

Analisis Klasifikasi Kualitas Udara Pada Tahun 2022 Menggunakan Metode *Support Vector Machine*

1st M.Habib Jauhari

Fakultas Rekayasa Industri

Universitas Telkom

Bandung, Indonesia

mhabibjauhari@student.telkomuniversi
ty.ac.id

2nd Faqih Hamami

Fakultas Rekayasa Industri

Universitas Telkom

Bandung, Indonesia

faqihhamami@telkomuniversity.ac.i
d

3rd Riska Yanu Fa'rifah

Fakultas Rekayasa Industri

Universitas Telkom

Bandung, Indonesia

riskayanu@telkomuniversity.ac.id

Abstrak - Kondisi udara merujuk pada kondisi atmosfer di sekeliling kita, baik itu murni atau terkontaminasi. Kualitas atmosfer yang optimal tidak hanya sangat penting bagi manusia, melainkan juga memiliki arti yang besar bagi makhluk hidup lainnya seperti flora, air, dan lingkungan tanah. Dengan adanya kebijakan penduduk Jakarta banyak melakukan kegiatan diluar ruangan. Maka, dibutuhkan klasifikasi untuk mendapatkan informasi kualitas udara pada Kota Jakarta. Jadi, salah satu cara melakukannya klasifikasi untuk mengetahui informasi kualitas udara ialah menggunakan data mining. Data mining adalah langkah-langkah untuk mengambil data atau pola yang sebelumnya diketahui, secara tersirat, dan dianggap tidak memiliki nilai sebagai informasi atau pengetahuan berharga dari data dalam jumlah yang besar. Metode data mining klasifikasi digunakan karena dapat mengubah data parameter ISPU menjadi informasi yang menunjukkan kualitas udara harianya. Data kualitas udara dikumpulkan dalam penelitian ini pada Bulan Januari hingga November tahun 2022 dan data tersebut diuji dengan algoritma Support Vector Machine untuk melakukan klasifikasi kualitas udara Kota Jakarta berdasarkan nilai akurasi, presisi, recall, dan skor F1. Hasil didapat dari algoritma SVM dengan rasio terbaik untuk klasifikasi kualitas udara Kota Jakarta dengan perbandingan 80:20 mendapatkan nilai precision 15%, nilai recall 54%, accuracy 90%.

Kata kunci— Data Mining, Klasifikasi, Kualitas Udara, Support Vector Machine

I. PENDAHULUAN

Perkembangan dalam teknologi informasi telah berdampak pada semua aspek kehidupan manusia, termasuk pendidikan, bisnis, politik, dan seni budaya. Kehidupan manusia dan spesies lain dipengaruhi oleh udara, sumber daya bersama yang terbatas. Dalam Keputusan Pemerintah telah ditetapkan definisi polusi udara yang mencakup berbagai faktor manusia maupun alam dapat menyebabkan penurunan kualitas udara, aktivitas gunung berapi adalah salah satu faktor alam yang dapat mencemari udara. Namun, faktor manusia terbesar signifikan adalah transportasi, aktivitas industri, dan pertumbuhan populasi [1].

DKI Jakarta merupakan ibu kota dengan populasi yang padat, dan semakin bertambahnya jumlah masyarakat berarti aktivitas manusia yang meningkat. Pada tahun 2022, tercatat bahwa kendaraan bermotor di wilayah DKI Jakarta sejumlah

26.370.535 unit dan jumlah ini meningkat dari tahun-tahun sebelumnya [2]. Hal ini menyebabkan pada tahun 2022, kandungan PM2,5 di udara wilayah DKI Jakarta sangat tinggi sebesar $148\mu/m^3$ [3].

Dengan mempertimbangkan latar belakang di atas, peneliti ingin melakukan *monitoring* kualitas udara menggunakan indeks standar pencemaran udara. Jadi, diperlukannya teknik *data mining* untuk memprediksi kualitas udara. Studi literatur ini menggunakan klasifikasi dalam teknik *data mining*. Pencarian model, *template* atau fungsi dapat memaparkan objek data, sehingga dapat dikelompokkan ke dalam kategori yang telah disediakan

II. KAJIAN TEORI

Terdapat beberapa landasan atau kajian teori yang berkaitan serta digunakan pada karya ilmiah ini :

A. Pencemaran Udara

Pencemaran udara disebabkan oleh alam secara alamiah serta aktivitas masyarakat. Berbagai bentuk gangguan fisik dimana meliputi sebagai berikut: pencemaran bunyi, suhu tinggi, pancaran atau pencemaran sinar, juga dipandang sebagai polusi udara. Dampak pencemaran udara dapat memiliki dampak baik secara langsung dan terbatas pada suatu wilayah tertentu, dapat juga memiliki pengaruh yang mencakup daerah yang lebih luas atau bahkan mempengaruhi seluruh planet secara global karena sifat alami udara [4].

B. Kualitas Udara dan ISPU

Kondisi udara merujuk pada kondisi atmosfer di sekeliling kita, baik itu murni atau terkontaminasi. Kualitas atmosfer yang optimal tidak hanya sangat penting bagi manusia, melainkan juga memiliki arti yang besar bagi makhluk hidup lainnya seperti flora, air, dan lingkungan tanah. Evaluasi umum atas kualitas atmosfer biasanya dilakukan berdasarkan tingkat nilai variabel polusi atmosfer terpantau, apakah mencapai konsentrasi sangat banyak atau sangat sedikit dari nilai yang ditetapkan sebagai standar kualitas udara nasional. Standar kualitas atmosfer adalah standar rentang atau tingkat unsur pencemaran atmosfer bisa diterima di *atmosfer ambience*. ISPU ditampilkan dalam bentuk nomor tanpa unit dan digunakan sebagai representasi keadaan[6].

kualitas udara di sekitar lokasi terbatas dan berdasarkan dalam hal akibat pada kesehatan masyarakat, keindahan, dan organisme berkehidupan, informasi dapat digunakan sebagai pertimbangan khusus untuk wilayah yang rentan terhadap kebakaran hutan dan lahan sistem untuk memberikan peringatan cepat kepada komunitas di sekitarnya. Tujuan dari ISPU adalah untuk memberikan informasi yang terpadu kepada komunitas mengenai kualitas udara di sekitar pada tempat dan dalam jangka waktu tertentu, bersama dengan hal-hal yang perlu diperhatikan dalam upaya pemerintah pusat untuk mengurangi pencemaran udara

C. Data Mining

Pada saat ini, terdapat banyaknya data yang ada, tetapi kekurangannya adalah bahwa kita mengalami kekurangan informasi dan pengetahuan yang bermanfaat [7]. Agar data dapat digunakan dan diperoleh informasi dari dalamnya, maka dapat menggunakan sebuah proses pengolahan data dalam jumlah besar menjadi bermakna menggunakan suatu metode yang dikenal sebagai *data mining*. Definisi *data mining* adalah langkah-langkah untuk mengambil data atau pola yang sebelumnya tidak diketahui secara tersirat dan dianggap tidak memiliki nilai sebagai informasi atau pengetahuan berharga dari data yang sangat banyak (*Big Data*). Teknik dalam *Data Mining* [7].

D. Machine Learning

Machine Learning adalah pemrograman komputer yang memproses standar kinerja yang ditargetkan dengan memanfaatkan kumpulan data latihan atau (pengalaman sebelumnya) (Chazar, 2020). Pembelajaran mesin merupakan satu di antara Implementasi kecerdasan buatan kecerdasan buatan (AI) yang berorientasi dalam proses tata cara yang mampu melakukan otodidak tanpa memerlukan pemrograman berulang kali [8].

Machine Learning mengandalkan beragam algoritma untuk mengatasi berbagai permasalahan data. Para ahli data berusaha menunjukkan bahwa tidak ada satu jenis algoritma tunggal yang universal cocok untuk semua situasi, melainkan terdapat beragam jenis algoritma yang lebih optimal dalam menyelesaikan berbagai tipe masalah. Jenis algoritma yang diterapkan akan bervariasi sesuai dengan jenis masalah yang dihadapi, jumlah variabel yang terlibat, model yang paling sesuai, dan lain sebagainya. Dibawah ini adalah gambaran singkat mengenai beberapa algoritma umum yang sering digunakan dalam bidang pembelajaran mesin (ML) [9].

E. Algoritma SVM

Algoritma SVM adalah sistem pengajaran yang mengandalkan asumsi dalam bentuk beberapa fungsi *linear* pada fitur-fitur dengan dimensi yang tinggi dan dilatih dengan teknik pembelajaran yang menggunakan prinsip atau teori optimasi sebagai dasar. Pada tahun 1992, SVM dipublikasikan sebagai serangkaian konsep yang menjadi terkemuka dalam bidang pengenalan pola. Tingkat keakuratan hasil dari model SVM sangat terpengaruh oleh jenis fungsi kernel dan nilai parameter yang digunakan. Dengan ciri-ciri ini, metode SVM dapat dipisahkan menjadi dua kategori: SVM *linear* dan SVM yang bukan termasuk dalam kategori. SVM *linear* dimanfaatkan jika data dapat dipisahkan metode *linear* yang artinya dua kelas dapat dipisahkan oleh *hyperplane* dengan batas lembut.

Sedangkan SVM yang bukan termasuk dalam kategori digunakan ketika data tidak dapat dipisahkan secara linier, dan situasi ini digunakan fungsi dari teknik kernel untuk mentransformasi data menjadi representasi di ruang berdimensi tinggi. SVM adalah metode yang cepat dan efektif dalam tantangan mengklasifikasikan teks. Dalam konteks teks, pengelompokan biner dapat diinterpretasikan sebagai *hyperplan* di dalam ruang fitur yang memisahkan titik-titik yang melambangkan contoh positif dari kelompok yang mewakili keadaan negatif. SVM memilih *hyperplane* yang khusus dan dapat memisahkan contoh positif yang diketahui dari contoh negatif selama proses pelatihan. Dalam klasifikasi SVM, keunggulannya terletak pada pendekatan teorinya yang mendukung penyelesaian masalah dengan baik [8].

F. Confusion Matrix

Kinerja sistem penjabaran mendeskripsikan seberapa baik sistem pada mengkategorikan data. Tujuan evaluasi ini adalah untuk menguji model yang digunakan untuk klasifikasi *data mining* untuk menentukan performa sistem [10]. Keterangan dari tabel 1 diatas adalah sebagai berikut :

1. TP : *True Positive* menunjukkan tingkat ketepatan klasifikasi
2. FN : *False Negative* menunjukkan kondisi dimana ramalan tanda hasil tidak tepat dan angka sesungguhnya tepat
3. FP : *False Positive* menunjukkan kondisi dimana tanda ramalan hasil tepat dan angka sesungguhnya tidak tepat
4. TN : *True Negative* menunjukkan tingkat ketepatan klasifikasi

Pada tabel 2 merupakan rentang nilai berdasarkan hasil performa yang didapatkan untuk mengklasifikasi keakuratan pengujian [10]

TABEL 1
CONFUSION MATRIX

	Prediksi Baik	Prediksi Sedang	Prediksi Tidak Sehat
Aktual Baik	TP	FP	FP
Aktual Sedang	TN	TN	TN
Aktual Tidak Sehat	FN	TN	TN

TABEL 2
KLASIFIKASI PERFORMA

Rentang Nilai	Klasifikasi Performa
90% - 100%	Sangat Baik
80% - 100%	Baik
70% - 100%	Cukup
60% - 100%	Buruk
<= 60%	Sangat Buruk

Terdapat beberapa parameter perhitungan yang dapat digunakan pada *confusion matrix* yaitu presisi, *recall*, akurasi dan nilai F1 atau *F1-Score*.

5. resisi, presisi menunjukkan tingkat ketepatan atau ketelitian dalam pengklasifikasian (Arifin & Sasongko, 2018). Berikut adalah rumus untuk menghitung *presisi*:

6. *Recall*, *recall* digunakan untuk menghitung proporsi positif aktual yang benar ditemukan (Arifin & Sasongko, 2018). Berikut adalah rumus untuk menghitung *recall*:

7. Nilai *F1*, dapat didefinisikan sebagai alternatif dari metode akurasi yang diperoleh dari hasil perhitungan antara *presisi* dan *recall*:

8. Akurasi, akurasi bisa didefinisikan menjadi mutu atau taraf kedekatan antara nilai prediksi menggunakan nilai aktual (Arifin & Sasongko, 2018). Berikut adalah rumus untuk menghitung akurasi:

III. METODE

Metode penelitian ini menggunakan metode yang terdiri dari 3 tahapan utama yaitu identifikasi masalah dan solusi, pengolahan data serta tahapan terakhir adalah hasil. Semua tahapan tersebut digambarkan pada gambar 1 dibawah ini.

A. Identifikasi Masalah dan Solusi

Pada tahapan ini mulanya Mengidentifikasi suatu masalah yang sudah dianggap penting bagi penduduk ibukota untuk memiliki kendaraan yang mana manfaat dari kendaraan sangat penting untuk memudahkan mobilitas orang. Meskipun ada banyak transportasi umum yang tersedia di wilayah DKI Jakarta, banyak orang yang percaya bahwa memiliki kendaraan pribadi lebih efisien untuk melakukan perjalanan sehari-hari. Ini diamati dari peningkatan jumlah mobil di jalan raya ada yang di provinsi DKI Jakarta dalam tiga tahun sebelumnya. Setelah melakukan identifikasi masalah, adapun solusi dari masalah yang nantinya digunakan sebagai objek penelitian yang dilakukan dalam penelitian akhir ini dapat dimanfaatkan sebagai data pengujian (*testing*) dan data pelatihan (*training*), hasil penerapan algoritma SVM penggunaan data mining dalam mengklasifikasikan tingkat pencemaran udara di daerah DKI Jakarta serta kategori pengklasifikasian tingkat polusi udara di wilayah DKI Jakarta. Setelah menentukan solusi masalah, menentukan tujuan dari penelitian ini adalah untuk membersihkan data kualitas udara

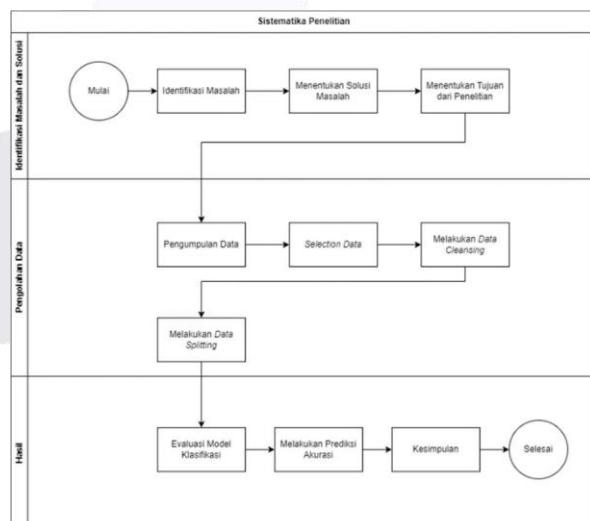
DKI Jakarta agar dapat digunakan dalam pembuatan model klasifikasi, klasifikasi indeks kualitas udara DKI Jakarta pada masa mendatang dengan mengimplementasi algoritma SVM.

B. Pengolahan Data

Setelah tahapan pertama dilakukan, maka tahapan selanjutnya adalah pengolahan data. Informasi yang dikumpulkan adalah data dari Dinas Lingkungan Hidup (DLH). Data yang dikumpulkan mulai dari Januari hingga November 2022. Ukuran data yang didapatkan adalah 114kb (terdiri dari 10 kolom dan 1825 baris). Data disimpan pada format Comma Separated Values (CSV). Setelah melakukan pengumpulan data, peneliti melakukan selection data yang merupakan proses awal dalam data mining. Sebelum langkah penggalian informasi di KDD (*Knowledge Discovery in Database*), kumpulan data operasional harus dilakukan seleksi agar mendapatkan data yang akurat. Proses *data mining* akan menggunakan hasil seleksi data, yang disimpan dalam *file* terpisah. Setelah proses seleksi data dilakukan, langkah berikutnya adalah mengolah data yang telah diperoleh dari proses *cleansing* data. Karena data yang didapatkan masih belum sempurna. Setelah semua data dilakukan proses *cleansing* agar menjadi data yang siap untuk dianalisis. Kemudian, masuk ke tahap proses *data splitting* yaitu proses pembagian antara data latih dan data uji. Dimana, jumlah persentase data latih harus lebih banyak daripada data yang di uji

C. Hasil

Setelah tahapan sebelumnya selesai maka dilanjutkan ke tahapan terakhir yaitu tahapan hasil. Setelah melewati proses *Data Splitting* kemudian masuk ke tahap confusion matrix yaitu mengevaluasi kinerja model klasifikasi untuk mengukur sejauh mana model tersebut berhasil mengklasifikasikan data dengan benar sehingga nantinya ada tiga kelas (sesuai data yang ada) yang ingin diidentifikasi. Setelah dilakukan evaluasi model klasifikasi, nantinya dilakukan prediksi dari hasil evaluasi model klasifikasi. Munculnya itu nanti seperti nilai akurasi, presisi, *recall*, *f1-score*. Pada tahap akhir, menyusun kesimpulan dari hasil penelitian yang telah dilaksanakan tentang hasil klasifikasi indeks kualitas udara di DKI Jakarta dan memberikan saran untuk penelitian ini agar lebih baik lagi untuk kedepannya.



GAMBAR 1 METODE PENELITIAN

IV. HASIL DAN PEMBAHASAN

Pada tahap analisis ruang lingkup, dilakukannya analisis terhadap ruang lingkup penelitian klasifikasi penggunaan algoritma *Support Vector Machine* untuk mengklasifikasikan kualitas udara di DKI Jakarta. Ruang lingkup yang akan dijadikan studi kasus adalah Provinsi DKI

Jakarta. Informasi yang dikumpulkan untuk penelitian ini diambil dari Dinas Lingkungan Hidup (DLH). Terdapat 5 lokasi monitoring kualitas udara di DKI Jakarta yang berarti di DKI 1 Bundaran HI, DKI 2 Kelapa Gading, DKI 3 Jagakarsa, DKI 4 Lubang Buaya, dan DKI 5 Kebun Jeruk. Data yang digunakan mulai dari Januari hingga November tahun 2022. Tahap pengumpulan data adalah pengumpulan informasi yang merupakan langkah-langkah yang dilakukan dari mulai pengambilan hingga penyimpanan yang nantinya dilakukan untuk membuat dan mengatur data di media penyimpanan komputer, sehingga dapat digunakan dengan baik. Pengambilan data dilakukan dengan menggunakan *Open Data* yang didapatkan merupakan data dalam format CSV (*Comma Separated Values*) yang berisikan data kualitas udara dari Januari hingga November tahun 2022 seperti informasi yang terdapat dalam Gambar 2.

ISPU-DKI-2022-01	15/01/2023 18.16	Microsoft Excel W...	32 KB
ISPU-DKI-2022-02	16/01/2023 00.50	Microsoft Excel W...	17 KB
ISPU-DKI-2022-03	21/12/2022 13.27	Microsoft Excel W...	235 KB
ISPU-DKI-2022-04	21/12/2022 13.27	Microsoft Excel W...	25 KB
ISPU-DKI-2022-05	21/12/2022 13.27	Microsoft Excel W...	28 KB
ISPU-DKI-2022-06	21/12/2022 13.27	Microsoft Excel W...	25 KB
ISPU-DKI-2022-07	21/12/2022 13.27	Microsoft Excel W...	28 KB
ISPU-DKI-2022-08	21/12/2022 13.27	Microsoft Excel W...	25 KB
ISPU-DKI-2022-09	21/12/2022 13.27	Microsoft Excel W...	25 KB
ISPU-DKI-2022-10	21/12/2022 13.27	Microsoft Excel W...	25 KB
ISPU-DKI-2022-11	21/12/2022 13.27	Microsoft Excel W...	25 KB

GAMBAR 2
DATA KUALITAS UDARA

Setelah data diperoleh, tahapan selanjutnya adalah melakukan proses *data processing*. Proses ini dilakukan untuk membuat data mentah menjadi data berkualitas yang siap untuk diolah. Dimana nantinya data dapat menjadi informasi yang berguna dan dapat diproses oleh *tools* pengolah data, karena pada data sudah tidak ada lagi yang kosong nilainya dan rusak setelah mengalami tahap pra-pemrosesan data. Tahapan pertama yang dilakukan adalah *data integration* dengan tujuan agar data tersebut diintegrasikan menjadi satu *server* basis data yang seragam. Data yang digabungkan menjadi satu kesatuan, yaitu data dari Januari hingga November 2022 menjadi satu *database*. Sebelum tahap penggalian informasi, data dari sekumpulan data harus dipilih operasional dan digunakan dalam operasi penggalian data. Selanjutnya, himpunan data yang diinginkan dapat dihasilkan melalui seleksi data. Pilihan lain adalah mengutamakan kelompok kecil variabel atau sampel data yang akan digunakan untuk penemuan. Semua *dataset* memiliki total 10 kolom. Tujuan dari seleksi data pada tahap ini adalah mengambil kolom sebelumnya berjumlah 10 kolom menjadi 6 kolom dikarenakan data yang diolah cukup menggunakan atribut dari 6 kolom tersebut.

Selanjutnya dilakukan proses *data cleansing* merupakan tahap pembersihan, menghapus serta memperbaiki data mentah yang didapatkan. Data bersih juga merupakan proses analisis kualitas data dengan mengubah, mengkoreksi, atau menghapus data yang tidak akurat, salah, atau tidak lengkap. Dimulai dari memeriksa apakah ada data *null* (*missing*

value). Setelah ditemukannya data *null*, dilanjutkan dengan mengubah *value* yang tidak sesuai untuk data yang kita butuhkan. Setelah itu, dilakukan pengubahan tipe data agar dapat dijalankan dengan sebagaimana fungsinya. Pada Gambar 3 adalah contoh data yang telah dilakukan proses *cleansing*.

No.	tanggal	stasiun	Pm10	So2	co	O3	No2	max	critical	kategori
0	01/12/2022	DKI 1 (Bunderan HI)	43	48	12	16	5	54	Pm25	Sedang
1	02/12/2022	DKI 1 (Bunderan HI)	22	45	7	17	5	45	So2	Baik
2	03/12/2022	DKI 1 (Bunderan HI)	34	49	11	17	8	54	Pm25	Sedang
3	04/12/2022	DKI 1 (Bunderan HI)	51	51	13	32	12	70	Pm25	Sedang
4	05/12/2022	DKI 1 (Bunderan HI)	73	53	16	37	16	106	Pm25	Sedang

GAMBAR 3
DATA SETELAH DILAKUKAN PROSES CLEANSING

Setelah itu, dilakukan proses *data splitting* untuk memisahkan *dataset* menggunakan jumlah rasio data tertinggi yang digunakan buat pelatihan. Sebagai contoh, data mampu dipisah sebagai rasio 80:20 atau 70:30 masing-masing untuk *training* (pelatihan) dan *testing* (pengujian).

Selanjutnya, Pelatihan model SVM tujuannya untuk menemukan *hyperplane* (batas keputusan) yang dapat memisahkan kelas-kelas dengan baik. Lalu dilanjutkan ke tahap evaluasi, tujuan dari tahap evaluasi ini adalah untuk mengevaluasi kinerja model SVM dan mengevaluasi metrik yang umumnya digunakan termasuk akurasi, presisi, *recall*, *F1-Score*. Setelah didapatkan hasil dari evaluasi, peneliti dapat melakukan suatu pengambilan kesimpulan dari hasil evaluasi klasifikasi kualitas udara di kota DKI Jakarta.

Setelah diperoleh hasil pelatihan, dilakukan *import GridSearchCV* untuk pencarian lintas validasi berulang dengan berbagai nilai parameter yang telah ditentukan sebelumnya. Setelah data berhasil dikumpulkan serta disesuaikan dengan kebutuhan. Langkah selanjutnya adalah melakukan proses implementasi dan evaluasi. Dalam proses ini, penulis menggunakan tiga skenario dalam proses memperoleh hasil dan mendapatkan kesimpulan. Setiap skenario diberi nama skenario 1, skenario 2 dan skenario 3 dimana pada skenario 1 proses yang dilakukan adalah proses *data splitting*, skenario 2 melakukan perhitungan dengan menggunakan *confusion matrix* dan pada skenario 3 proses yang dilakukan adalah proses *GridSearchCV*. Pada gambar 4 adalah hiperparameter dan parameter terbaik yang dapat digunakan.

```
print(grid.best_params_)
print(grid.best_estimator_.get_params())
['C': 10, 'gamma': 0.1, 'kernel': 'rbf']
['C': 10, 'break_ties': False, 'cache_size': 200, 'class_weight': None, 'coef0': 0.0, 'decision_function_shape': 'ovr', 'degree': 3, 'gamma': 0.1}
# Set the best parameters
best_params = grid.best_estimator_.get_params()
print(best_params)
['C': 10, 'gamma': 0.1, 'kernel': 'rbf']
```

GAMBAR 4 HIPERPARAMETER DAN PARAMETER TERBAIK

Selanjutnya dilakukan proses pengujian menggunakan *confusion matrix* dengan skenario rasio 80:20 dan 70:30. Pada tabel 3 adalah hasil perhitungan *confusion matrix* dengan rasio 80:20, dengan menghasilkan akurasi sebesar 76.8%. Selain itu, diperoleh juga nilai presisi sebesar 6.52, *recall* 4.87 dan *F1-Score* sebesar 34,71%.

TABEL 1
CONFUSION MATRIX RASIO 80:20

Akurasi 76,80%			
	Prediksi Baik	Prediksi Sedang	Prediksi Tidak Sehat
Aktual Baik	5	18	0
Aktual Sedang	4	250	0
Aktual Tidak Sehat	4	51	0

Selanjutnya adalah perhitungan *confusion matrix* dengan rasio 70:30. Pada tabel 3 adalah hasil perhitungan *confusion matrix*, dengan menghasilkan akurasi sebesar 76,84%. Selain itu, diperoleh juga nilai presisi sebesar 81,7, *recall* 4,18 dan *F1-Score* sebesar 26%.

V. KESIMPULAN

Berdasarkan analisis data dan pembahasan yang telah dilakukan, dapat disimpulkan penelitian adalah algoritma SVM dalam membantu pemantauan kualitas udara dengan cara melakukan *splitting* data, melakukan model SVM, dan melakukan evaluasi pada *confusion matrix*. Tingkat akurasi pada algoritma SVM dalam melakukan prediksi pada data kualitas udara dari kota Jakarta terdapat pada rasio 80:20 dengan akurasi 92,05%, presisi 21,73%, dan recall 55,55%. Penelitian ini menghasilkan model terbaik yang didapatkan dari klasifikasi algoritma SVM untuk data kualitas udara tahun 2022 pada Provinsi DKI Jakarta. Penelitian selanjutnya disarankan untuk penggunaan model SVM yang mencari *hyperplane* terbaik agar dapat memisahkan dua buah kelas melihat kembali data yang didapatkan apakah data yang didapatkan memiliki dua buah kelas atau lebih dari dua kelas. Kemudian, untuk penelitian selanjutnya disarankan agar membuat model dengan metode lain seperti *Naive Bayes*, *Random Forest*, atau yang lainnya agar dapat membandingkan performa terbaik dari model yang didapatkan.

REFERENSI

- [1] . P. I. Agista, N. Gusbini, and M. D. D. Maharani, "Analisis Kualitas Udara Dengan Indeks Standar Pencemar Udara (Ispu) Dan Sebaran Kadar Polutannya Di Provinsi DKI

Jakarta," Sustainable Environmental and Optimizing Industry Journal, Sep. 2020, doi: 10.36441/seoi.v2i2.491.

[2] Badan Pusat Statistik Provinsi DKI Jakarta."https://jakarta.bps.go.id/indicator/17/786/1/jumlah-kendaraan-bermotor-menurut-jenis-kendaraan-unit-di-provinsi-dki-jakarta.html

[3] A. M. Puspitasari, "Klasifikasi penyakit gigi dan mulut menggunakan metode Support Vector machine," 2017. [Online]. Available: <http://repository.ub.ac.id/id/eprint/1998>

[4] R. Teguh, E. D. Oktaviyani, and K. A. Mempun, ".Rancang Bangun Desain Internet Of Things Untuk Pemantauan Kualitas Udara Pada Studi Kasus Polusi Udara," Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika, Aug. 2018, doi: 10.47111/jti.v12i2.532.

[5] N. Nawasyarif, M. Julkarnain, and K. R. Ananda, "Sistem Informasi Pengolahan Data Ternak Unit Pelaksana Teknis Produksi Dan Kesehatan Hewan Berbasis Web," Jurnal Informatika Teknologi Dan Sains (Jinteks), vol. 2, no. 1, pp. 32–39, Feb. 2020, doi: 10.51401/jinteks.v2i1.556.

[6] Kusumaningtyas, M. K. (2019, November). Analisis Dampak Diterapkannya Kebijakan Working From Home Saat Pandemi Covid-19 Terhadap Kondisi Kualitas Udara Di Jakarta. Jurnal Meteorologi Klimatologi dan Geofisika, 6, 6-14.Sudipa (2023), Data Mining

[7] Isnain, S. A. (2021). Sentimen analisis publik terhadap kebijakan lockdown pemerintah jakarta menggunakan algoritma SVM, 2, 31-37.

[8] Mahesh, B. (2019, January). Machine Learning Algorithms -A Review. International Journal of Science and Research (IJSR), 9, 381-386. doi:10.21275/ART20203995

[9] Arifin, O., & Sasongko, T. B. (2018). Analisa Perbandingan Tingkat Performansi Metode Support Vector Machine Dan Naive Bayes Classifier Untuk Klasifikasi Jalur Minat Sma. Semnasteknomedia Online, 6(1). <https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/download/2059/1868>