

## Analisis Kualitas Udara Jakarta dan Prediksi Tingkat Polusi dengan Metode Mesin Pembelajaran SVM

### *Analysis of Jakarta's Air Quality and Prediction of Pollution Levels using Support Vector Machine Machine Learning Method*

Soni Agung Wahyudiyanta<sup>1</sup>, Supriyati<sup>2</sup>

<sup>1,2</sup> Program Studi Teknik Industri, Fakultas Teknik, Universitas Pelita Bangsa

sonyagung308@gmail.com , supriyati@pelitabangsa.ac.id

#### **Abstract**

*The advancement of information technology has significantly impacted various aspects of life, including the economy, politics, arts, culture, and education. A good environment, especially air quality, is crucial for human survival. Air pollution in urban areas, such as DKI Jakarta, one of the cities in Asia, has become a serious concern. The increase in the number of motor vehicles in the city contributes to high levels of air pollution. This study utilizes a conceptual model with a design science and behavior science approach to classify air quality in DKI Jakarta. The Standard Air Pollution Index (ISPU) parameters are used, including Carbon Monoxide (CO), Sulfur Dioxide (SO<sub>2</sub>), Nitrogen Dioxide (NO<sub>2</sub>), Surface Ozone (O<sub>3</sub>), and Particulate Matter (PM<sub>10</sub>). The research process involves data analysis using the Support Vector Machine (SVM) algorithm. The results of data exploration show that air quality is predominantly moderate, with a fairly high level of good air quality. Time-based analysis indicates variations in air quality, influenced by events such as lockdowns and the new normal. The dataset from Jakarta Open Data in 2020 is processed and cleaned to ensure good data quality. Distribution analysis reveals that moderate air quality dominates, with specific locations like Bundaran HI having good air quality, while West Jakarta has the highest air pollution levels. The classification process using the SVM algorithm provides good performance evaluation results, with an accuracy of 95.47%, precision of 96.09%, recall of 97.70%, and an f1-score of 96.88%. This model can be relied upon to make predictions and process new data related to air quality in DKI Jakarta. Keywords: supply chain management, inventory, artificial intelligence. This app predict Machine Learning : <https://predict-air-pollution.streamlit.app/>*

**.Keywords:** *Exporatory Data Analyts, Clustering, Air Quality, SVM*

#### **Abstrak**

Kemajuan teknologi informasi memberikan dampak signifikan pada berbagai aspek kehidupan, termasuk ekonomi, politik, seni, budaya, dan pendidikan. Lingkungan yang baik, khususnya kualitas udara, menjadi krusial bagi kelangsungan hidup manusia. Pencemaran udara di perkotaan, seperti DKI Jakarta, salah satu kota di Asia, menjadi perhatian serius. Peningkatan jumlah kendaraan bermotor di kota tersebut berkontribusi pada tingginya tingkat pencemaran udara. Penelitian ini menggunakan model konseptual dengan pendekatan design science dan behavior science untuk mengklasifikasikan kualitas udara di DKI Jakarta. Paramater Indeks Standar Pencemar Udara (ISPU) digunakan, mencakup Karbon Monoksida (CO), Sulfur Dioksida (SO<sub>2</sub>), Nitrogen Dioksida (NO<sub>2</sub>), Ozon Permukaan (O<sub>3</sub>), dan Partikel Debu (PM<sub>10</sub>). Proses penelitian melibatkan analisis data dengan menggunakan algoritma Support Vector Machine (SVM). Hasil eksplorasi data menunjukkan bahwa kualitas udara dominan pada tingkat sedang, dengan tingkat udara baik yang cukup tinggi. Analisis berdasarkan waktu menunjukkan variasi kualitas udara, dipengaruhi oleh peristiwa seperti lockdown dan new normal. Dataset dari Jakarta Open Data tahun 2020 diolah dan dibersihkan untuk membuat kualitas data yang baik. Analisis distribusi menunjukkan bahwa kualitas udara sedang mendominasi, dengan lokasi tertentu seperti Bundaran HI memiliki udara baik, sementara Jakarta Barat memiliki tingkat pencemaran udara tertinggi. Proses klasifikasi menggunakan algoritma SVM memberikan hasil evaluasi kinerja yang baik, dengan akurasi sebesar 95.47%, presisi 96.09%, recall 97.70%, dan f1-score 96.88%. Model ini dapat diandalkan untuk

membuat prediksi dan mengolah data baru terkait kualitas udara di DKI Jakarta. Berikut app prediksi Machine Learning : <https://predict-air-polution.streamlit.app/>

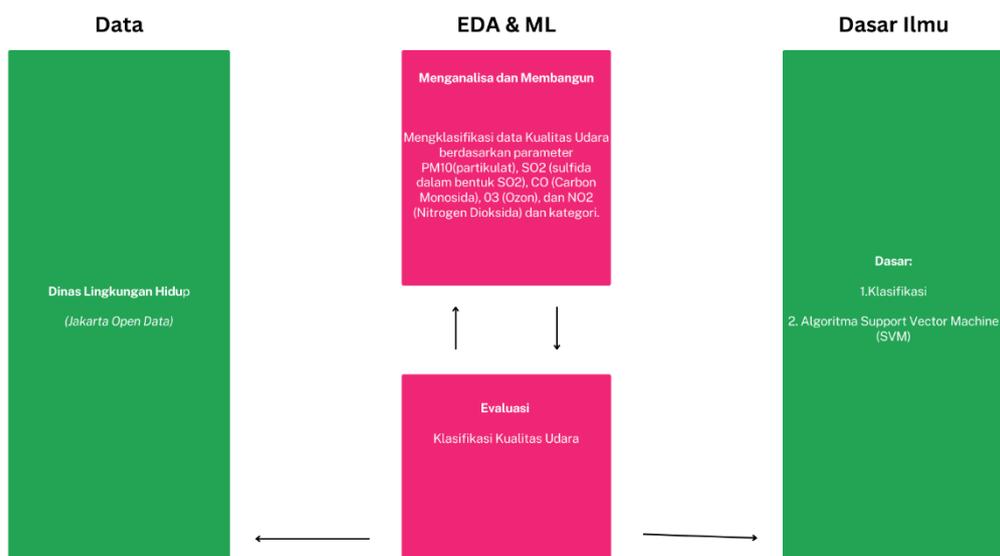
**Kata kunci:** *Exporatory Data Analyts*, Klasifikasi, Kualitas Udara, SVM

## Pendahuluan

Kemajuan teknologi informasi telah berpengaruh dalam segala aspek kehidupan baik di bidang ekonomi, politik, seni dan budaya bahkan didunia pendidikan [1]. Lingkungan yang baik merupakan kebutuhan paling mendasar bagi seluruh umat manusia untuk mempertahankan kehidupan, maka dari itu manusia sangat membutuhkan udara yang baik [2]. Suatu kota yang memiliki pertumbuhan dan perkembangan padat dan pesat menjadi salah satu faktor terjadinya pencemaran udara di suatu perkotaan. Pencemaran udara terjadi karena kualitas udara sudah tergabung dengan berbagai komponen senyawa. Pencemaran udara yang merugikan kesehatan masyarakat merupakan masalah yang tersebar luas di banyak negara di dunia [3]. Kota dengan kualitas udara tidak sehat salah satunya kota yang berada di benua Asia yaitu DKI Jakarta sebagai ibukota Indonesia. Udara (ISPU) digunakan sebagai parameter untuk mengukur kualitas udara. Pada pengukuran ISPU menetapkan lima parameter pencemaran udara yang digunakan untuk pengamatan, yaitu Karbon Monoksida (CO), Sulfur Dioksida (SO<sub>2</sub>), Nitrogen Dioksida (NO<sub>2</sub>), Ozon Permukaan (O<sub>3</sub>), Partikel Debu (PM<sub>10</sub>). Pencemaran udara dapat disebabkan dari perkembangan kendaraan bermotor. Banyaknya masyarakat yang memiliki kendaraan pribadi mengakibatkan pencemaran udara semakin meningkat. Pertumbuhan kendaraan bermotor di DKI Jakarta dari tahun 2018 hingga tahun 2020 mengalami peningkatan yang signifikan sebesar 19% atau sebanyak 8.381.000 kendaraan dari tahun sebelumnya [5].

## Metode Penelitian

Model konseptual digunakan untuk mengenali dan mengevaluasi permasalahan dalam penelitian sistem informasi dengan memanfaatkan pendekatan design science dan behavior science. Pendekatan design science membahas penelitian melalui konstruksi dan evaluasi artefak yang didesain untuk memenuhi kebutuhan bisnis, sedangkan behavior science mencakup penelitian melalui pengembangan dan justifikasi teori yang terkait dengan fenomena yang relevan dengan kebutuhan bisnis [36].



**Gambar 1. Metode Konseptual**

Pada gambar 1 merupakan model konseptual yang berisi tentang data-data yang dibutuhkan dalam proses penelitian yang digunakan untuk menggambarkan konsep permasalahan yang akan diteliti agar mudah dipahami. Berdasarkan Gambar 1 penelitian ini mengenai klasifikasi kualitas udara di DKI Jakarta. Lingkungan yang terdapat pada penelitian ini adalah DKI Jakarta karena penelitian ini menggunakan dataset yang didapat dari website *Jakarta Open Data* untuk melakukan klasifikasi Kualitas Udara di DKI Jakarta. Pada bagian Dasar Ilmu ditampilkan metode data analisa untuk mengklasifikasi data kualitas udara berdasarkan bulan wilayah, parameter PM10, SO2, CO, O3, NO2. Untuk melakukan klasifikasi yang akurat maka algoritma yang digunakan yaitu Algoritma SVM.

## Hasil dan Pembahasan

### Data Preprocessing dan Data Cleansing

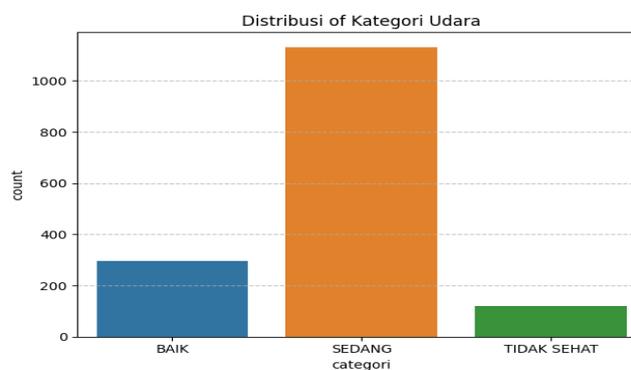
Dataset yang digunakan berasal dari sumber Jakarta Open Data pada tahun 2020. Data tersebut awalnya tersedia dalam format CSV (Comma Separated Values) dan diunduh terpisah untuk setiap bulan dari Januari hingga Desember. Setelah itu, data tersebut telah dikonversi ke format .xlsx atau Excel, menghasilkan 1663 baris data dengan atribut tanggal, wilayah, parameter PM10, CO2, CO, O3, NO2, Maximum, dan Critical. Data kemudian telah disatukan ke dalam satu basis data server yang sama yang kemudian dicari missing valuenya untuk membuat kualitas data yang baik. Contoh dari data yang telah digabungkan dapat ditemukan dalam Tabel 2.

**Table 2. Dataset**

	tanggal	stasiun	pm10	so2	co	o3	no2	max	critical	kategori
0	2020-01-01 00:00:00	DKI1 (Bunderan HI)	30	20	10	32	9	32.0	O3	BAIK
1	2020-01-02 00:00:00	DKI1 (Bunderan HI)	27	22	12	29	8	29.0	O3	BAIK
2	2020-01-03 00:00:00	DKI1 (Bunderan HI)	39	22	14	32	10	39.0	PM10	BAIK
3	2020-01-04 00:00:00	DKI1 (Bunderan HI)	34	22	14	38	10	38.0	O3	BAIK
4	2020-01-05 00:00:00	DKI1 (Bunderan HI)	35	22	12	31	9	35.0	PM10	BAIK

### Exploratory Data Analyst

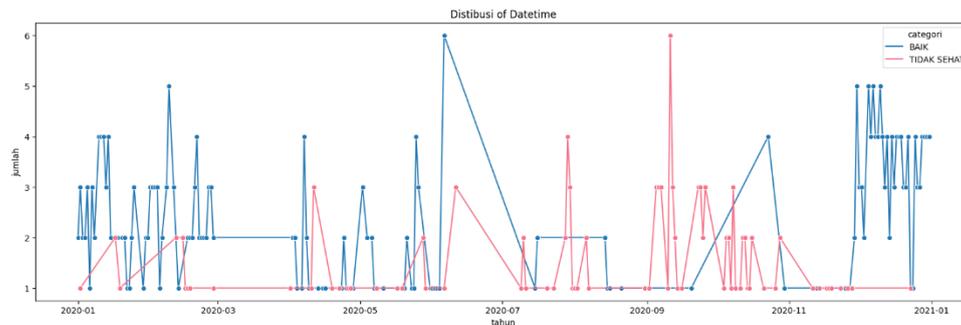
Dari dataset dapat di Analisa distribusi pada tahun 2020 kategori udara manakah yang lebih dominan, apakah Baik, Sedang, Tidak sehat yang diambil dari semua stasiun . Berikut hasil Plot Bar dalam gambar 2.



**Gambar 2. Barplot Distribusi Kategori Udara**

Dari hasil plot diatas dapat dilihat bahwa Kategori udara dengan kualitas sedang merupakan paling banyak terjadi dan kemudian disusul kualitas udara baik, dapat disimpulkan bahwa Tingkat kualitas udara masih di level aman.

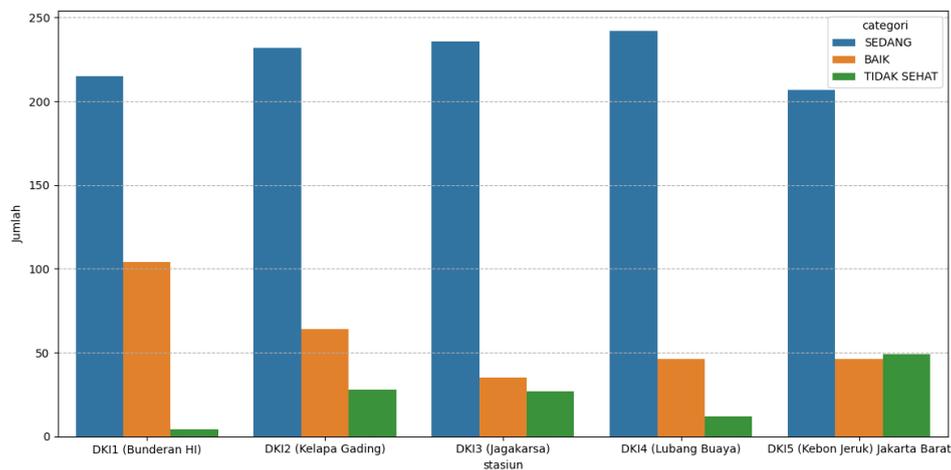
Sekarang kita bandingkan dengan berdasarkan waktunya dari bulan januari 2020 – bulan januari 2023 dimanakah letak kualitas udara yang baik dan tidak sehatnya. Dapat dilihat pada plot Line Pada Gambar 3 dibawah ini.



**Gambar 3. Lineplot Distribusi Berdasarkan Kategori Kualitas Udara**

Dari data plot diatas dapat dilihat pada tahun 2020 januari hingga bulan juli 2020 udara Tidak sehat cenderung rendah dibandingkan kualitas udara baik, mungkin karena terjadinya lockdown yang membuat tidak ada aktivitas diluar rumah ,namun pada bulan September 2020 mengalami lonjakan dikarenakan new normal. Setelah bulan desember hinggann 2021 terdapat penurunan pada udara tidak sehat mungkin karena adanya teknologi terbaru pengurangan emisi pada Perusahaan.

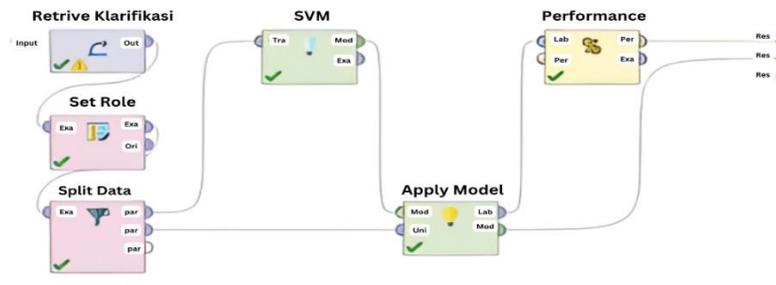
Kemudian kita bandingkan dengan berbagai lokasinya / Stasiun berdasarkan kualitas udaranya. Mari kita lihat pada gambar 4 dibawah.



**Gambar 4. Distribusi Lokasi berdasarkan Kualitas Udara**

Dapat dilihat dari data plot diatas dari semua lokasi memiliki jumlah terbanyak pada kualitas udara sedang, udara dengan kualitas baik ada pada lokasi Bunderan HI , mungkin karena fasilitas lingkungan yang terjaga dan tempat yang bersih yang merupakan kota besar. Pada lokasi Jakarta Barat terdapat kualitas udara tidak sehat tertinggi, dikarenakan padat pemukiman yang kumuh dan fasilitas lingkungan yang tidak terjaga.

## Proses Klasifikasi



Gambar 5. Proses Klasifikasi menggunakan Algoritma SVM

Pada gambar 3. Adalah algoritman menggunakan SVM, Proses pertama yaitu input dataset yang digunakan kemudian dihubungkan dengan set role untuk penandaan kolom sebagai label. Kemudian split data yang bertujuan untuk membagi menjadi 2 data yaitu data uji Selanjutnya dihubungkan ke Apply Model untuk menguji data testing dan data training. Selanjutnya pada operator performance akan melakukan evaluasi kinerja dan akan memberikan nilai dari klasifikasi.

## Evaluasi Performance Algoritma SVM

Dalam menghitung performa terdapat perhitungan persamaan sebagai berikut :

$$\begin{aligned}
 \text{precision} &= \frac{TP}{TP + FP} \\
 \text{recall} &= \frac{TP}{TP + FN} \\
 \text{F1 - score} &= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \\
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}
 \end{aligned}$$

Setelah dilakukan klasifikasi terdapat hasil Menggunakan Algoritma SVM pada gambar 6. Dibawah:

```

=====
Accuracy Score: 95.47%
-----
CLASSIFICATION REPORT:
      1          2          3  accuracy  macro avg
precision  0.947115  0.960957  0.912500  0.954713  0.940191
recall     0.895455  0.976953  0.901235  0.954713  0.924547
f1-score   0.920561  0.968889  0.906832  0.954713  0.932094
support    220.000000  781.000000  81.000000  0.954713  1082.000000
  
```

Gambar 6. Hasil Akurasi evaluasi SVM

Dari hasil Evaluasi diatas pada gambar 6. terdapat hasil Akurasi model 95.47 % , presicison 96.09 % , recall 97.70 % , f1-score 96.88 % [6]. Dapat disimpulkan bahwa performa model tersebut memiliki kualitas yang baik untuk membuat prediksi dan dapat digunakan untuk mengolah data baru.

## Kesimpulan

Kemajuan teknologi informasi telah mengubah lanskap kehidupan sehari-hari, mempengaruhi berbagai aspek dari ekonomi hingga pendidikan. Dalam konteks lingkungan, fokus pada kualitas udara, khususnya di perkotaan seperti DKI Jakarta, menjadi suatu keharusan untuk mendukung kelangsungan hidup manusia. Penelitian ini menggunakan model konseptual dan pendekatan design science dan behavior science untuk mengklasifikasikan kualitas udara di DKI Jakarta. Melibatkan parameter Indeks Standar Pencemar Udara (ISPU), penelitian ini memberikan pemahaman mendalam tentang tingkat pencemaran udara di kota tersebut. Hasil eksplorasi data menunjukkan bahwa kualitas udara cenderung moderat, dengan tingkat udara baik yang relatif tinggi. Analisis berdasarkan waktu mengidentifikasi variasi dalam kualitas udara, yang dipengaruhi oleh peristiwa-peristiwa seperti lockdown dan kebijakan new normal. Dataset dari Jakarta Open Data tahun 2020 diolah dengan seksama, dan hasil analisis distribusi menyoroti dominasi kualitas udara sedang. Lokasi spesifik seperti Bundaran HI menunjukkan tingkat kualitas udara baik, sementara Jakarta Barat menjadi sorotan dengan tingkat pencemaran udara tertinggi. Proses klasifikasi menggunakan algoritma SVM memberikan hasil evaluasi kinerja yang sangat baik, dengan akurasi tinggi, presisi, recall, dan f1-score yang memuaskan. Model ini dapat diandalkan untuk membuat prediksi dan mengolah data terkini terkait kualitas udara di DKI Jakarta. Dengan demikian, penelitian ini memberikan kontribusi penting dalam pemahaman dan pemantauan kualitas udara di perkotaan, yang dapat menjadi dasar untuk pengembangan kebijakan lingkungan yang lebih efektif dan strategi pengurangan pencemaran udara di masa depan.

## Daftar Rujukan

- [1] Y. M. Jamun, "Desain Aplikasi Pembelajaran Peta NTT Berbasis Multimedia," J. Pendidik. dan Kebud.
- [2] K. Prabowo and B. Muslim, *Penyehat Udara, Pertama*. Jakarta Selatan: Pusat Pendidikan Sumber Daya
- [3] H. Zheng, Y. Cheng, and H. Li, "Investigation of model ensemble for fine-grained air quality prediction,"
- [4] IQAir, "World Air Quality Report," 2020 World Air Qual. Rep., no. August, pp. 1–35, 2020, [Online].
- [5] D. P. Sari, "Peningkatan Jumlah Kendaraan Bermotor di DKI Jakarta," 2020. <https://jakarta.bps.go.id/indicator/17/786/1/jumlah-kendaraan-bermotor-menurut-jenis-kendaraan-unitdi-provinsi-dki-jakarta.html>. P. Helo and Y. Hao, "Artificial intelligence in operations management and supply chain management: an exploratory case study," *Production Planning and Control*, vol. 33, no. 16, pp. 1573–1590, 2022, doi: 10.1080/09537287.2021.1882690.
- [6] N. Giarsyani, A. F. Hidayatullah, and R. Rahmadi, "KLASIFIKASI KUALITAS UDARA DENGAN METODE SUPPORT VECTOR MACHINE," *Jire*, vol. 3, no. 1, pp. 48–57, 2020. N. Svetlana, N. Anna, M. Svetlana, G. Tatiana, and M. Olga, "Artificial intelligence as a driver of business process transformation," *Procedia Comput Sci*, vol. 213, pp. 276–284, 2022, doi: 10.1016/j.procs.2022.11.067.
- [7] Badan Pengelolaan Lingkungan Hidup Daerah. *Status Lingkungan Hidup Daerah Provinsi DKI Jakarta*. 2015. BBPLHD, Jakarta.
- [8] Dinas Lingkungan Hidup (DLH) DKI Jakarta. *Indeks Standar Pencemar Udara*. 2019. Jakarta
- [9] Wiryono, Singgih. *Data Pemprov DKI, Kendaraan Bermotor Jadi Masalah Utama Pencemaran Udara Jakarta*, 2021. Kompas.com. Available at: <https://megapolitan.kompas.com/read/2021/11/11/14321471/data-pemprov-dki-kendaraan-bermotor-jadi-masalahutama-pencemaran-udara?page=all>.

- [10] Keputusan Menteri Negara Lingkungan Hidup. Indeks Standar Pencemar Udara. 1997. KEPMENLH, Jakarta.
- [11] PERATURAN PEMERINTAH. Pengendalian Pencemaran Udara PP RI No.41/1999, Jakarta.
- [12] PERMENLHK. Pedoman Nomenklatur Perangkat Daerah Provinsi Dan Kabupaten/Kota Yang Melaksanakan Urusan Pemerintahan Bidang Lingkungan Hidup Dan Urusan Pemerintahan Bidang Kehutanan. 2016. MENLHKRI.