

Disetujui oleh:

Prof. Dr. Imas Sukaesih Sitanggang, S.Si., M.Kom. Dr. Hendra Rahmawan, S.Kom., M.T.



**SEMINAR HASIL
PROGRAM STUDI SARJANA ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER
INSTITUT PERTANIAN BOGOR**

Nama	: Lutfiah Nursabiliyanti
NIM	: G6401211041
Program Studi	: Sarjana Ilmu Komputer
Judul Penelitian	: Model <i>Spatio-Temporal Clustering</i> untuk Estimasi PM2.5 di Jakarta
Pembimbing	: 1. Prof. Dr. Imas Sukaesih Sitanggang, S.Si., M.Kom. 2. Dr. Hendra Rahmawan, S.Kom., M.T.
Hari / Tanggal	: Rabu / 4 Juni 2025
Waktu	: 13.00 s/d 14.00 WIB
Tempat	: RKU 2.01
Pembahas	: 1. Halida Fiadnin: G6401211142 2. Dwinanda Rizkiansyah: G6401211117 3. Syifa Ainul Qolbiyah: G6401211007

MODEL *SPATIO-TEMPORAL CLUSTERING* UNTUK ESTIMASI PM2.5 DI JAKARTA

Lutfiah Nursabiliyanti¹, Imas Sukaesih Sitanggang², Hendra Rahmawan³

¹Mahasiswa Sarjana Program Studi Ilmu Komputer, IPB

²Pembimbing 1, staf pengajar Sekolah Sains Data Matematika dan Informatika IPB

³Pembimbing 2, staf pengajar Sekolah Sains Data Matematika dan Informatika IPB

Abstrak

LUTFIAH NURSABILIYANTI. Model *Spatio-Temporal Clustering* untuk Estimasi PM2.5 di Jakarta. Dibimbing oleh IMAS SUKAESIH SITANGGANG dan HENDRA RAHMAWAN.

Polusi udara memiliki dampak buruk terhadap kesehatan manusia, terutama PM2.5 yang dapat menembus hingga alveoli paru-paru karena ukurannya yang sangat kecil. Saat ini pemantauan kualitas udara di Jakarta melalui Stasiun Pemantau Kualitas Udara Ambien (SPKUA) dan sensor kualitas udara yang dikelola oleh Dinas Lingkungan Hidup (DLH) cakupannya belum menyeluruh, sehingga dibutuhkan pendekatan baru untuk memantau PM2.5 dengan cakupan lebih luas dan analisis sebaran yang lebih akurat. Oleh karena itu, penelitian ini bertujuan untuk membuat model *spatio-temporal clustering* menggunakan algoritma ST-DBSCAN untuk mengestimasi PM2.5 di Jakarta dengan data Himawari AOD pada periode 1 Januari 2022 hingga 31 Desember 2024. Tahapan penelitian ini fokus pada pemodelan estimasi PM2.5 menggunakan metode Random Forest Regressor (RFR) dan *clustering* menggunakan ST-DBSCAN. Hasil penelitian menunjukkan model estimasi PM2.5 berhasil dibangun dan model terbaik didapatkan menggunakan parameter $n_estimator = 1.000$, $max_depth = 80$, $min_sample_split = 2$, dan $min_sample_leaf = 1$ dengan $R^2 = 0,682$ dan MAE = 7,429. Kemudian hasil *clustering* terbaik didapatkan dengan parameter $Eps1 = 0,15$, $Eps2 = 3$, dan $MinPts = 20$, menghasilkan koefisien Silhouette sebesar 0,747 pada sub *dataset* DJF 2023. *Cluster* yang terbentuk yaitu 3 *cluster* dan 45 titik *noise* dengan rata-rata PM2.5 tiap *cluster* yaitu 70,61 $\mu\text{g}/\text{m}^3$, 62,24 $\mu\text{g}/\text{m}^3$, dan 52,07 $\mu\text{g}/\text{m}^3$. Hasil penelitian ini diharapkan dapat membantu pemerintah dan pihak lain yang terkait dalam merancang kebijakan dan strategi pengendalian polusi udara di Jakarta.

Kata Kunci: DKI Jakarta, Himawari AOD, PM2.5, *Spatio-Temporal Clustering*, ST-DBSCAN.

PENDAHULUAN

Latar Belakang

Polusi udara terdiri dari berbagai macam gas yang berbahaya bagi kesehatan manusia seperti Ozon (O_3), Karbon Monoksida (CO), Nitrogen Oksida (NO_3), Sulfur Dioksida (SO_3) dan *Particulate Matter* (PM) (Glencross *et al.* 2020; WHO 2021). PM2.5 sangat berbahaya bagi kesehatan manusia karena ukurannya yang kurang atau sama dengan 2.5 mikrometer sehingga memungkinkan untuk masuk hingga ke dalam alveoli pada paru-paru (Panuju dan Usman 2023). Hingga saat ini, berdasarkan portal Kualitas Udara Jakarta (<https://udara.jakarta.go.id/>), Provinsi DKI Jakarta memiliki 31 Stasiun Pemantau Kualitas Udara Ambien (SPKUA) dan sensor kualitas udara untuk memantau kualitas udara secara kontinu oleh Dinas Lingkungan Hidup (DLH) DKI Jakarta. SPKU dan sensor tersebut belum mencakup keseluruhan area DKI Jakarta.

Teknik penginderaan jauh telah banyak dimanfaatkan oleh peneliti untuk mengukur polusi udara dengan cakupan spasial yang lebih luas melalui pemanfaatan data satelit seperti data *Aerosol Optical Depth* (AOD). Li *et al.* pada tahun 2022 menggunakan metode Random Forest Regressor (RFR) untuk estimasi konsentrasi PM2.5 pada tahun 2020 hingga 2022 menggunakan data AOD dari satelit *Multi-Angle Implementation of Atmospheric Correction* (MAIAC).

Spatio-temporal clustering merupakan teknik *clustering* yang mengelompokkan data berdasarkan kemiripan spasial dan temporalnya (Rodriguez *et al.* 2019; Ansari *et al.* 2020). Salah satu algoritma *spatio-temporal clustering* adalah Spatio-Temporal Density-Based Spatial Clustering of Applications with Noise (ST-DBSCAN). Penelitian oleh Manalu *et al.* (2021) menggunakan ST-DBSCAN untuk mengelompokkan titik gempa di Pulau Sulawesi pada tahun 2019. Namun, belum ada penelitian yang secara khusus menggunakan algoritma ST-DBSCAN untuk melakukan *spatio-temporal clustering* PM2.5. Oleh karena itu, penelitian ini dilakukan untuk menerapkan metode ST-DBSCAN untuk melihat pola sebaran PM2.5 di Jakarta.

Perumusan Masalah

Provinsi DKI Jakarta hanya memiliki 31 SPKUA dan sensor kualitas udara yang digunakan untuk pemantauan PM2.5 di Jakarta. SPKUA tersebut belum mencakup seluruh area di Jakarta sehingga dibutuhkan pemetaan data PM2.5 menggunakan citra satelit untuk cakupan wilayah yang lebih menyeluruh dan analisis sebaran PM2.5 yang lebih akurat menggunakan pendekatan *spatio-temporal clustering*.

Tujuan

Tujuan penelitian ini adalah:

1. Membuat model estimasi PM2.5 di Jakarta menggunakan metode RFR berbasis data Meteorologis dan Himawari AOD.
2. Menerapkan metode *spatio-temporal clustering* menggunakan algoritma ST-DBSCAN untuk melihat pola sebaran PM2.5 di Jakarta.

Manfaat

Pola penyebaran PM2.5 di Jakarta yang dihasilkan dari penelitian ini diharapkan dapat membantu pemerintah dan pihak lain yang terkait dalam merancang kebijakan dan strategi pengendalian polusi udara di Jakarta.

Ruang Lingkup

Ruang lingkup penelitian ini adalah:

1. Wilayah penelitian ini hanya mencakup wilayah Jakarta.
2. Data PM2.5 dan meteorologis yang digunakan hanya dari delapan titik SKPUA di Jakarta.

METODE

Data Penelitian

Data yang digunakan pada penelitian ini adalah data Indeks Standar Pencemar Udara (ISPU) PM2.5 dari delapan SPKUA yang tercatat di situs Rendah Emisi Jakarta dan data meteorologi di lokasi yang sama dengan delapan SPKUA tersebut. SPKUA tersebut diantaranya berlokasi di Bundaran HI, Jagakarsa, Jakarta GBK, Kebun Jeruk, Kelapa Gading, Lubang Buaya, US Embassy 1, dan US Embassy 2. Selain itu, penelitian ini juga menggunakan data AOD Level 2 dari satelit Himawari-8 dan Himawari-9 dengan resolusi spasial 0.05 derajat dengan periode waktu harian. Data yang digunakan yaitu data dalam rentang waktu 1 Januari 2022 hingga 31 Desember 2024.

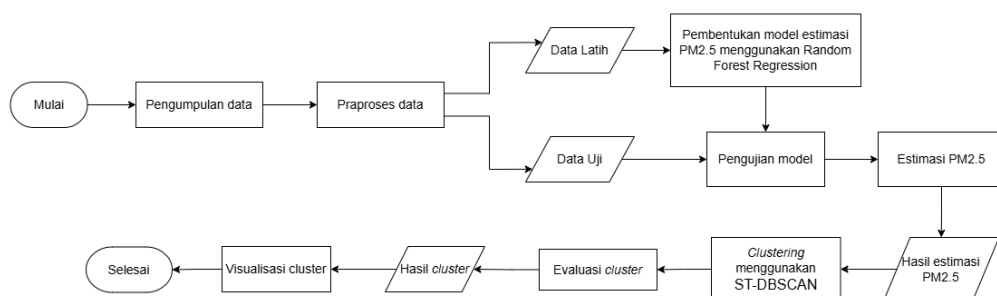
Peralatan Penelitian

Penelitian ini dilakukan menggunakan perangkat keras dan perangkat lunak sebagai berikut:

- a. Perangkat keras berupa komputer dengan spesifikasi:
 - AMD Ryzen 3 3250U, RAM 8 GB dan SSD 512 GB
- b. Perangkat lunak yang digunakan antara lain:
 - Sistem Operasi Windows 11 64-bit
 - Bahasa Pemrograman Python versi 3.12

Tahapan Penelitian

Tahapan yang dilakukan pada penelitian ini fokus pada pemodelan estimasi PM2.5 menggunakan metode RFR dan *clustering* menggunakan ST-DBSCAN.



Gambar 1 Tahapan penelitian

Pengumpulan Data

Data historis PM2.5 di Jakarta dalam periode 1 Januari 2022 hingga 31 Desember 2024 tersedia dan dapat diakses dari situs Rendah Emisi Jakarta milik DLH Jakarta (<https://rendahemisi.jakarta.go.id/ispu>). Sedangkan data meteorologi tersedia di situs Visual Crossing (<https://www.visualcrossing.com/>). Data AOD tersedia di situs JAXA Himawari Monitor (<https://www.eorc.jaxa.jp/ptree/>).

Praproses Data

Praproses data dilakukan untuk memastikan data yang digunakan untuk analisis adalah data yang berkualitas. Tahapan ini terdiri dari *feature selection*, *data selection*, *data cleaning*, dan *data transformation*.

Pemodelan RFR

Data yang dibangun dari praproses data akan dibagi menjadi data uji dan data latih secara acak menggunakan perbandingan 80% data latih dan 20% data uji. Model RFR dibentuk menggunakan fungsi ‘RandomForestRegressor’ dari modul `sklearn.ensemble`, salah satu modul dari Scikit-learn. Selama pemodelan akan dilakukan *hyperparameter tuning* untuk mendapatkan parameter optimal menggunakan ‘GridSearchCV’ dari modul `sklearn.model_selection` untuk mendapatkan parameter yang menghasilkan model dengan R2 terbaik. Model RFR menggunakan data AOD dan meteorologi sebagai prediktor dan PM2.5 sebagai targetnya.

Pengujian Model RFR

Model yang dibangun dari data latih akan divalidasi terhadap data uji menggunakan Mean Absolute Error (MAE) dan uji korelasi menggunakan R2. Nilai MAE dihitung menggunakan persamaan 1 (Pelánek 2015) dan R2 dihitung menggunakan persamaan 2 (Gao 2024).

$$MAE = \frac{\sum_{i=1}^n |y_i - y_p|}{n} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_p)^2}{\sum_{i=1}^n (y_i - y_a)^2} \quad (2)$$

dimana y_i merupakan nilai aktual, y_p merupakan nilai estimasi, y_a merupakan rata-rata nilai aktual, dan n merupakan jumlah data yang diestimasi.

Estimasi PM2.5

Model RFR dengan R2 optimal akan digunakan untuk mengestimasi PM2.5 di wilayah yang nilai PM2.5-nya tidak diketahui.

Clustering ST-DBSCAN

Data PM2.5 dari SPKUA dan hasil estimasi model RFR akan digunakan untuk *spatio-temporal clustering* menggunakan algoritma ST-DBSCAN yang dikembangkan oleh Birant dan Kut (2007). Algoritma ST-DBSCAN memerlukan empat parameter *input* yaitu Eps1, Eps2, MinPts, dan Δt . Eps1 merupakan parameter jarak untuk atribut spasial (lintang dan bujur). Eps2 merupakan

parameter jarak untuk atribut *non* spasial. MinPts merupakan jumlah minimum anggota titik dalam Eps1 dan Eps2 untuk membentuk sebuah *cluster*. Parameter $\Delta\epsilon$ digunakan untuk mencegah penemuan pengelompokan gabungan akibat perbedaan kecil nilai *non* spasial dari lokasi tetangga. *Clustering* dilakukan menggunakan fungsi ‘ST_DBSCAN’ dari *package* *st_dbscan*. Fungsi ini hanya menggunakan tiga parameter input yaitu Eps1, Eps2, dan MinPts. Oleh karena itu, dibutuhkan evaluasi *cluster* menggunakan koefisien Silhouette. Selama pemodelan akan dilakukan *hyperparameter tuning* untuk mendapatkan parameter optimal untuk *clustering*.

Evaluasi Cluster

Hasil *clustering* akan dievaluasi menggunakan koefisien Silhouette untuk menghitung seberapa baik *cluster* yang terbentuk. Koefisien Silhouette dihitung menggunakan persamaan 3 (Rousseeuw 1987).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

dimana $a(i)$ merupakan jarak rata-rata objek i ke semua objek lain di dalam *clusternya* dan $b(i)$ merupakan nilai minimum rata-rata jarak objek i ke semua *cluster* lain. Nilai koefisien Silhouette bervariasi dalam rentang -1 hingga 1. Semakin $s(i)$ mendekati 1, artinya objek i sudah berada pada *cluster* yang tepat. Sebaliknya, jika $s(i)$ mendekati -1, artinya objek i sebaiknya dimasukkan ke dalam *cluster* lain. Pada penelitian ini, implementasi evaluasi menggunakan koefisien Silhouette dilakukan menggunakan fungsi ‘silhouette_score’ dari modul *sklearn.metrics*.

Visualisasi Cluster

Visualisasi *cluster* dilakukan untuk melihat sebaran titik PM2.5 setiap *cluster*, di mana hasil *clustering* ditampilkan melalui perbedaan warna antar *cluster* menggunakan *plotly.subplots*.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Dataset historis PM2.5 diperoleh dengan interval waktu per jam dalam periode harian untuk delapan SPKUA di Jakarta. *Dataset* ini terdiri dari 16 atribut, yaitu waktu, ISPU PM10, ISPU PM2.5, ISPU SO2, ISPU CO, ISPU O3, ISPU NO2, *Status by* PM10, *Status by* PM2.5, *Status by* SO2, *Status by* CO, *Status by* O3, *Status by* NO2, *Critical Parameter*, *Overall ISPU*, dan *Datetime*. *Dataset* meteorologis diperoleh dengan interval waktu harian untuk delapan lokasi yang sama dengan kedelapan SPKUA di Jakarta. *Dataset* meteorologis terdiri dari 18 atribut, yaitu *Name (Longitude, Latitude)*, *Datetime*, *Tempmax*, *Tempmin*, *Temp*, *Feelslikemax*, *Feelslikemin*, *Feelslike*, *Dew*, *Humidity*, *Precip*, *Precipcover*, *Windgust*, *Windspeed*, *Winddir*, *Sealevelpressure*, *Cloudcover*, dan *Visibility*. Sedangkan data AOD diperoleh dalam bentuk *file* NC yang kemudian dikonversi menjadi *file* CSV dengan interval waktu harian. Data AOD terdiri dari empat atribut, yaitu *Datetime*, *Longitude*, *Latitude*, dan AOD. Seluruh *dataset* memiliki rentang waktu 1 Januari 2022 hingga 31 Desember 2024.

Praproses Data

Pada *dataset* PM2.5, dilakukan *feature selection* dengan hanya memilih atribut ISPU PM2.5 dan *Datetime*. Selanjutnya ditambahkan atribut lokasi dari tiap stasiun berupa koordinat *Longitude* dan *Latitude*. Data PM2.5 yang memiliki interval waktu per jam kemudian ditransformasikan menjadi data harian dengan menghitung rata-rata ISPU PM2.5 setiap harinya. Selanjutnya interpolasi linear dilakukan untuk menghilangkan *missing value* pada data ISPU PM2.5.

Pada *dataset* meteorologi, atribut *Name* ditransformasikan menjadi dua atribut baru, yaitu *Longitude* dan *Latitude*. Interpolasi linear kemudian dilakukan pada seluruh atribut kecuali atribut *Longitude*, *Latitude*, dan *Datetime*. Pada *dataset* AOD, dilakukan pembersihan *missing value* dari atribut AOD dengan menghapus baris yang memiliki nilai AOD sama dengan 0. Kemudian atribut *datetime* ditransformasikan menjadi atribut baru bernama *day_of_year* yang berisi nilai hari ke berapa dalam satu tahun dengan rentang 0 sampai 365.

1. Praproses data untuk pemodelan RFR

Dataset PM2.5 digabungkan dengan data meteorologi berdasarkan atribut *Datetime*, *Longitude*, dan *Latitude* yang sama. Selanjutnya *dataset* tersebut digabungkan kembali dengan data AOD. Penggabungan dilakukan dengan cara mencocokkan titik *Longitude* dan *Latitude* dari SPKUA yang berada dalam batas kotak (*grid cell*) AOD yang memiliki resolusi spasial 0,05 derajat. Untuk menangani *outlier*, proses *smoothing* dilakukan dengan cara mengganti nilai PM2.5 yang berada di luar rentang kuartil 1 hingga kuartil 3 dengan nilai batas bawah dan batas atasnya. *Dataset* ini memiliki total 6.506 baris.

2. Praproses data untuk *clustering* ST-DBSCAN

Dataset AOD yang lokasinya tidak beririsan dengan SPKUA akan digabungkan dengan data meteorologis dari SPKUA terdekat. *Dataset* ini akan digunakan untuk mengestimasi PM2.5 di lokasi-lokasi Jakarta yang nilai PM2.5-nya tidak diketahui menggunakan model RFR yang telah dibuat. Setelah proses estimasi PM2.5, *dataset* hasil estimasi ini akan digabungkan dengan *dataset* PM2.5 yang sudah ada untuk mendapatkan *dataset* lengkap PM2.5 di wilayah Jakarta. Atribut *Datetime* kemudian ditransformasikan menjadi data numerik dan menjadi atribut baru bernama *Day_num* untuk mempermudah proses *clustering*. Selanjutnya *dataset* akan dibagi menjadi 12 sub *dataset* yang mewakili tiap musim di tahun 2022 hingga 2024. Pembagian musim dibagi menjadi empat jenis, yaitu musim hujan (Desember - Februari, disebut DJF), peralihan musim hujan ke kemarau (Maret – Mei, disebut MAM), kemarau (Juni – Agustus, disebut JJA), dan peralihan musim kemarau ke musim hujan (September – November, disebut SON). Jumlah baris pada tiap sub *dataset* dapat dilihat pada Tabel 1.

Tabel 1 Jumlah baris sub *dataset* untuk *clustering*

Tahun	Bulan/Musim	Jumlah Baris
2022	DJF	2.751
	MAM	9.042
	JJA	10.580
	SON	8.016
2023	DJF	4.433
	MAM	10.347
	JJA	12.288
	SON	11.925

2024	DJF	7.515
	MAM	9.704
	JJA	11.805
	SON	11.061

Pemodelan RFR

Dataset untuk pemodelan RFR terdiri dari 5.204 data latih dan 1.302 data uji. Selama pemodelan, dilakukan *hyperparameter tuning* dengan menguji parameter *n_estimator*, *max_depth*, *min_sample_split*, dan *min_sample_leaf*. Nilai *n_estimator* yang diujikan yaitu 100, 300, 500, dan 1.000. Nilai *max_depth* yang diujikan yaitu 80, 85, 90, 95, dan 100. Nilai *min_sample_split* yang diujikan yaitu 2 dan 5. Nilai *min_sample_leaf* yang diujikan yaitu 1 dan 2. Total kombinasi parameter yaitu 80 kombinasi.

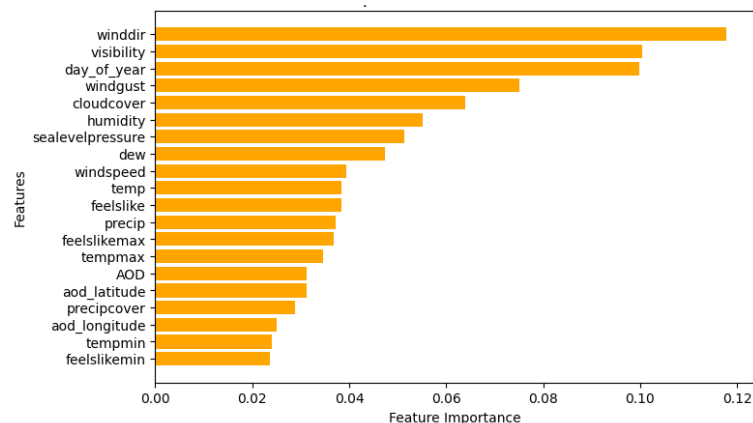
Evaluasi Model RFR

Setelah pembuatan model estimasi PM2.5 menggunakan 80 kombinasi parameter, parameter yang menghasilkan R2 terbaik tersaji pada Tabel 2.

Tabel 2 Hasil evaluasi model RFR

dataset	Parameter				R2	MAE
	<i>n_estimator</i>	<i>max_depth</i>	<i>min_sample_split</i>	<i>min_sample_leaf</i>		
<i>Smoothing</i>	1.000	80	2	1	0,682	7,429
Tanpa <i>smoothing</i>	500	85	2	1	0,660	7,669

Feature importance pada model RFR terbaik dilakukan untuk mengukur seberapa besar kontribusi setiap fitur terhadap prediksi model. Pada Gambar 2 terlihat bahwa nilai AOD memiliki kontribusi yang kurang signifikan pada model estimasi PM2.5.



Gambar 2 *Feature importance* model RFR terbaik

Estimasi PM2.5

Model RFR dengan teknik *smoothing* memiliki R2 optimal. Oleh karena itu, model ini digunakan untuk mengestimasi PM2.5 di wilayah yang nilai PM2.5-nya

tidak diketahui. Hasil estimasi PM2.5 kemudian digunakan pada praproses data untuk *clustering* menggunakan ST-DBSCAN.

Clustering ST-DBSCAN

Dataset PM2.5 dari SPKUA dan PM2.5 estimasi yang telah dibagi menjadi 12 sub *dataset* akan digunakan untuk *clustering* menggunakan ST-DBSCAN. Selama *clustering* dilakukan *Hyperparameter tuning* pada parameter Eps1, Eps2, dan MinPts untuk mendapat parameter yang menghasilkan koefisien Silhouette terbaik. Nilai Eps1 yang dicobakan yaitu 0,7; 0,8; 0,9; 0,1; dan 0,15. Nilai Eps2 yang diujikan yaitu 2, 3 dan 7. Nilai MinPts yang diujikan yaitu 5, 10, dan 20. Total kombinasi parameter yang diuji yaitu 45 kombinasi. Satuan untuk Eps1 adalah derajat ($^{\circ}$), yang dapat dikonversi ke kilometer (km) dengan mengalikan nilai Eps1 dengan 111. Sedangkan satuan waktu pada Eps2 adalah hari. Atribut yang digunakan sebagai fitur untuk *clustering* yaitu *Day_num*, *Longitude*, *Latitude*, dan PM2.5.

Evaluasi Cluster

Setelah dilakukan *clustering* untuk setiap sub *dataset*, dilakukan evaluasi *cluster* menggunakan koefisien Silhouette. Hasil evaluasi berkisar antara -1 hingga 1. Semakin koefisien Silhouette mendekati 1 artinya *cluster* yang terbentuk semakin baik, dan semakin mendekati -1 artinya *cluster* yang terbentuk semakin buruk. Parameter terbaik yang menghasilkan nilai koefisien Silhouette tertinggi untuk setiap sub *dataset* disajikan pada Tabel 3.

Tabel 3 Kombinasi parameter ST_DBSCAN terbaik setiap sub *dataset* PM2.5

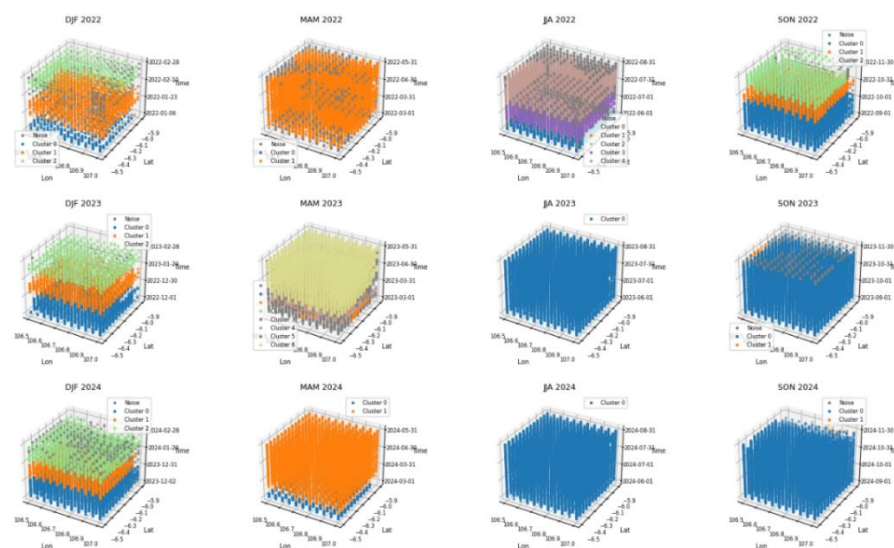
<i>Dataset</i>	Eps1	Eps2	MinPts	Jumlah <i>Cluster</i>	Jumlah <i>Noise</i>	Koefisien Silhouette
DJF 2022	0,1	2	20	3	180	0,557
MAM 2022	0,07	3	20	2	433	0,122
JJA 2022	0,07	2	20	5	796	0,362
SON 2022	0,08	2	20	3	132	0,525
DJF 2023	0,15	3	20	3	45	0,747
MAM 2023	0,07	2	20	7	902	0,150
JJA 2023	-	-	-	1	0	-
SON 2023	0,07	3	20	2	295	0,257
DJF 2024	0,08	2	20	3	261	0,587
MAM 2024	0,1	2	5	2	0	0,421
JJA 2024	-	-	-	1	0	-
SON 2024	0,08	2	5	2	3	0,188

Pada Tabel 3 dapat dilihat bahwa *dataset* DJF 2023 mendapatkan koefisien Silhouette tertinggi yaitu 0,747 dengan jumlah *cluster* 3 dan *noise* 45. Masing-masing *cluster* sub *dataset* DJF 2023 terdiri 2073, 1444, dan 871 titik data. *Cluster* 0 memiliki rata-rata PM2.5 70,61 $\mu\text{g}/\text{m}^3$ dengan periode waktu 1 Desember 2022 hingga 27 Desember 2022. *Cluster* 1 memiliki rata-rata PM2.5 62,24 $\mu\text{g}/\text{m}^3$ dengan periode waktu 31 Desember 2022 hingga 26 Januari 2023. Sedangkan *cluster* 3 memiliki rata-rata PM2.5 52,07 $\mu\text{g}/\text{m}^3$ dengan periode waktu 5 Februari 2023 hingga 28 Februari 2022. Sedangkan *dataset* JJA 2023 dan JJA 2024 hanya membentuk satu *cluster* dengan rata-rata PM2.5 86,68 $\mu\text{g}/\text{m}^3$ dan 86,66 $\mu\text{g}/\text{m}^3$ sehingga tidak memiliki nilai koefisien Silhouette. Proses *clustering* menggunakan ST-DBSCAN yang hanya membentuk satu *cluster* disebabkan oleh rapatnya data

PM2.5 baik secara spasial maupun temporalnya. Algoritma ST-DBSCAN menggunakan parameter Eps1 sebagai jarak spasial dan Eps2 sebagai jarak *non* spasial maksimum agar dua titik dikatakan berada pada kelompok yang sama. Saat data terlalu rapat atau jarak spasial dan temporalnya terlalu kecil, maka seluruh titik data akan dikelompokkan menjadi satu *cluster*.

Visualisasi Cluster

Visualisasi hasil *clustering* dilakukan dengan memberikan warna yang berbeda untuk setiap *cluster* dan warna abu khusus untuk *cluster* berlabel -1 atau *noise*. Rata-rata PM2.5 tiap *cluster* dari seluruh sub *dataset* disajikan pada Tabel 4.



Gambar 3 Visualisasi *cluster*

Tabel 4 Rata-rata PM2.5 tiap *cluster* seluruh sub *dataset*

<i>Dataset</i>	<i>Cluster</i>	Periode	Rata-rata PM2.5 ($\mu\text{g}/\text{m}^3$)
DJF 2022	0	6 Januari 2022 – 12 Januari 2022	65,94
	1	17 Januari 2022 – 11 Februari 2022	66,14
	2	14 Februari 2022 – 28 Februari 2022	61,47
MAM 2022	0	2 Maret 2022 – 8 Maret 2022	48,67
	1	2 Maret 2022 – 31 Mei 2022	74,83
	2	1 Juni 2022 – 14 Juni 2022	78,68
JJA 2022	0	1 Juni 2022 – 5 Juni 2022	74,83
	1	1 Juni 2022 – 5 Juni 2022	69,21
	2	16 Juni 2022 – 12 Juli 2022	87,02
	3	17 Juli 2022 – 31 Agustus 2022	83,78
SON 2022	0	1 September 2022 – 20 Oktober 2022	77,82
	1	20 Oktober 2022 – 11 November 2022	68,49
	2	9 November 2022 – 30 November 2022	61,71
DJF 2023	0	1 Desember 2022 – 27 Desember 2022	70,61
	1	31 Desember 2022 – 26 Januari 2023	62,24
	2	5 Februari 2023 – 28 Februari 2023	52,07
MAM 2023	0	3 Maret 2023 – 7 Maret 2023	61,09
	1	3 Maret 2023 – 14 Maret 2023	66,43
	2	3 Maret 2023 – 7 Maret 2023	55,91
	3	3 Maret 2023 – 7 Maret 2023	55,5

	4	15 Maret 2023 – 29 Maret 2023	75,81
	5	29 Maret 2023 – 2 April 2023	65
	6	29 Maret 2023 – 31 Mei 2023	78,36
JJA 2023	0	1 Juni 2023 – 31 Agustus 2023	86,68
SON 2023	0	1 September 2023 – 30 November 2023	84,79
	1	26 November 2023 – 30 November 2023	75,2
DJF 2024	0	2 Desember 2023 – 4 Januari 2024	80,19
	1	4 Januari 2024 – 29 Januari 2024	61,44
	2	29 Januari 2024 – 28 Februari 2024	66,24
MAM 2024	0	1 Maret 2024 – 8 Maret 2024	59,13
	1	11 Maret 2024 – 31 Mei 2024	80,03
JJA 2024	0	1 Juni 2024 – 31 Agustus 2024	86,66
SON 2024	0	1 September 2024 – 30 November 2024	78,7
	1	26 November 2024 – 26 November 2024	67,76

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil membuat Model estimasi PM2.5 di Jakarta menggunakan data AOD Himawari dan meteorologis. Model terbaik didapatkan dengan nilai R^2 sebesar 0,682 dan MAE sebesar 7,429, yang menunjukkan kemampuan model menjelaskan variasi data PM2.5 dengan cukup baik. Namun, tingkat akurasi model masih perlu ditingkatkan karena nilai MAE yang tergolong tinggi. Teknik *smoothing* pada nilai PM2.5 terbukti dapat meningkatkan performa model. Selain itu, nilai AOD Himawari dinilai memiliki kontribusi yang kurang signifikan pada model estimasi PM2.5. *Clustering* menggunakan ST-DBSCAN berhasil diimplementasikan pada data PM2.5 dari SPKUA dan PM2.5 estimasi periode 2022 hingga 2024. Hasil *clustering* terbaik menghasilkan koefisien Silhouette sebesar 0,747 pada sub *dataset* DJF 2023. Selain itu, terdapat dua sub *dataset* yang hanya membentuk satu *cluster*, yaitu JJA 2023 dan JJA 2024, yang disebabkan oleh jarak spasial dan temporal yang terlalu kecil.

Saran

Penelitian ini hanya menggunakan data PM2.5 dari delapan SPKUA di Jakarta. Penelitian selanjutnya disarankan dapat menambah jumlah data PM2.5 dari seluruh SPKUA dan sensor kualitas udara di Jakarta untuk meningkatkan performa estimasi PM2.5. Mengacu pada hasil *feature importance*, fitur AOD dari satelit Himawari memiliki kontribusi yang kurang signifikan pada model estimasi PM2.5. Oleh karena itu, penelitian selanjutnya dapat mengeksplorasi data AOD dari satelit lain seperti MODIS atau MAIAC yang mungkin dapat memberikan kontribusi lebih signifikan terhadap model estimasi PM2.5. Selain itu, karena data AOD Himawari cenderung seragam secara spasial, eksplorasi algoritma *clustering* berbasis data *non* spasial dapat dilakukan untuk melihat pola sebaran PM2.5 yang lebih baik.

DAFTAR PUSTAKA

- Ansari MY, Ahmad A, Khan SS, Bhushan G, Mainuddin. 2020. Spatiotemporal clustering: a review. *Artif Intell Rev.* 53(4):2381–2423. doi:10.1007/s10462-019-09736-1.
- Birant D, Kut A. 2007. ST-DBSCAN: an algorithm for clustering spatial-temporal data. *Data Knowl Eng.* 60(1):208–221. doi:10.1016/j.datak.2006.01.013.
- Gao J. 2024. R-Squared (R²) – How much variation is explained? . *Res Methods Med Heal Sci.* 5(4):104–109. doi:10.1177/26320843231186398.
- Glencross DA, Ho TR, Camiña N, Hawrylowicz CM, Pfeffer PE. 2020. Air pollution and its effects on the immune system. *Free Radic Biol Med.* 151 January:56–68. doi:10.1016/j.freeradbiomed.2020.01.179.
- Li X, Li L, Chen Longgao, Zhang T, Xiao J, Chen Longqian. 2022. Random Forest Estimation and Trend Analysis of PM2.5 Concentration over the Huaihai Economic Zone, China (2000–2020). *Sustain.* 14(14). doi:10.3390/su14148520.
- Manalu DJ, Rahmawati R, Widiharis T. 2021. Pengelompokan titik gempa di Pulau Sulawesi menggunakan algoritma ST-DBSCAN (Spatio Temporal-Density Based Spatial Clustering Application with Noise). *J Gaussian.* 10(4):554–561. doi:10.14710/j.gauss.v10i4.29499.
- Panuju AYT, Usman M. 2023. PM2.5 concentration pattern in ASEAN countries based on population density. *Procedia Eng Life Sci.* 4 June. doi:10.21070/pels.v4i0.1385.
- Pelánek R. 2015. Metrics for Evaluation of Student Models. *JEDM - J Educ Data Min.* 7(2):1–19. <http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/JEDM087>.
- Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa L da F, Rodrigues FA. 2019. *Clustering algorithms: a comparative approach*. Volume ke-14.
- Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 20 C:53–65. doi:10.1016/0377-0427(87)90125-7.
- World Health Organization. 2021. *WHO Global Air Quality Guidelines*.