

ANALISIS KLASIFIKASI DATASET INDEKS STANDAR PENCEMARAN UDARA (ISPU) DI MASA PANDEMI MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Ihda Innar Ridho¹ Galih Mahalisa²

Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin

Email: ihdainnaridho@gmail.com

Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin

Email: galih.mahalisa@gmail.com

ABSTRAK

Kualitas udara atau air quality merupakan kadar kandungan udara berdasarkan konsentrasi polutan di lokasi tertentu. Pada masa Pandemi terdapat adanya perubahan kualitas udara yang terjadi di Indonesia. Penelitian ini akan melakukan analisis klasifikasi tingkat kualitas udara berdasarkan dataset ISPU. Data mining dikelompokkan dalam dua kategori, yakni supervised dan unsupervised. Algoritma Support Vector Machine (SVM) adalah suatu metode yang menggunakan algoritma supervised, SVM digunakan untuk mencari hyperplane terbaik dengan memaksimalkan jarak antar kelas. Hyperplane adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas. Metode yang digunakan dalam penelitian ini yaitu algoritma Support Vector Machine (SVM), dimana merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Didapatkan hasil pengujian tertinggi dengan menggunakan support vector dengan nilai accuracy sebesar dengan akurasi 97%. Hal ini membuktikan bahwa pada masa pandemi covid-19 kualitas udara menjadi lebih sehat atau dalam kondisi sedang.

Kata Kunci : *Kualitas Udara, Klasifikasi, Support Vector Machine (SVM), Akurasi*

1. PENDAHULUAN

Kualitas udara atau air quality merupakan kadar kandungan udara berdasarkan konsentrasi polutan di lokasi tertentu [1]. Pada masa Pandemi terdapat adanya perubahan kualitas udara yang terjadi di Indonesia [2]. Selama masa pandemi Covid-19 yang mengakibatkan banyak daerah di Indonesia melakukan Pembatasan Sosial Berskala Besar (PSBB) serta memaksa para pegawai bekerja dari rumah atau work from home (WFH), sedikit banyak memengaruhi kualitas udara di beberapa kota di Indonesia khususnya kota-kota besar [3]. Hal ini dipengaruhi oleh berkurangnya aktivitas kendaraan bermotor di jalan raya, yang merupakan sumber pencemar/emisi udara paling besar. ISPU

merupakan angka tanpa satuan, digunakan untuk menggambarkan kondisi mutu udara ambien di lokasi tertentu dan didasarkan kepada dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup lainnya [4]. Pengklasifikasian dalam penelitian ini digunakan sebagai alat bantu bahkan dijadikan suatu bahan pertimbangan dalam menghasilkan outcome yang akurat.

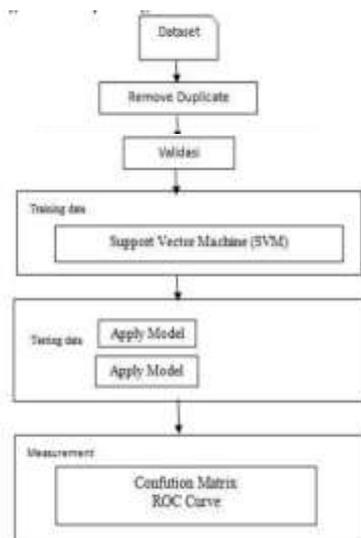
Referensi yang berkaitan dengan penelitian ini seperti berikut Analisis Kualitas Udara Menggunakan Algoritma K-Means mendapatkan hasil pengujian monitoring sensor gas menghasilkan nilai accuracy, sensitivity, specificity, precision dan recall sebesar 77,78 % [5]. Referensi tentang ANALISIS DAN KOMPARASI ALGORITMA KLASIFIKASI DALAM INDEKS PENCEMARAN UDARA DI DKI

JAKARTA, bahwa Neural Network Backpropagation, K-Nearest Neighbors, Support Vector Machine, dan Naive Bayes juga masih dapat digunakan sebagai model klasifikasi yang baik karena mendapatkan nilai akurasi yang tinggi di atas 90% dan nilai kappa di atas 0.8 dan nilai RMSE di bawah 0.3 [4].

Dalam penelitian ini menggunakan metode SVM dengan beberapa fitur yang ada agar prediksi yang didapatkan bisa lebih akurat dan spesifik

2. METODE PENELITIAN

Model yang diusulkan pada penelitian ini adalah menggunakan algoritma support vector machine dan algoritma support vector machine, yang terlihat pada gambar 1 dibawah ini.



Gambar 1 . Tahap Metode Penelitian

Karakteristik dataset terdiri dari 7 variabel. 6 fitur yang dijadikan variabel X atau disebut variabel dependen dan 1 variabel yang dijadikan variabel Y atau variabel independen, dan data sebanyak 181, namun tidak semua data dapat digunakan dan tidak semua atribut digunakan karena harus melalui beberapa tahap pengolahan awal data (preparation data) [1]. Beberapa teknik

digunakan untuk mendapatkan data yang berkualitas tinggi. Yaitu, menghapus data duplikat, data nilai setara, atau menghapus data yang berlebihan. Validasi data untuk mengidentifikasi dan menghilangkan data ganjil (outlier/noise), data tidak konsisten, dan data tidak lengkap (missing value) [9]. Integrasi dan transformasi data untuk meningkatkan akurasi dan efisiensi algoritme. Data yang digunakan dalam buku putih ini memiliki nilai kategoris . Data diubah dengan perangkat lunak Jupyter

3. HASIL DAN PEMBAHASAN

a. Tahap Pengolahan Data

Pengumpulan dataset dengan cara pengunduhan pada API Command kaggle (kaggle datasets download -d adityakadiwal/water-potability).

Proses awal yang dilakukan adalah pengolahan dataset seperti pada gambar

	pm10	pm25	so2	co	o3	no2	kategori
0	73	126	38	26	46	34	TIDAK SEHAT
1	53	70	40	14	55	25	SEDANG
2	32	53	40	11	42	19	SEDANG
3	36	59	40	14	47	24	SEDANG
4	29	51	40	14	45	35	SEDANG
...
176	82	140	56	13	41	35	TIDAK SEHAT
177	82	145	53	18	40	45	TIDAK SEHAT
178	78	140	52	18	53	39	TIDAK SEHAT
179	90	154	54	15	81	35	TIDAK SEHAT
180	63	100	50	18	65	29	SEDANG

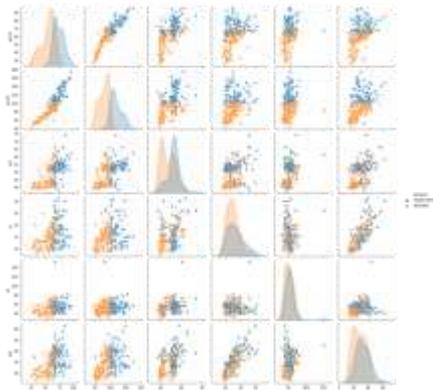
181 rows x 7 columns

Gambar 2. Dataset

Pengolahan data dimulai dengan mengidentifikasi kumpulan data yang berisi tipe data integer dan data floating-point. Setelah menganalisis data, langkah selanjutnya adalah preprocessing dan splitting dataset. Artinya, sebaran data latih dan uji seperti terlihat pada gambar di bawah ini :

id	berbunyi	sehat	sharpening	buffer	connectivity	organ_carbon	tikasmenahan	stability	penalty
0	True	False	False	False	False	False	False	False	False
1	False	False	False	True	False	False	False	False	False
2	False	False	False	False	True	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	True	False	False	False	False
...
3271	False	False	False	False	False	False	False	False	False
3272	False	False	False	False	True	False	False	True	False
3273	False	False	False	False	True	False	False	False	False
3274	False	False	False	False	True	False	False	False	False
3275	False	False	False	False	True	False	False	False	False

Gambar 3. Pengolahan Data



Gambar 4 pairplot kategori data

Pada gambar 4 menjelaskan visualisasi berbentuk pairplot untuk mengkategorikan data, kuning melambangkan kategori udara yang sedang, dan biru melambangkan kategori udara yang tidak sehat. Dari gambar tersebut terlihat bahwa udara sedang lebih dominan dibandingkan dengan udara yang tidak sehat.

b. Tahap Pengujian

	precision	recall	f1-score	support
SEDANG	1.00	0.95	0.98	21
TIDAK SEHAT	0.94	1.00	0.97	16
accuracy			0.97	37
macro avg	0.97	0.98	0.97	37
weighted avg	0.97	0.97	0.97	37

Gambar 5. Hasil akurasi

Pada gambar 6 terlihat menunjukkan nilai akurasi yang baik yaitu sekitar 97%, sehingga dapat dikatakan bahwa

berdasarkan pengujian dataset kualitas udara lebih banyak yang sedang dibandingkan dengan kualitas udara yang tidak sehat pada masa pandemic covid

4. KESIMPULAN

Berdasarkan pengolahan data yang sudah dilakukan dengan membandingkan algoritma Support Vector Machine dengan menggunakan dataset dari ISPU. Kemudian di uji untuk mendapatkan nilai accuracy dari setiap algoritma sehingga didapatkan hasil pengujian tertinggi dengan menggunakan support vector dengan nilai accuracy sebesar dengan akurasi 97%. Hal ini membuktikan bahwa pada masa pandemi covid-19 kualitas udara menjadi lebih sehat atau dalam kondisi sedang

5. REFERENSI

[1] H. Situs et al., “Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV,” *Bul. Ilm. Sarj. Tek. Elektro*, vol. 4, no. 1, pp. 12–21, 2022, doi: 10.12928/biste.v4i1.6079.

[2] M. S. Mukharrim, S. Nengsi, and R. Rahmatilla, “Alat Cuci Tangan Tanpa Sentuh sebagai Produk Pengabdian kepada Masyarakat Desa Rea,” *SIPISSANGNGI J. Pengabd. Kpd. Masy.*, vol. 1, no. 1, p. 34, Feb. 2021, doi: 10.35329/sipissangngi.v1i1.1817.

[3] T. S. Sabrila, V. R. Sari, and A. E. Minarno, “Analisis Sentimen Pada Tweet Tentang Penanganan Covid-19 Menggunakan Word Embedding Pada Algoritma Support Vector Machine Dan K-Nearest Neighbor,” *Fountain Informatics J.*, vol. 6, no. 2, p. 69, Jul. 2021, doi: 10.21111/fij.v6i2.5536.

[4] S. Syihabuddin Azmil Umri, “Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 98–104, 2021, doi: 10.33387/jiko.v4i2.2871.

[5] K. Auliasari, M. Kertaningtyas, and J. Raya Karanglo Km, “ANALISIS

- KUALITAS UDARA
MENGUNAKAN ALGORITMA K-
MEANS,” 2021.
- [6] Y. Nurdiansyah, “Informal : informatics journal,” vol. 2, no. 2, pp. 114–122, Jul. 2017.
- [7] P. A. Octaviani, Y. Wilandari, and D. Ispriyanti, “PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) PADA DATA AKREDITASI SEKOLAH DASAR (SD) DI KABUPATEN MAGELANG,” *ejournal3.undip.ac.id*, vol. 3, no. 4, pp. 811–820, 2014.
- [8] V. K. Chauhan, K. Dahiya, · Anuj Sharma, and K. A. In, “Problem formulations and solvers in linear SVM: a review,” Springer, vol. 52, no. 2, pp. 803–855, Aug. 2019, doi: 10.1007/s10462-018-9614-6.
- [9] L. A. Wardani, I. G. P. S. Wijaya, and F. Bimantoro, “KLASIFIKASI JENIS DAN TINGKAT KEMATANGAN BUAH PEPAYA BERDASARKAN FITUR WARNA, TEKSTUR DAN BENTUK MENGGUNAKAN SUPPORT VECTOR MACHINE,” *J. Teknol. Informasi, Komputer, dan Apl. (JTIKA)*, vol. 4, no. 1, pp. 75–87, Mar. 2022, doi: 10.29303/JTIKA.V4I1.171.