

Penilaian Kualitas Udara dan Polusi Menggunakan Algoritma *Support Vector Machine*

*Ridho Sholehurrohman¹, Ivani Valentine²

¹Jurusan Ilmu Komputer, Universitas Lampung

Jl. Soemantri Brojonegoro No. 1, Kota Bandar Lampung, Provinsi Lampung, 35145, Indonesia

²Jurusan Magister Teknik Informatika, Institut Informatika dan Bisnis Darmajaya

Jl. ZA. Pagar Alam No.93, Kota Bandar Lampung, Provinsi Lampung, 35141, Indonesia

Email: ¹ridho.sholehurrohman@fmipa.unila.ac.id*, ²ivani.valentin@gmail.com

ABSTRACT

Good air quality is essential for healthy and sustainable living, but increasing air pollution caused by industrialization, urbanization, and motor vehicles has become a serious global issue. Air pollution negatively affects health, the environment, and the quality of life, making air quality monitoring and assessment a priority. The complexity of air quality data renders conventional analytical approaches less effective; therefore, machine learning methods such as Random Forest and Neural Networks have been applied to address these challenges. However, these methods have limitations in handling non-linear patterns or computational efficiency. This study employs the Support Vector Machine (SVM) algorithm with various kernels to classify air quality based on pollution and environmental parameters into categories of Good, Moderate, Poor, and Hazardous. The results indicate that the Polynomial Kernel performs best for the Good category, while the RBF Kernel is also competitive but less optimal for the Hazardous and Poor categories. With parameter optimization using GridSearchCV, the combination of $C=10$ and $\Gamma=0.1$ or scale achieved an average accuracy of 90.75%. CO concentration and proximity to industrial areas proved to be significant features in classification. This study aims to support air pollution management and mitigate its impacts on society.

Keywords : air quality; air pollution; SVM algorithm

ABSTRAK

Kualitas udara yang baik sangat penting untuk kehidupan yang sehat dan berkelanjutan, tetapi meningkatnya polusi udara akibat industrialisasi, urbanisasi, dan kendaraan bermotor menjadi masalah global yang serius. Polusi udara berdampak buruk pada kesehatan, lingkungan, dan kualitas hidup masyarakat, sehingga pemantauan dan penilaian kualitas udara menjadi prioritas. Kompleksitas data pada kualitas udara membuat pendekatan analitik konvensional kurang efektif, sehingga metode machine learning seperti *Random Forest* dan *Neural Networks* telah diterapkan untuk mengatasi tantangan tersebut. Namun, metode ini memiliki keterbatasan dalam menangani pola non-linear atau efisiensi komputasi. Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dengan berbagai *kernel* untuk mengklasifikasikan kualitas udara berdasarkan parameter polusi dan lingkungan ke dalam kategori *Good*, *Moderate*, *Poor*, dan *Hazardous*. Hasil penelitian menunjukkan *Polynomial Kernel* memberikan performa terbaik pada kategori *Good*, sedangkan *RBF Kernel* juga kompetitif meskipun kurang optimal pada kategori *Hazardous* dan *Poor*. Dengan optimasi parameter menggunakan *GridSearchCV*, kombinasi $C=10$ dan $\Gamma=0.1$ atau scale menghasilkan akurasi rata-rata 90,75%. Konsentrasi CO dan kedekatan dengan area industri terbukti sebagai fitur signifikan dalam klasifikasi. Penelitian ini diharapkan dapat mendukung pengelolaan polusi udara dan mitigasi dampaknya terhadap masyarakat.

Kata kunci : kualitas udara; polusi udara; algoritma SVM

1. PENDAHULUAN

Kualitas udara yang baik merupakan salah satu faktor penting untuk menunjang kehidupan yang sehat dan berkelanjutan. Namun, peningkatan aktivitas manusia seperti industrialisasi, urbanisasi, dan penggunaan kendaraan bermotor telah menyebabkan polusi udara menjadi salah satu isu global yang serius (Chaudhary, M.T, 2022). Menurut laporan WHO, lebih dari 90% populasi dunia tinggal di area dengan tingkat polusi udara yang melebihi ambang batas yang direkomendasikan (B.M. Cintoro et al., 2023). Dampak buruk dari polusi udara mencakup gangguan kesehatan seperti penyakit pernapasan dan kardiovaskular, kerusakan lingkungan, hingga penurunan kualitas hidup Masyarakat (IQAIR, 2023).

Selama beberapa dekade terakhir, pemantauan dan penilaian kualitas udara telah menjadi prioritas untuk mendeteksi dan mencegah dampak negatif polusi udara (Chaudhary, M.T, 2022). Indikator seperti PM2.5, PM10, NO2, SO2, dan CO sering digunakan untuk mengukur tingkat polusi. Namun, data kualitas udara cenderung kompleks, bersifat non-linear, dan dipengaruhi oleh berbagai

faktor lingkungan seperti suhu, kelembapan, jarak ke area industri, dan kepadatan populasi (Biao Sun et al, 2023). Hal ini membuat pendekatan analitik konvensional, seperti regresi linier, kurang efektif dalam menangkap pola-pola yang kompleks (Addis Bikis et al, 2023).

Berbagai penelitian telah mencoba mengatasi masalah ini dengan metode *machine learning* (Julien Vachonet al., 2024). Algoritma seperti *Random Forest* dan *Neural Networks* sering digunakan untuk klasifikasi kualitas udara (Julien Vachon et al., 2024). Meskipun *Random Forest* mampu memberikan hasil yang baik pada dataset yang besar dan heterogen, algoritma ini kurang efisien dalam menangani data dengan pola non-linear yang kompleks (Kamel Maaloul and Lejdel Brahim, 2022). Di sisi lain, *Neural Networks* dapat menangkap pola non-linear, tetapi sering memerlukan data dalam jumlah besar dan sumber daya komputasi yang tinggi (Abdulwaheed Tella et al., 2021).

Untuk mengatasi keterbatasan tersebut, penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) sebagai pendekatan utama (E.A. Zanaty, 2012). SVM telah terbukti

efektif dalam menangani hubungan non-linear melalui fungsi *kernel* seperti RBF (*Radial Basis Function*) dan *Polynomial* (Rameswar Debnath, and Haruhisa Takahashi, 2004). Dengan memetakan data ke dalam ruang dimensi yang lebih tinggi, SVM mampu menemukan batas keputusan optimal antara kelas-kelas data yang kompleks (Wardatus Sholihahet al., 2023). Algoritma ini juga lebih efisien dibandingkan metode lain dalam memproses dataset dengan ukuran sedang (Wardatus Sholihahet al., 2023).

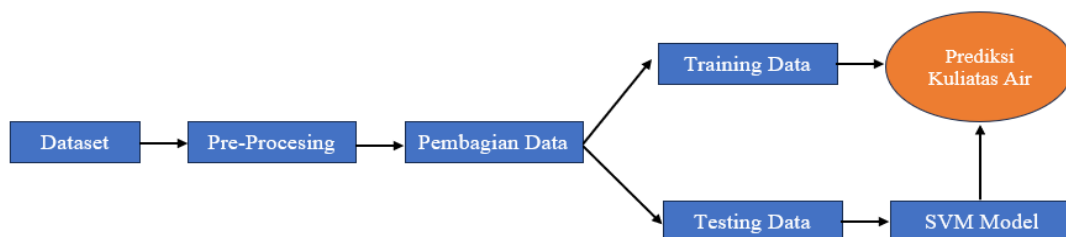
Penelitian ini bertujuan untuk mengembangkan sistem penilaian dan klasifikasi kualitas udara berdasarkan data polusi dan parameter lingkungan. Dengan memanfaatkan algoritma SVM, sistem ini akan mengklasifikasikan kualitas udara ke dalam kategori seperti *Good*, *Moderate*, *Poor*, dan *Hazardous*. Penelitian ini diharapkan tidak hanya menghasilkan model yang akurat, tetapi juga menjadi dasar pengambilan keputusan dalam pengelolaan polusi

udara dan mitigasi dampaknya terhadap Masyarakat.

2. METODE

2.1 Pengumpulan Dataset

Penelitian ini menggunakan Python versi 3.7 sebagai bahasa pemrograman untuk mengimplementasikan model prediksi kesehatan tubuh umum diagnostik. Selanjutnya, dataset yang digunakan adalah data kualitas udara diberbagai wilayah yang berisi 5000 sampel. Detail data mencakup konsentrasi polutan seperti PM2.5, PM10, NO2, SO2, dan CO, serta parameter lingkungan seperti *Temperature*, *Humidity*, *Proximity to Industrial Areas*, dan *Population Density*. Data ini diklasifikasikan ke dalam empat kategori kualitas udara, yaitu *Good*, *Moderate*, *Poor*, dan *Hazardous*. Dataset tersebut diperoleh dari tautan <https://www.kaggle.com/datasets/mujtabamatin/air-quality-and-pollution-assessment>. Berikut Gambar 1 alur metode penelitian.



Gambar 1. Metode Penelitian

2.2 Preprocessing dan Pembagian Data

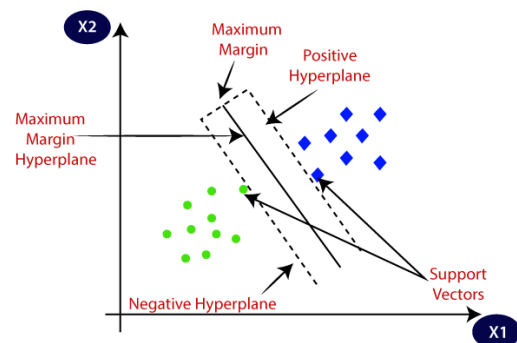
Sebelum digunakan, data dilakukan preprocessing yang meliputi normalisasi menggunakan *Standard-Scaler* untuk memastikan kontribusi fitur yang seimbang terhadap model (Schoëlkopf B and Smola AJ., 2002). Selain itu, jika distribusi kelas target tidak seimbang, metode SMOTE (*Synthetic Minority Oversampling Technique*) digunakan untuk meningkatkan representasi kelas minoritas (Schoëlkopf BS et al., 1999). Data kemudian dibagi menjadi 80% data pelatihan dan 20% data pengujian dengan stratifikasi untuk menjaga proporsi kategori.

2.3 Support Vector Machine (SVM)

Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) untuk membangun model klasifikasi. Berbagai fungsi *kernel* seperti *Linear*, *Polynomial*, *RBF* (*Radial Basis Function*), dan *Sigmoid* diuji untuk menentukan *kernel* terbaik berdasarkan akurasi klasifikasi (Burges CJ, 1997). Parameter-parameter seperti *C*, *gamma*, dan *degree* (untuk *kernel Polynomial*) dioptimalkan menggunakan teknik *GridSearchCV* untuk menemukan

kombinasi yang paling optimal (Schoëlkopf B et al, 1998). SVM dipilih karena kemampuannya menangani hubungan non-linear antara fitur dengan memetakan data ke ruang dimensi yang lebih tinggi melalui fungsi *kernel* (Burges CJ et al, 1997)

Selanjutnya, Gambar 2 menunjukkan prinsip dasar *Support Vector Machine* (SVM) yaitu algoritma *machine learning* berbasis margin yang digunakan untuk klasifikasi (Vishal Sharma et al., 2023). SVM bekerja dengan mencari *hyperplane* yang memaksimalkan margin antara dua kelas data (dalam gambar ini, kelas hijau dan biru), sehingga menghasilkan generalisasi yang baik untuk data baru (Vishal Sharma et al., 2023).



Gambar 2. Algoritma SVM
(Sumber : javatpoint.com, 2023)

Hyperplane didefinisikan oleh *support vectors*, yaitu titik-titik data yang paling dekat dengan margin. Keberadaan *support vectors* sangat

penting karena menentukan posisi *hyperplane* yang optimal dan mempengaruhi kemampuan klasifikasi model (Pedro Guerra et al, 2006).

Dalam kasus Gambar 2, memisahkan dua kategori data, hijau dan biru, berdasarkan fitur x_1 dan x_2 . SVM mencari *hyperplane* (garis pemisah) yang memaksimalkan jarak atau margin antara kedua kategori, dengan *support vectors* sebagai titik data terdekat yang menentukan posisi *hyperplane* (Vishal Sharma et al., 2023).

Hyperplane keputusan berada di tengah margin, sedangkan garis *Positive Hyperplane* dan *Negative Hyperplane* menunjukkan batas margin untuk setiap kategori. Terlihat bahwa model dapat mengklasifikasikan pasangan koordinat (x_1, x_2) dalam warna hijau atau biru dengan memaksimalkan margin (Pedro Guerra et al, 2006). Jika data tidak dapat

dipisahkan secara *linier*, SVM menggunakan fungsi *kernel* untuk memetakan data ke dimensi yang lebih tinggi, sehingga pemisahan menjadi lebih mudah dan akurat (Vishal Sharma et al., 2023).

3. HASIL DAN PEMBAHASAN

Studi ini mengevaluasi algoritma SVM yang melakukan analisis komprehensif menggunakan beberapa metrik kinerja. Metrik ini mengukur waktu pelatihan dan pengujian, akurasi, presisi, *recall*, F1- skor, *Area Under the Curve* (AUC), *Receiver Operating Characteristic* (ROC), dan validasi silang. Penulis juga mencermati bahwa seluruh proses konfigurasi, parameter untuk setiap *kernel* ditetapkan ke *default*.

3.1. Dataset

Langkah pertama, dideskripsikan data awal pada dataset seperti Tabel 1 berikut.

Tabel 1. Deskripsi Tampilan 5 baris pertama Dataset

<i>Temperature</i>	<i>Humidity</i>	<i>PM2.5</i>	<i>PM10</i>	<i>NO2</i>	<i>SO2</i>	<i>CO</i>	<i>Proximity to Industrial Areas</i>	<i>Population Density</i>	<i>Air Quality</i>
29.8	59.1	5.2	17.9	18.9	9.2	1.72	6.3	319	<i>Moderate</i>
28.3	75.6	2.3	12.2	30.8	9.7	1.64	6.0	611	<i>Moderate</i>
23.1	74.7	26.7	33.8	24.4	12.6	1.63	5.2	619	<i>Moderate</i>
27.1	39.1	6.1	6.3	13.5	5.3	1.15	11.1	551	<i>Good</i>

Tabel 1 menunjukkan data awal mengenai faktor lingkungan dan metrik kualitas udara untuk berbagai area. Suhu

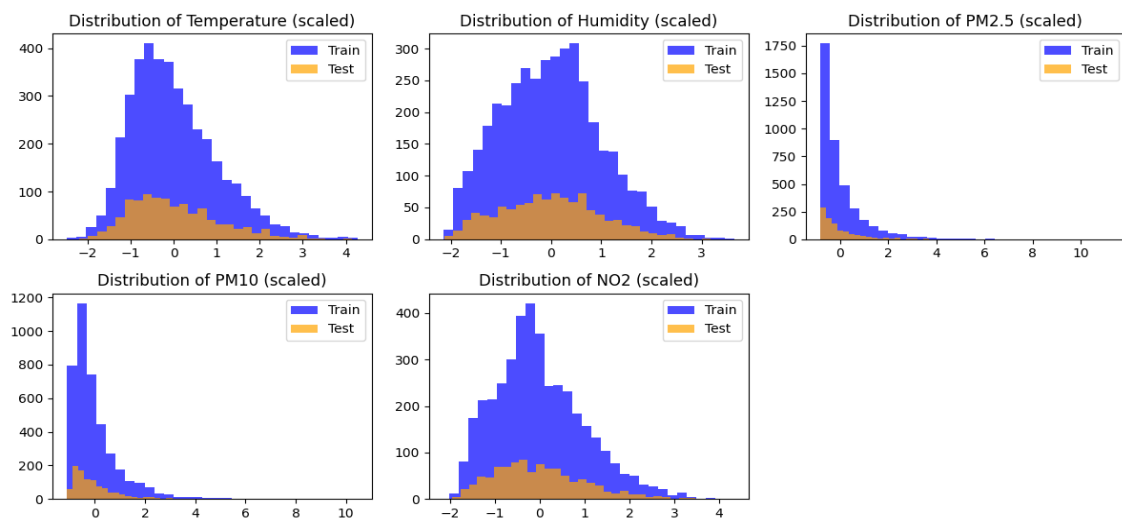
berkisar antara 23,1°C hingga 29,8°C, sementara kelembaban berada di antara 39,1% dan 75,6%. Partikel debu PM2.5

tercatat antara 2,3 hingga 26,7 $\mu\text{g}/\text{m}^3$, dan PM10 antara 6,3 hingga 33,8 $\mu\text{g}/\text{m}^3$. Konsentrasi nitrogen dioksida (NO₂) berkisar antara 13,5 hingga 30,8 $\mu\text{g}/\text{m}^3$, sedangkan sulfur dioksida (SO₂) antara 5,3 hingga 12,6 $\mu\text{g}/\text{m}^3$. Karbon dioksida (CO) tercatat antara 1,15 hingga 1,72 ppm. Jarak ke area industri bervariasi antara 5,2 km hingga 11,1 km, dan kepadatan penduduk antara 319 hingga 619 orang per km². Kualitas udara sebagian besar dikategorikan sebagai "Moderate," dengan satu entri menunjukkan kualitas "Good." Metrik

ini memberikan gambaran kondisi lingkungan dan faktor-faktor yang mempengaruhi kualitas udara di setiap area.

3.2 Visualisasi Preprocessing Data

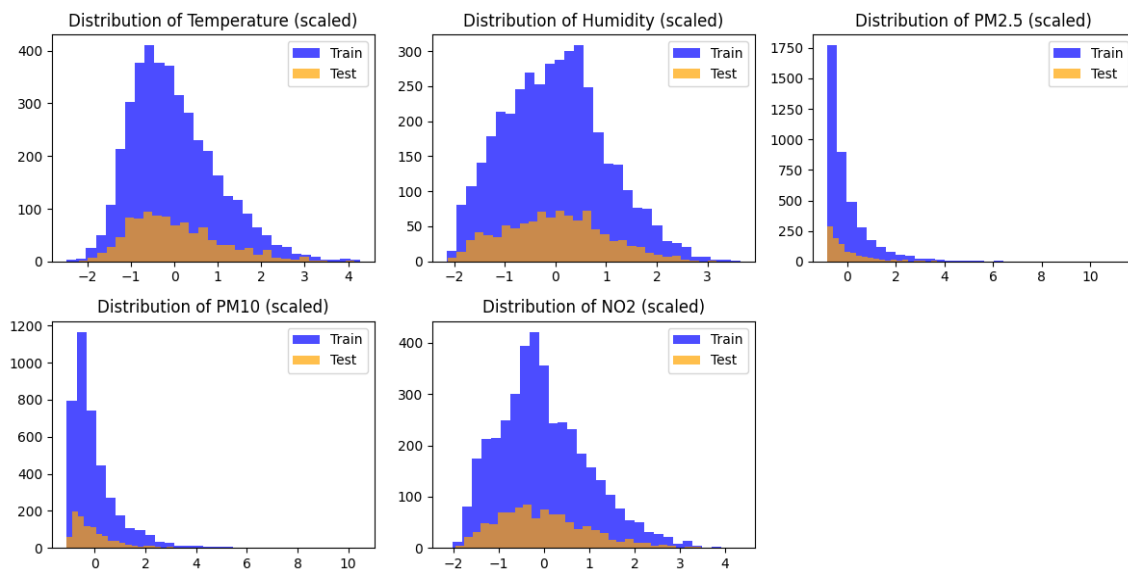
Selanjutnya, untuk memberikan gambaran awal terkait data yang akan digunakan, diberikan visualisasi sebelum dilakukan proses *pre-processing* dataset. Visualisasi ini bertujuan untuk membantu memahami pola, distribusi, atau potensi masalah pada data mentah sebelum tahap pengolahan lebih lanjut dilakukan.



Gambar 3. Visualisasi data sebelum *pre-processing*

Gambar 3 menunjukkan distribusi parameter lingkungan yang sebagian besar konsisten antara data pelatihan dan pengujian. Temperatur, kelembaban, dan NO₂ mendekati distribusi normal, sementara PM_{2.5} dan PM₁₀ skewed ke kanan. Distribusi diatas

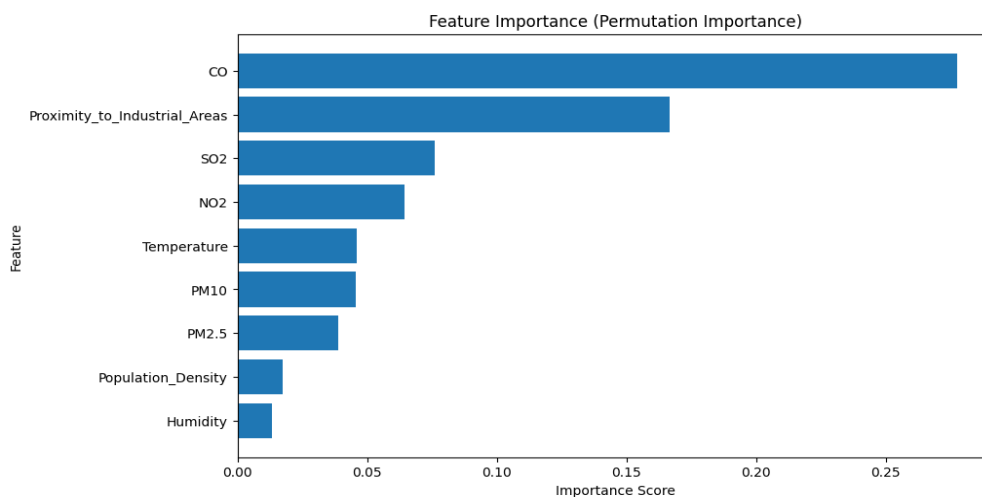
mencerminkan pembagian data yang representatif, namun distribusi yang tidak normal pada PM_{2.5} dan PM₁₀, maka Penulis melakukan transformasi data untuk hasil model yang optimal dengan melakukan *pre-processing* data.



Gambar 4. Visualisasi data hasil *pre-procesing scaling* data

Gambar 4 terlihat bahwa distribusi parameter lingkungan yang telah diskalakan untuk data pelatihan dan pengujian. Temperatur, kelembaban, dan NO2 memiliki distribusi mendekati normal, sementara PM2.5 dan PM10 menunjukkan pola skewed ke kanan

dengan konsentrasi pada nilai rendah. Distribusi data pelatihan dan pengujian terlihat konsisten, menunjukkan pembagian data yang representatif. Selanjutnya, diberikan hasil proses *feature importance* seperti pada Gambar 4 berikut.



Gambar 5. Visualisasi data hasil *feature importance*

Gambar 5 menunjukkan hasil analisis *feature importance* menggunakan metode permutasi.

Berdasarkan *importance score* yang ditampilkan, fitur CO memiliki skor tertinggi, lebih dari 0.25, menunjukkan

bahwa konsentrasi karbon monoksida adalah faktor paling relevan dalam memengaruhi variabel target. Fitur *Proximity to Industrial Areas* juga memiliki skor penting di atas 0.15, mengindikasikan bahwa kedekatan dengan area industri sangat mempengaruhi prediksi, mungkin terkait dengan polusi lingkungan. SO2 dan NO2 memiliki skor menengah (di sekitar 0.1), menandakan relevansi moderat terhadap target, sedangkan fitur-fitur seperti *Temperature*, PM10, dan PM2.5 memiliki kontribusi yang lebih rendah (skor di bawah 0.1), tetapi masih signifikan dalam analisis. Di sisi lain, *Population Density* dan *Humidity* memiliki skor paling kecil, menunjukkan bahwa pengaruhnya terhadap model minimal. Berdasarkan importance score ini, model dapat difokuskan pada fitur dengan skor tinggi untuk meningkatkan efisiensi prediksi tanpa kehilangan akurasi.

3.3 Training Data

Selanjutnya, dilakukan *training* pada dataset yang telah diproses preprosesing. Dalam penelitian ini simulasi pembagian pada dataset menjadi 80 % *training* dan 20% *testing*

```
svm_model = SVC(kernel='rbf', C=1.0,
gamma='scale', random_state=42)
svm_model.fit(X_train_scaled, y_train)
```

3.4 Testing Data dan Evaluasi Model

Setelah didapatkan model SVM yang telah *training*, model tersebut digunakan untuk memprediksi klasifikasi data *testing*. Sehingga didapatkan performa metrik evaluasi seperti Tabel 2 berikut.

Tabel 2. performa matrik evaluasi model

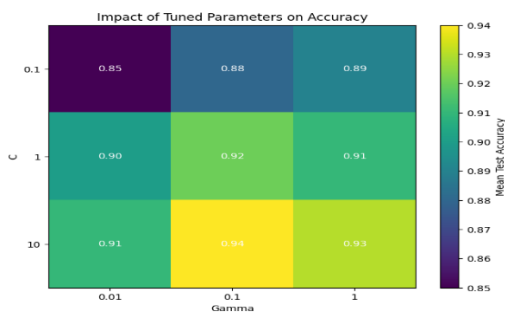
Class	Akurasi	Presisi	Recall	F1-Score
<i>Good</i>	0.99	0.99	0.99	0.99
<i>Hazardous</i>	0.82	0.88	0.82	0.85
<i>Moderate</i>	0.96	0.94	0.96	0.95
<i>Poor</i>	0.86	0.86	0.86	0.86

Berdasarkan Tabel 2 terlihat bahwa model menunjukkan performa yang sangat baik untuk kelas *Good*, dengan nilai akurasi, presisi, *recall*, dan *F1-Score* sebesar 0.99. Hal ini menunjukkan bahwa model dapat dengan sangat andal mengklasifikasikan data ke dalam kelas ini tanpa kesalahan yang berarti.

Untuk kelas *Moderate*, performa model juga tergolong tinggi, dengan akurasi sebesar 0.96, presisi 0.94, *recall* 0.96, dan *F1-Score* 0.95. Hasil ini mencerminkan kemampuan model yang konsisten dalam mengidentifikasi kelas ini dengan hanya sedikit kesalahan klasifikasi. Namun, untuk kelas *Hazardous*, meskipun presisi model relatif tinggi di angka 0.88, akurasi dan

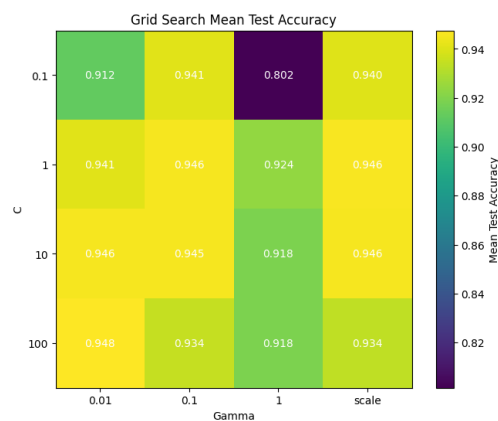
recall hanya mencapai 0.82, yang menghasilkan *F1-Score* sebesar 0.85. Hal ini mengindikasikan bahwa model kurang mampu menangkap semua instance dari kelas ini, meskipun sebagian besar prediksi yang dibuat untuk kelas ini akurat.

Kelas *Poor* menunjukkan performa yang cukup seimbang dengan semua metrik bernilai 0.86. Meskipun hasil ini tidak setinggi kelas *Good* atau *Moderate*, model tetap menunjukkan kinerja yang stabil dalam membedakan data pada kelas ini. Secara keseluruhan performansi algoritma SVM memiliki rata-rata akurasi adalah 0.9075, rata-rata presisi 0.9175, rata-rata *recall* 0.9075 dan rata-rata F-1Score 0.9125. Dalam *impact parameter tuned parameter on Accuracy* pada Gambar 6 menunjukkan hubungan antara dua parameter yang di *tuning*, yaitu nilai *C* dan *Gamma*, terhadap rata-rata akurasi pengujian model.



Gambar 6. Visualisasi *impact parameter tuned parameter on Accuracy*

Berdasarkan *heatmap* ini, terlihat bahwa peningkatan nilai *Gamma* dan *C* secara umum menghasilkan akurasi yang lebih tinggi, dengan puncak akurasi tercapai pada kombinasi nilai *C* = 10 dan *Gamma* = 0.1, yaitu 0.94. Di sisi lain, pada nilai *Gamma* yang rendah (0.01), akurasi cenderung lebih rendah terlepas dari nilai *C*, dengan rata-rata akurasi berkisar antara 0.85 hingga 0.91.



Gambar 7. *Grid Search Mean Test Accuracy*

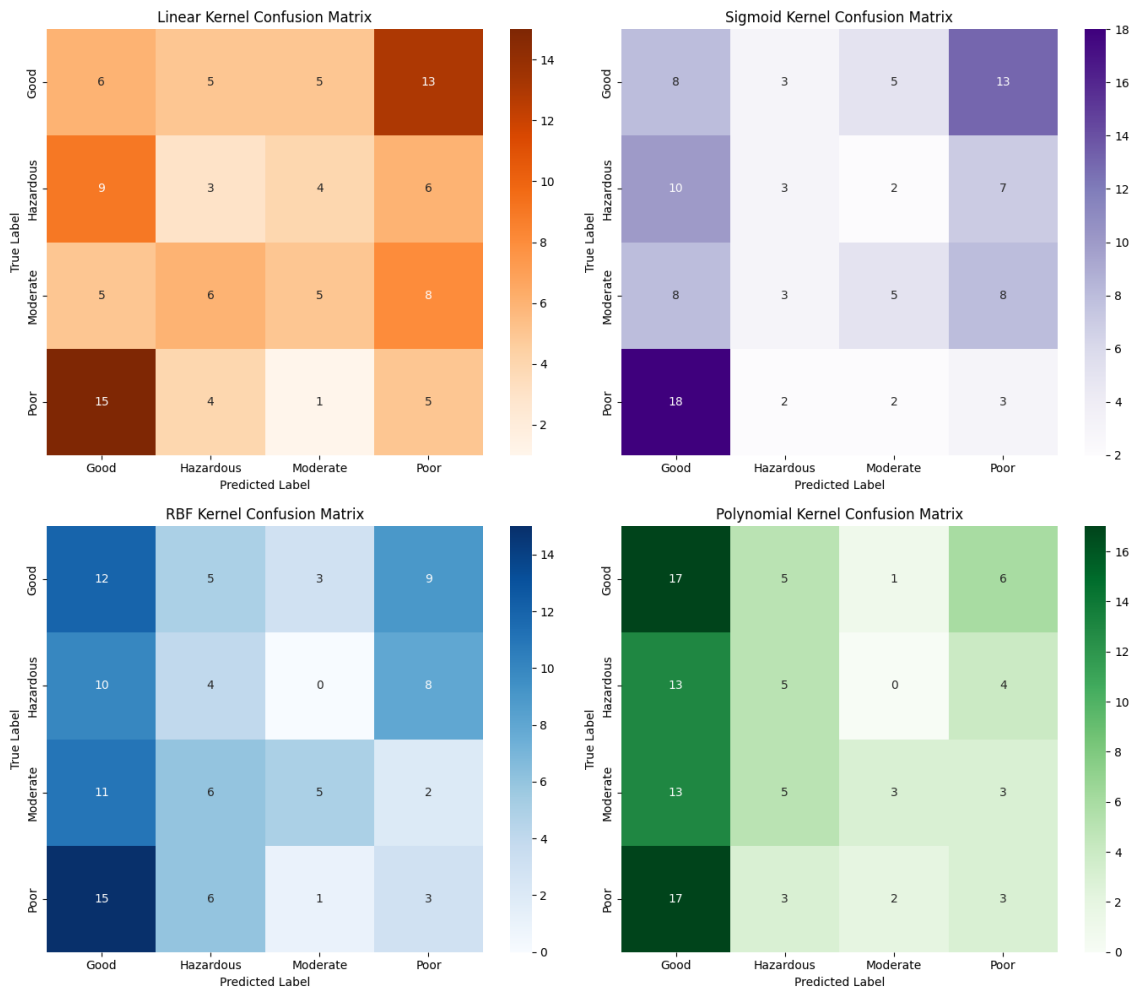
Berdasarkan Gambar 7 *Grid Search Mean Test Accuracy* menunjukkan peningkatan *C* secara konsisten meningkatkan akurasi, terutama pada *Gamma* = 0.1 atau *Gamma* = *scale*, dengan akurasi tertinggi mencapai 0.946. Namun, pada *Gamma* = 1, akurasi menurun drastis, terutama ketika *C* = 0.1 (akurasi 0.802). Analisis ini menekankan bahwa kombinasi $C \geq 10$ dan *Gamma* = 0.1 atau *scale* memberikan performa terbaik, semen-tara pengaturan

Gamma yang tidak optimal dapat menghambat akurasi, menegaskan pentingnya optimasi parameter.

3.5 Evaluasi Model dengan Kernel

Dalam analisis pengujian berbagai *kernel* yang digunakan dalam penelitian ini,

yaitu *RBF Kernel*, *Polynomial Kernel*, *Sigmoid Kernel*, dan *Linear Kernel* dengan pemilihan dan parameter *default*, visualisasi confusion matrik seperti Gambar 8 berikut.



Gambar 8. Confusion Matrix dengan berbagai kernel

Berdasarkan analisis *Confusion Matrix* pada Gambar 8 dengan *RBF Kernel*, *Polynomial Kernel*, *Linear Kernel*, dan *Sigmoid Kernel*, terlihat bahwa setiap *kernel* memiliki kelebihan dan kekurangan dalam klasifikasi data

pada kategori *Good*, *Hazardous*, *Moderate*, dan *Poor*.

Polynomial Kernel menunjukkan performa terbaik dibandingkan *kernel* lainnya, khususnya dalam mengklasifikasikan kategori *Good*,

dengan 17 prediksi benar. *RBF Kernel* juga memberikan hasil yang kompetitif, meskipun memiliki kelemahan pada kategori *Hazardous* dan *Poor*. Sebaliknya, *Linear Kernel* dan *Sigmoid Kernel* menunjukkan performa yang lebih rendah secara keseluruhan, dengan tingkat kesalahan prediksi yang tinggi di semua kategori, terutama pada kategori *Poor*.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa Support Vector Machine (SVM) dengan Polynomial *Kernel* memberikan performa terbaik dalam klasifikasi kualitas udara, khususnya pada kategori *Good*, sementara *RBF Kernel* cukup kompetitif tetapi kurang optimal dalam kategori *Hazardous* dan *Poor*. Optimasi parameter menggunakan GridSearchCV dengan kombinasi $C = 10$ dan $\Gamma = 0.1$ atau scale menghasilkan akurasi rata-rata 90,75%. Sedangkan pada penelitian terdahulu, metode yang diusulkan lebih efisien dalam menangani pola non-linear dibandingkan Random Forest, serta lebih ringan secara komputasi dibandingkan Neural Networks (Vachon et al., 2024; Maaloul & Brahim, 2022). Hasil ini menegaskan pentingnya pemilihan *kernel* dan tuning parameter dalam

meningkatkan akurasi model, serta menunjukkan bahwa konsentrasi CO dan kedekatan dengan area industri memiliki pengaruh signifikan terhadap klasifikasi. Namun, masih terdapat kelemahan dalam mengklasifikasikan kategori *Hazardous* dan *Poor* sehingga mengindikasikan perlunya eksplorasi lebih lanjut terhadap teknik optimasi model dan penggunaan fitur tambahan untuk meningkatkan akurasi secara keseluruhan.

DAFTAR PUSTAKA

- Addis Bikis, "Urban Air Pollution and Greenness in Relation to Public Health," Mizan-Tepi University, Tepi, Ethiopia. January 2023. <https://doi.org/10.1155/2023/8516622>
- Abdulwaheed Tella, Abdul-Lateef Balogun, Naheem Adebisi, and Samsuri Abdullah, "Spatial assessment of PM10 hotspots using Random Forest, K-Nearest Neighbour and Naïve Bayes", Atmospheric Pollution Research. Volume 12, Issue 10, October 2021. <https://doi.org/10.1016/j.apr.2021.101202>
- Biao Sun, Chuanglin Fang, Xia Liao, Xiaomin Guo, and Zhitao Liu, "The relationship between urbanization and air pollution affected by intercity factor mobility: A case of the Yangtze River Delta region" Environmental Impact Assessment Review, Volume

- 100, May 2023.
<https://doi.org/10.1016/j.eiar.2023.107092>
- B.M. Cintoro, S.A. Mulya, R. Sulistyorini and D. herianto, "Analisis emisi polutan kendaraan akibat kegiatan transportasi di lingkungan Universitas Lampung," *Rekayasa: Jurnal Ilmiah Fakultas Teknik Universitas Lampung*, 27(3), Desember 2023, 22-24.
- Burges CJ. A tutorial on *Support Vector Machines* for pattern recognition. *Data Min Knowledge Discovery* 1998;2(2):121–67.
- Burges CJ, Schoˆlkopf B. Improving the accuracy and speed of *Support Vector Machines*. *Neural Inform Proc Syst* 1997;9:375–81
- Chaudhary, M.T. "Urban Air Pollution, Urban Heat Island and Human Health: A Review of the Literature," *Sustainability* 2022, 14, 9234.
<https://doi.org/10.3390/su14159234>
- E.A. Zanaty, "*Support Vector Machines* (SVMs) versus Multilayer Perception (MLP) in data classification", *Egyptian Informatics Journal*, Volume 13, Issue 3, November 2012, Pages 177-183.
<https://doi.org/10.1016/j.eij.2012.08.002>
- IQAIR: Anggota Well Keystone (2023)
<https://www.iqair.com/id/world-most-polluted-cities>.
- Julien Vachon, Jules Kerckhoffs, Stéphane Buteau, and Audrey Smargiassi, "Do machine learning methods improve prediction of ambient air pollutants with high spatial contrast? A systematic review," *Environmental Research*, Volume 262, Part 2, 1 December 2024.
<https://doi.org/10.1016/j.envres.2024.119751>
- Kamel Maaloul, and Lejdel Brahim, "Comparative Analysis of Machine Learning for Predicting Air Quality in Smart Cities", *WSEAS TRANSACTIONS on COMPUTERS*, Volume 21, 2022.
<http://dx.doi.org/10.37394/23205.2022.21.30>
- Pedro Guerra, José Manuel Bravo, Ari Ingimundarson, and Vicenç Puig, "Robust Fault Detection Based on Zonotope-Based Set-Membership Parameter Consistency Test", *A Proceedings Volume from the 6th IFAC Symposium, SAFEPROCESS 2006, Beijing, P.R. China, August 30–September 1, 2006*.
<https://doi.org/10.1016/B978-008044485-7/50178-0>
- Rameswar Debnath, and Haruhisa Takahashi, "*Kernel* selection for the *Support Vector Machine*", *IEICE Transactions on Information and Systems*, Vol. E87-D, No 12, Desember 2004.
- Schoˆlkopf B, Burges C, Smola A. *Advances in kernel methods support vector learning*. MIT Press; 1998.
- Schoˆlkopf BS, Mika CJ, Burges P, Knirsch RR, Muller KG. Input space versus feature space in *kernel*-based methods. *IEEE Trans Neural Network* 1999;10(5):1000–17

Scho“lkopf B, and Smola AJ. “Learning with *kernels*”. Cambridge, MA: The MIT Press; 2002.

Vishal Sharma, et al., “Insights into the recent advances of agro-industrial waste valorization for sustainable biogas production”, *Bioresource Technology*, Volume 390, December 2023. <https://doi.org/10.1016/j.biortech.2023.129829>

Wardatus Sholihah, Ade Silvia Handayani, and Sarjana, "Revolutionizing Healthcare: Comprehensive Evaluation and Optimization of SVM *Kernels* for Precise General Health Diagnosis", *Scientific Journal of Informatics*, Vol. 10, No. 4, Nov 2023. doi: <https://journal.unnes.ac.id/nju/index.php/sji/article/view/46430>