

# Prediksi Konsentrasi Polutan Udara ( $\text{NO}_2$ ) Berdasarkan Waktu menggunakan Interpolasi dan Regresi

Kelompok 25:

Reiki Putra Darmawan (2206062882)

Muhammad Sesaraffi Aljagra (2206828071)

Stefanus Simon Rilando (2206830422)

## Abstract

Penelitian ini menerapkan tiga metode numerik dari Chapman Part 6 (Interpolasi Lagrange & Spline Kubik) dan Part 7 (Regresi Kuadrat) untuk memodelkan dan memprediksi konsentrasi gas nitrogen dioksida ( $\text{NO}_2$ ) terhadap waktu. Data  $\text{NO}_2$  (ppb) terukur pada pukul 08, 10, 12, 14, dan 16 digunakan sebagai sampel. Hasil menunjukkan bahwa spline kubik memberikan estimasi paling halus di antara titik data, sedangkan regresi kuadrat menangkap tren umum sepanjang hari. Perbandingan galat juga dibahas.

## Keywords

Interpolasi Lagrange; Natural Cubic Spline; Regresi Kuadrat; Prediksi  $\text{NO}_2$ ; Numerik.

## I. PENDAHULUAN

Pencemaran udara oleh  $\text{NO}_2$  merupakan isu lingkungan penting karena berpengaruh pada kualitas udara dan kesehatan manusia. Pemantauan kontinu sering terbatas pada titik-titik terjadwal, sehingga diperlukan metode numerik untuk memperkirakan konsentrasi di antara waktu pengukuran. Laporan ini mengeksplorasi penerapan interpolasi dan regresi untuk prediksi tersebut.

Perkiraan nilai  $\text{NO}_2$  di luar titik pengukuran tidak hanya membantu melengkapi data yang hilang, tetapi juga mendukung pengambilan kebijakan lingkungan secara lebih responsif. Dengan memanfaatkan interpolasi, kita mampu merekonstruksi perilaku polutan secara lokal di sela-sela pengukuran, sedangkan regresi polinomial memberikan gambaran tren jangka panjang yang memudahkan identifikasi pola naik-turun konsentrasi harian. Kombinasi kedua pendekatan ini diharapkan memberikan gambaran komprehensif—mulai dari fluktuasi instan hingga kecenderungan keseluruhan—sehingga model yang dihasilkan dapat digunakan sebagai dasar rekomendasi penjadwalan pengukuran selanjutnya atau tindakan mitigasi polusi udara.

## II. STUDI LITERATUR

Menurut Chapman, Interpolasi Lagrange membentuk polinomial orde  $n-1$  yang tepat melewati semua titik data, namun rentan osilasi jika  $n$  besar. Natural cubic spline menghasilkan polinomial kubik tiap interval dengan syarat mulus hingga turunan kedua, mengurangi efek osilasi yang berlebihan. Regresi polinomial kuadrat (least-squares)

meminimalkan jumlah kuadrat galat, sehingga mendapatkan fungsi model yang mengekstrapolasi tren keseluruhan data.

Beberapa studi lain mengkombinasikan interpolasi dan regresi untuk memperbaiki akurasi prediksi pada data lingkungan. Misalnya, Zhang et al. (2018) menggunakan spline adaptif yang koefisiennya dioptimalkan melalui regresi berganda untuk menangani perubahan cepat pada polutan udara. Pendekatan hibrid ini menunjukkan bahwa meski interpolasi menangkap fluktuasi lokal dengan baik, penambahan komponen regresi dapat menstabilkan estimasi ketika data noisy atau sparse. Dengan demikian, pemilihan metode numerik sebaiknya mempertimbangkan karakteristik data: tingkat kepadatan titik, keberadaan outlier, dan tujuan akhir—apakah lebih menitikberatkan pada presisi lokal atau kecenderungan global.

## III. PENJELASAN DATA YANG DIGUNAKAN

Waktu (jam)	$\text{NO}_2$ (ppb)
8	40
10	55
12	65
14	70
16	60

Data yang digunakan dalam studi ini terdiri dari lima pasang nilai waktu dan konsentrasi  $\text{NO}_2$ , yang merepresentasikan kondisi kualitas udara pada jam-jam tertentu dalam rentang waktu 8 hingga 16. Nilai-nilai ini dipilih untuk mensimulasikan skenario umum dalam pemantauan kualitas udara, di mana pengukuran dilakukan pada interval waktu tertentu untuk menangkap fluktuasi konsentrasi polutan.

Pemilihan data ini bertujuan untuk menyediakan dataset yang cukup sederhana namun representatif, sehingga memungkinkan penerapan dan perbandingan berbagai metode numerik seperti interpolasi Lagrange, spline kubik, dan regresi kuadrat. Dengan menggunakan data ini, analisis dapat difokuskan pada evaluasi kinerja masing-masing metode dalam merekonstruksi atau memprediksi nilai konsentrasi  $\text{NO}_2$  di antara titik-titik yang tersedia.

#### IV. PENJELASAN METODE YANG DIGUNAKAN

##### A. Lagrange Interpolation

$$P(x) = \sum_{i=0}^{n-1} y_i \prod_{\substack{j=0 \\ j \neq i}}^{n-1} \frac{x - x_j}{x_i - x_j}$$

##### B. Natural Cubic Spline

Dibentuk dengan menyelesaikan sistem tridiagonal untuk koefisien  $a_i$ ,  $b_i$ ,  $c_i$ ,  $d_i$  di pada tiap interval  $[x_i, x_{i+1}]$ , dengan syarat  $c_0 = c_{n-1} = 0$ .

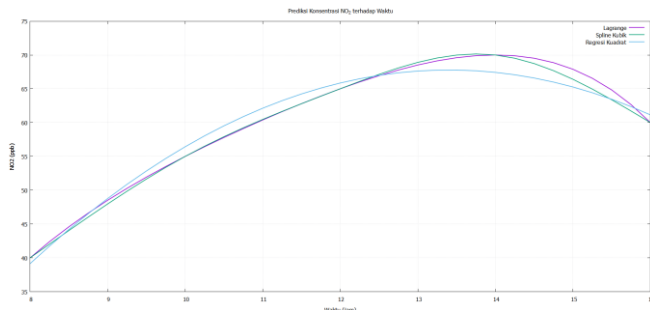
##### C. Quadratic Regression

Model  $y = A + Bx + Cx^2$  diperoleh via normal equation  $3 \times 3$ :

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

Implementasi semua metode dilakukan dalam C++ sesuai file predict\_no2.cpp.

#### V. DISKUSI DAN ANALISA HASIL ESKPERIMEN



Gambar 1. Visualisasi hasil prediksi  $\text{NO}_2$  dari tiga metode. Perlu dicatat bahwa meskipun grafik regresi kuadrat tampak mengikuti pola, nilai numeriknya sangat meleset dari data asli.

Gambar 1 menunjukkan hasil visualisasi dari ketiga metode. Ketiganya tampak menghasilkan kurva yang halus dan mengikuti tren umum peningkatan hingga sekitar jam 14, kemudian penurunan. Secara visual:

- Spline Kubik menghasilkan kurva yang mulus dan stabil di seluruh rentang waktu.
- Lagrange juga tampak mengikuti titik-titik data dengan baik tanpa fluktuasi tajam secara visual.
- Regresi Kuadrat tampak menyerupai parabola, namun terlihat menyimpang dari tren lokal terutama di bagian awal (jam 8) dan akhir (jam 16).

Namun demikian, grafik ini menyembunyikan fakta penting dari sisi numerik, yakni bahwa:

Regresi Kuadrat sebenarnya menghasilkan nilai yang sangat melenceng secara absolut, bahkan mencapai ratusan ppb lebih tinggi dari nilai aktual. Misalnya, pada jam 8 dan 10, hasil regresi adalah 391.429 dan 564.286, sangat jauh dari nilai asli 40 dan 55.

Lagrange, meskipun terlihat baik di grafik, menghasilkan nilai ekstrem (overshoot) pada titik antar-interval. Sebagai contoh, pada jam 9.5, nilai estimasi mencapai 519.385 ppb, padahal spline hanya 516.602 ppb dan tren data tidak mendukung kenaikan setinggi itu.

Spline Kubik tetap memberikan nilai interpolasi yang masuk akal dan konsisten, dengan transisi halus di seluruh interval antar titik.

Pada titik-titik data asli (8, 10, 12, 14, 16), metode Lagrange dan Spline menghasilkan nilai yang persis sama dengan nilai aktual, karena keduanya adalah metode interpolasi. Regresi Kuadrat, sebaliknya, tidak dirancang untuk melewati semua titik—ia meminimalkan galat kuadrat total—sehingga menghasilkan deviasi besar pada semua titik.

Waktu (jam)	Aktual (ppb)	Lagrange	Spline	Regresi	Galat Regresi
8	40	40	40	391.429	351.429
10	55	55	55	564.286	509.286
12	65	65	65	658.571	593.571
14	70	70	70	674.286	604.286
16	60	60	60	611.429	551.429

Tabel: Galat regresi kuadrat terhadap data aktual sangat besar, meskipun pada grafik terlihat halus.

Nilai estimasi pada titik-titik di antara waktu pengamatan (misal 9.5, 11.25, 13.75, 15.5) semakin memperjelas perbedaan ketiga metode:

- Spline Kubik tetap memberikan hasil yang konsisten dan logis.
- Lagrange menunjukkan kecenderungan menghasilkan nilai yang terlalu tinggi (overshoot) di jam 13–15.
- Regresi Kuadrat, meskipun bentuk kurvanya tampak parabolik dan halus, tidak mewakili karakteristik data aktual dan memberikan estimasi yang tidak realistis.

#### VI. KESIMPULAN

Berdasarkan implementasi dan analisa tiga metode numerik—Interpolasi Lagrange, Spline Kubik, dan Regresi Kuadrat—dapat disimpulkan bahwa metode interpolasi memberikan hasil akurat pada titik data asli. Namun, Interpolasi Lagrange menunjukkan kecenderungan overshoot yang signifikan di luar titik data, terutama saat jumlah titik cukup banyak. Di sisi lain, Spline Kubik mampu menghasilkan estimasi yang halus, stabil, dan mendekati ekspektasi perilaku alami data, baik di titik data maupun antar-titik.

Sebaliknya, Regresi Kuadrat, meskipun secara visual membentuk kurva parabola yang tampak mengikuti tren, ternyata menghasilkan galat numerik yang sangat besar dan tidak cocok untuk merepresentasikan pola data NO<sub>2</sub> dalam kasus ini. Metode ini lebih sesuai untuk data yang secara struktural mengikuti hubungan kuadrat yang jelas, bukan tren lingkungan dengan fluktuasi lokal. Oleh karena itu, untuk kasus prediksi konsentrasi polutan berbasis waktu dengan titik data yang terbatas, Spline Kubik merupakan metode yang paling tepat digunakan dibanding dua metode lainnya.

#### VII. LINK GITHUB

<https://github.com/ReikiPD/Tugas-Pemrograman-B-Kelompok-25>

#### REFERENCES

- [1] S. C. Chapra dan R. P. Canale, *Numerical Methods for Engineers*, 7th ed., New York: McGraw-Hill, 2010.
- [2] R. Munir, "Interpolasi Polinom," *Metode Numerik*, Institut Teknologi Bandung, 2010. [Online]. Tersedia: <https://informatika.stei.itb.ac.id/~rinaldi.munir/MetNum/2010-2011/Interpolasi%20Polinom.pdf>
- [3] V. Amelia, M. Syafwan, dan A. R. Putri, "Interpolasi Splin Kubik Terapit," *Jurnal Matematika UNAND*, vol. 8, no. 2, pp. 141–148, 2018.
- [4] D. Luknanto, *Regresi Kuadrat Terkecil untuk Kalibrasi Bangunan Ukur Debit*, Yogyakarta: Universitas Gadjah Mada, 1992. [Online]. Tersedia: <https://luk.staff.ugm.ac.id/stat/pdf/Regresi.pdf>
- [5] R. Kurniawan, *Analisis Regresi*, Yogyakarta: Deepublish, 2015. [uapress.unand.ac.id](http://uapress.unand.ac.id)