januari 2023. Konsentrasi ISPU memiliki interval 9,29 – 22,18 dan berdasarkan kategori ISPU berada pada indeks kategori baik. Plot hasil peramalan NO_2 dapat dilihat pada **Gambar 11**.

e. Karbon Monoksida (CO)



Gambar 11 Hasil Peramalan CO

Plot grafik hasil peramalan CO di Kota Makassar mengalami kenaikan hingga 33,01 pada pertengahan Januari 2023 kemudian mengalami fluktuasi sampai akhir Januari 2023. Konsentrasi ISPU memiliki interval 1,34 – 33,01 dan berdasarkan kategori ISPU berada pada indeks kategori baik. Plot hasil peramalan CO dapat dilihat pada **Gambar 12**.

5. Kesimpulan

Berikut ini adalah kesimpulan dari peramalan Indeks Kualitas Udara di Kota Makassar dengan metode SVR.

1. Berdasarkan penelitian dan pembahasan yang telah dilakukan untuk meramalkan indeks kualitas udara di kota Makassar dengan metode SVR diperoleh kesimpulan bahwa model terbaik yang digunakan dari kelima variabel adalah dengan menggunakan kernel radial dengan parameter dan nilai RMSE masing-masing sebagai berikut :
 - a) Partikulat (PM10) $\epsilon = 0,1$, $C = 3$, dan $\gamma = 10$ dengan hasil akurasi RMSE sebesar 0,12352
 - b) Sulfur Dioksida (SO_2) $\epsilon = 0,1$, $C = 100$, dan $\gamma = 10$ dengan hasil akurasi RMSE sebesar 0,11502
 - c) Ozon (O_3) $\epsilon = 0,1$, $C = 10$, dan $\gamma = 10$ dengan hasil akurasi RMSE sebesar 0,13561
 - d) Nitrogen dioksida (NO_2) $\epsilon = 0,1$, $C = 100$, dan $\gamma = 10$ dengan hasil akurasi RMSE sebesar 0,11380
 - e) Karbon Monoksida (CO) $\epsilon = 0,1$, $C = 100$, dan $\gamma = 10$ dengan hasil akurasi RMSE sebesar 0,00699
2. Adapun hasil prediksi dari Indeks Kualitas Udara di Kota Makassar yakni sebagai berikut
 - a. Peubah PM10 cenderung naik sedangkan peubah SO_2 , O_3 , NO_2 , dan CO cenderung fluktuatif sampe akhir Januari 2023
 - b. Berdasarkan hasil prediksi ISPU di Kota Makassar dengan rentang nilai 1 – 33, kualitas udara tergolong baik

References

- Ahmar, A. S., Guritno, S., Abdurakhman, Rahman, A., Awi, Alimuddin, Minggu, I., Tiro, M. A., Aidid, M. K., Annas, S., Sutiksno, D. U., Ahmar, D. S., Ahmar, K. H., Ahmar, A. A., Zaki, A., Abdullah, D., Rahim, R., Nurdianto, H., Hidayat, R., ... Arifin, A. N. M. (2018). Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO). *Journal of Physics: Conference Series*, 954(1), 1–12.
- Aini, N., Ruktiari, R., Pratama, R., Fitrah Buana, A., & Lunak, R. P. (2019). Sistem Prediksi Tingkat Pencemaran Polusi Udara dengan Algoritma Naïve Bayes di Kota Makassar. *Journal Kominfo*, 3, 83–90.
- Aswi, & Sukarna. (2006). *Analisis Deret Waktu: Teori dan Aplikasi*. Andira Publisher.
- Beaumont, C., Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1984). *Forecasting: Methods and Applications*. Dalam *The Journal of the Operational Research Society* (Vol. 35, Nomor 1). JSTOR.
- Budiyono, A. (2003). Indeks kualitas udara. *Journal lapan*, 3(1), 6–9.

- Caraka, R. E., Yasin, H., & Basyiruddin, A. W. (2017). Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis Related papers. *Jurnal Matematika*, 7(1), 1693–1394.
- Fajarita, L., & Hati, E. N. (2018). Penerapan forecasting stright line method dalam pengadaan stok barang mendatang studi kasus : PT. Bina Karya Kusuma. *Proosiding SINTAK 2018*, 2, 310–317.
- Furi, R. P., Si, M., & Saepudin, D. (2015). Prediksi Financial Time Series Menggunakan Independent Component Analysis dan Support Vector Regression Studi Kasus : IHSG dan JII. *ISSN: 2355-9365 e-Proceeding of Engineering* :, 2(2), 1–10.
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). A Pedagogical Explanation A Pedagogical Explanation Part of the Computer Sciences Commons. *Departmental Technical Reports*, 1(1–6).
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann.
- Haque, I. I., Kallista, M., & Setianingsih, C. (2022). Prediksi Indeks Standar Pencemar Udara Menggunakan Metode Support Vector Regression Berbasis Web (Studi Kasus Pada Dki Jakarta). *e-Proceeding of Engineering*, 9, 1256–1266.
- Isnaeni, R., Sudarmin, & Rais, Z. (2022). Analisis Support Vector Regression (SVR) dengan Kernel Radial Basis Function (RBF) untuk Memprediksi Laju Inflasi di Indonesia. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(1), 30–38. <https://doi.org/10.35580/variansium13>
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4704–4713. <http://j-ptiik.ub.ac.id>
- Nasution Darnisa, A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma K-NN. *Journal of Computer Engineering System and Science*, 4(1), 2502–7131.
- Peraturan Pemerintah RI. (2020). *Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia No 14 Tahun 2020 tentang Indeks Standar Pencemaran Udara*. 1–16.
- Prahardis, R., Syauqi, D., & Akbar, S. R. (2018). Implementasi Sistem Monitoring Polusi Udara Berdasarkan Indeks Standar Pencemaran Udara Dengan Pemodelan Finite State Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(9), 3128–3137.
- Scholkopf, B., & Smola, A. J. (2018). *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press.
- Septiningrum, L., Yasin, H., & Sugito. (2015). Prediksi Indeks Harga Saham Gabungan Menggunakan Support Vector Regression (SVR) dengan Algoritma grid search. *Jurnal Gaussian*, 4(2), 315–321.
- Siregar, N. A. (2022). Peramalan Curah Hujan Di Kota Medan menggunakan Metode Support Vector Regression. *Journal of Informatics and Data Science (J-IDS)*, 1(1), 1–3.
- Wirawan I Nyoman Trisna, & Eksistyanto Ivan. (2015). PENERAPAN NAIVE BAYES PADA INTRUSION DETECTION SYSTEM DENGAN DISKRITISASI VARIABEL. *Jurnal Ilmiah Teknologi Informasi*, 13(2), 182–189.