Jurnal Komputer dan Informatika
Vol 19 No 1, Oktober 2024: hlm 1 - 8

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

PERBANDINGAN ALGORITMA CLUSTERING K-MEANS, GAUSSIAN MIXTURE MODEL, DAN DBSCAN PADA DATA INDEKS STANDAR PENCEMAR UDARA (ISPU) DI PROVINSI DKI JAKARTA

Apriyanto Chandra
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: apriyanto.535210032@stu.untar.ac.id.,

ABSTRAK

Tujuan dari penelitian ini adalah untuk membandingkan kinerja tiga algoritma clustering, yaitu *K-Means Clustering*, *Gaussian Mixture Model* (GMM), dan DBSCAN, saat menganalisis data Indeks Standar Pencemar Udara (ISPU) Provinsi DKI Jakarta. Parameter kualitas udara seperti nitrogen dioksida, ozon, sulfur dioksida, karbon monoksida, dan PM10 digunakan. Proses penelitian mencakup tahap preprocessing, di mana nilai yang tidak ada digunakan dengan metode imputasi rata-rata dan data dinormalisasi dengan *Standard Scaler*. Dengan nilai Silhouette Score sebesar 0.2784, *Calinski-Harabasz Score* sebesar 403.111, dan *Davies-Bouldin Index* sebesar 1.2481, *K-Means Clustering* menunjukkan hasil terbaik dari tiga metrik clustering. Sementara itu, dari ketiga matrik penilaian, GMM dan DBSCAN menunjukkan hasil yang lebih rendah. Hasilnya menunjukkan bahwa algoritma *K-Means Clustering* melakukan clustering data ISPU Jakarta dengan lebih baik daripada algoritma lain dengan keunggulan dalam kekompakan dan pemisahan cluster yang lebih baik. Dengan penggunaan algoritma clustering tambahan dan analisis data yang lebih mendalam, penelitian ini membuka peluang pengembangan lebih lanjut.

Kata kunci, *—K-Means Clustering, Gaussian Mixture Model, DBSCAN, Clustering, Data Mining*

ABSTRACT

The purpose of this study is to compare the performance of three clustering algorithms, namely K-Means Clustering, Gaussian Mixture Model (GMM), and DBSCAN, when analyzing the Air Pollution Standard Index (ISPU) data of DKI Jakarta Province. Air quality parameters such as nitrogen dioxide, ozone, sulfur dioxide, carbon monoxide, and PM10 were used. The research process includes a preprocessing stage, where missing values are used with the average imputation method and the data is normalized with Standard Scaler. With a Silhouette Score of 0.2784, Calinski-Harabasz Score of 403.111, and Davies-Bouldin Index of 1.2481, K-Means Clustering showed the best results of the three-clustering metrics. Meanwhile, of the threes coring metrics, GMM and DBSCAN showed inferior results. The results show that the K-Means Clustering algorithm performs better clustering of Jakarta ISPU data than other algorithms, with advantages in compactness and better cluster separation. With the use of additional clustering algorithms and more in-depth data analysis, this research opens up opportunities for further development.

Keywords— *K-Means Clustering, Gaussian Mixture Model, DBSCAN, Clustering, Data Mining*

1. PENDAHULUAN

Sekarang, polusi udara adalah masalah lingkungan umum di seluruh dunia. *World Health Organization* (WHO) melaporkan bahwa hingga 2 juta jiwa meninggal akibat polusi udara setiap tahun. Pada dasarnya, polusi udara disebabkan oleh proses alam dan aktivitas manusia. Mayoritas masalah pencemaran udara berasal dari penggunaan bahan bakar fosil dan aktivitas industri. Oleh karena itu, polusi udara saat ini juga merupakan masalah yang sangat penting bagi Indonesia. Karena populasi yang besar dan penggunaan kendaraan yang tinggi di Indonesia, gas buang yang dihasilkan menyebabkan pencemaran udara yang tinggi. Selain itu, ada banyak kawasan industri yang aktivitasnya tidak memenuhi regulasi emisi, yang menyebabkan kualitas udara di banyak tempat di Indonesia menjadi lebih buruk [1].

Kualitas udara di wilayah DKI Jakarta dapat diketahui melalui pengukuran yang dilakukan oleh Dinas Lingkungan Hidup Provinsi DKI Jakarta. Indeks Standar Pencemar Udara (ISPU) adalah indeks yang mengumpulkan hasil pengukuran ini. ISPU adalah laporan dari hasil pemantauan kualitas udara yang menunjukkan kualitas udara yang bersih atau tercemar serta dampaknya terhadap kesehatan. ISPU menyajikan informasi tentang kualitas udara dalam bentuk angka dan tidak dalam satuan. ISPU juga menunjukkan kondisi kualitas udara di lokasi tertentu.Indeks ini dapat digunakan untuk memberikan informasi tentang kualitas udara di DKI Jakarta, terutama untuk membantu dalam merencanakan cara untuk mengurangi pencemaran udara [2].

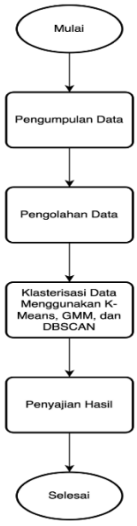
Indeks Standar Pencemar Udara adalah ukuran tanpa satuan yang menunjukkan kualitas udara di daerah tertentu. Indikator ini didasarkan pada efeknya terhadap kesehatan manusia, nilai estetika, dan keberadaan makhluk hidup di Bumi. Nilai ISPU diperoleh melalui sistem pemantauan kualitas udara otomatis, juga dikenal sebagai AQMS. Tujuan dari penggunaan ISPU adalah untuk memberikan informasi yang seragam tentang kualitas udara kepada masyarakat pada waktu dan lokasi tertentu, serta sebagai acuan bagi upaya pemerintah pusat dan daerah untuk mengendalikan pencemaran udara. Ada banyak zat polutan udara, termasuk karbon monoksida (CO), sulfur dioksida (SO2), Ozon permukaan (O3), oksida nitrogen (NOx), dan partikulat zat (PM10) [3].

Data mining adalah teknik yang dapat digunakan untuk melakukan pengelompokan data dalam skala besar. Tujuan data mining adalah untuk menemukan pola, hubungan, dan tren yang tersembunyi dalam data untuk mendapatkan informasi yang lebih baik tentang berbagai topik. Salah satu metode data mining adalah clustering. Pada dasarnya, clustering digunakan untuk mencari dan mengelompokkan data yang memiliki karakteristik yang serupa (*similarity*) satu sama lain. *K-Means, Fuzzy C-Means, Mixture Modeling K-Medoids, Single Linkage, Complete Linkage, Average Linkage, dan Average Group Linkage* adalah beberapa dari banyak metode pengelompokan *clustering* yang dapat digunakan. Dalam penelitian ini, peneliti menggunakan algoritma *K-Means, Gaussian Mixture Model, dan DBSCAN* sebagai metode untuk mengatasi masalah tersebut [4].

2. METODE PENELITIAN

2.1 Rancangan Penelitian

Saat melakukan penelitian untuk klasterisasi, kami mengikuti diagram alur pada gambar 1 untuk tahapan penelitian [5].



Gambar 1 Tahapan Penelitian

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

2.2 Dataset

Untuk mendukung proses penelitian ini, jelas dibutuhkan data untuk melatih sistem. Data ini terletak disitus Satudata Jakarta dan disebut Data Indeks Standar Pencemar Udara (ISPU) di Provinsi DKI Jakarta [6].

2.3 Preprocessing

Serangkaian tindakan atau prosedur yang dilakukan pada data sebelum digunakan dalam pemodelan atau analisis. *Preprocessing* dilakukan untuk membersihkan, menormalkan, dan mempersiapkan data sehingga dapat diolah lebih baik oleh algoritma analisis atau model pembelajaran mesin. Data sebelumnya telah dibersihkan, diperbaiki, dan diubah tanpa mengubah informasi yang ada di dalamnya [7].

2.4 Algoritma

Dalam penelitian ini, algoritma *K-Means Clustering*, *Gaussian Mixture Model*, dan *DBSCAN* digunakan untuk *clustering* data indeks standar pencemar udara di provinsi DKI Jakarta [8].

2.5 Tahapan Clustering

Mencari dan mengelompokkan data yang memiliki ciri-ciri yang menyerupai satu sama lain dikenal sebagai *clustering*. Salah satu tujuan utama dari metode *clustering* adalah untuk mengelompokkan sejumlah besar data atau objek ke dalam kelompok, atau grup, sehingga setiap kelompok dapat berisi jumlah data yang seminimal mungkin. Tujuan utama dari metode *clustering* adalah untuk menempatkan objek dalam kelompok yang dekat satu sama lain dan membuat jarak antara kelompok sekecil mungkin. Ini berarti bahwa objek dalam kelompok tertentu sangat mirip satu sama lain dan berbeda dari yang ada di kelompok lain [9].

2.6 Model K-Means Clustering

Clustering dilakukan oleh algoritma *K-Means* berdasarkan data yang dikumpulkan dan hasil yang ingin dicapai di akhir proses. Oleh karena itu, ada aturan untuk jumlah cluster yang diperlukan untuk menggunakan algoritma ini. Algoritma ini diuji dalam penggalian ilmu pengetahuan dari *database* kehadiran pertemuan tenaga non-PNS. Data ini diolah dari semester sebelumnya dan dikelompokkan ke dalam beberapa klaster. *Clustering* Algoritma *K-Means* dapat memberitahu Anda tentang jadwal pertemuan yang akan diadakan dalam bulan berjalan. Studi ini digunakan sebagai dasar untuk menilai kinerja pendidikan dan pengajaran tenaga pendidik atau dosen tidak tetap (non-PNS) dalam rapat konsorsium keilmuan [10].

Algoritma *k-means* hanya mengambil beberapa sampel dari seluruh populasi komponen yang diperoleh untuk digunakan sebagai pusat *cluster* awal. Pusat *cluster* awal ini dipilih secara acak dari populasi data. Setelah menguji setiap komponen ke dalam jumlah populasi data tersebut, algoritma *k-means* akan menandai setiap komponen ke dalam salah satu pusat *cluster* yang telah dijelaskan sebelumnya, berdasarkan jarak minimum antar komponen [11].

2.7 Model Gaussian Mixture Model

Model statistik yang disebut *Gaussian Mixture Model* (GMM) menggabungkan berbagai distribusi *Gaussian* untuk menggambarkan distribusi data kompleks. Dalam model ini, data dianggap berasal dari berbagai bagian distribusi *Gaussian* yang berbeda. Digunakan dalam berbagai bidang, seperti *clustering*, pengurangan dimensi, pemodelan distribusi data, dan restorasi gambar, GMM bertujuan untuk melakukan clustering. Menggunakan GMM karena bentuk fungsi densitasnya yang sederhana, dan tidak membutuhkan banyak parameter. Ada banyak model yang dapat digunakan untuk clustering dalam GMM. Model ini bergantung pada bentuk geometris yang diciptakan oleh komponen *Gaussian*. Sebuah *cluster* diwakili oleh setiap angka distribusi Gaussian dalam GMM. Kombinasi dari rata-rata dan varian akan mewakili setiap distribusi Gaussian. Tujuan *clustering* GMM adalah untuk menemukan parameter model yang paling sesuai dengan data. Ada banyak model yang digunakan untuk clustering dengan memperhatikan bentuk geometris. Bentuk geometris ini terdiri dari bagian *Gaussian* dengan berbagai parameter. Pada persamaan (1) [12].

$$f(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}\,e^{\frac{-(X-\mu)^2}{2\sigma^2}}} \#(1)$$

2.8 Model DBSCAN

Salah satu algoritma *clustering*, *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), menggunakan kepadatan data untuk membentuk klaster. Ini juga dapat mengidentifikasi klaster dengan berbagai bentuk dan ukuran, dan juga dapat menangani suara yang ada dalam data. Dalam tahap clustering, algoritma DBSCAN menggunakan Epsilon dan MinPts. Konsep dasar DBSCAN adalah ketercapaian kepadatan dan keterhubungan kepadatan, yang bergantung pada parameter radius maksimal (epsilon) dan jumlah objek minimum dalam cluster.

1. Epsilon (eps): Jarak antara objek dan objek di sekitarnya. Objektif yang memenuhi nilai ini dianggap memiliki hubungan yang dekat dengan objek yang sedang diamati.
2. Jumlah Nilai Minimal (MinPts): Jumlah minimal objek yang diperlukan untuk membentuk sebuah cluster jika saling berdekatan.
3. Objek Tepi: Objek yang berdekatan dengan objek core, sedangkan objek core adalah objek yang berdekatan dengan nilai epsilon yang diberikan. Anomaly atau pencilan adalah objek yang tidak mencakup inti atau batas.

   Tahapan berikut menggunakan algoritma DBSCAN:

   a. Temukan nilai parameter minPts dan epsilon.
   b. Temukan titik awal atau p secara acak.
   c. Ulangi langkah 3 hingga 5 sampai semua titik di proses selesai.
   d. Temukan epsilon yang ketercapaian kepadatan terhadap p menggunakan

$$E(X,Y) = \sqrt{\sum_{i=0}^{n} (X_i - Y_i)} \#(2)$$

   e. Core point adalah titik p jika epsilon lebih besar dari minPts.
   f. Jika p adalah titik batas dan tidak ada titik lain yang dapat dicapai karena kepadatan p, proses dilanjutkan ke titik selanjutnya [13].

2.9 Metode Evaluasi

Untuk mengetahui keakuratan metode perhitungan klasterisasi yang digunakan, digunakan tiga metode untuk mengetahui keakuratan metode perhitungan klasterisasi yaitu *Silhouette Score*, *Calinski-Harabasz Index*, dan *Davies Bouldin* [5]

2.10 *Silhouette Score*

Silhouette Score, juga dikenal sebagai koefisien *Silhouette*, adalah ukuran yang digunakan untuk mengevaluasi keakuratan metode pengelompokkan data. Metrik ini ditampilkan dalam bentuk grafik. Siluet untuk setiap klaster dibuat berdasarkan perbandingan kerapatan dan pemisahnya. Keberadaan objek dalam kelompok ditunjukkan dalam siluet ini. Plot yang terdiri dari beberapa siluet menunjukkan klasifikasi. Kualitas relatif kluster dan gambaran umum konfigurasi data ditunjukkan dalam plot ini. Lebar siluet rata-rata membantu mengevaluasi validitas pengelompokan dan menentukan jumlah cluster yang tepat [14].

2.11 *Calinski-Harabaz Index*

Pertama, nilai rentang K yang diinginkan dihitung untuk memperoleh indeks *Calinski Harabasz.* Kemudian kita melakukan prediksi klaster berdasarkan nilai setiap K. Nilai *Calinski* akan diberikan kepada setiap nilai K. Kita akan membuat peta dengan nilai ini. Cluster terbaik berdasarkan pendekatan ini memiliki nilai terkecil dari indeks *Calinski* [15].

2.12 Davies Bouldin

Dalam analisis data, indeks *Davies-Bouldin* digunakan untuk mengukur kualitas klastering. Tujuan indeks ini adalah untuk mengukur seberapa baik klaster yang dibuat oleh algoritma klastering memisahkan kelompok data yang berbeda dan mendekati pusat klasternya; semakin rendah nilai indeks, semakin baik klaster yang dihasilkan [16].

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

3. HASIL DAN PEMBAHASAN

Untuk mendapatkan arsip data pemantauan Indeks Standar Pencemar Udara (ISPU) Kota Jakarta, penelitian ini menggunakan data publik yang tersedia secara online di https://satudata.jakarta.go.id/[17]. Kemudian, eksperimen clustering dilakukan menggunakan tiga algoritma: *K-Means, Gaussian Mixture Model (GMM), dan DBSCAN*. Tabel 1 menunjukkan rincian eksperimen dan parameter yang diuji untuk masing-masing algoritma.

Tabel 1 Tabel Eksperimen

No	Algoritma	Parameter
1.	K-Means Clustering	n_clusters = 3
2.	K-Means Clustering	n_clusters = 4
3.	K-Means Clustering	n_clusters = 5
4.	K-Means Clustering	n_clusters = 6
5.	K-Means Clustering	n_clusters = 7
6.	K-Means Clustering	n_clusters = 8
7.	K-Means Clustering	n_clusters = 9
8.	K-Means Clustering	n_clusters = 10
9.	K-Means Clustering	n_clusters = 11
10.	K-Means Clustering	n_clusters = 12
11.	GMM	n_components = 3
12.	GMM	n_components = 4
13.	GMM	n_components = 5
14.	GMM	n_components = 6
15.	GMM	n_components = 7
16.	GMM	n_components = 8
17.	GMM	n_components = 9
18.	GMM	n_components = 10
19.	GMM	n_components = 11
20.	GMM	n_components = 12
21.	DBSCAN	eps = 0.5
22.	DBSCAN	eps = 0.7
23.	DBSCAN	eps = 0.9
24.	DBSCAN	eps = 0.11
25.	DBSCAN	eps = 0.13
26.	DBSCAN	eps = 0.15
27.	DBSCAN	eps = 0.17
28.	DBSCAN	eps = 0.19
29.	DBSCAN	eps = 0.21
30.	DBSCAN	eps = 0.23

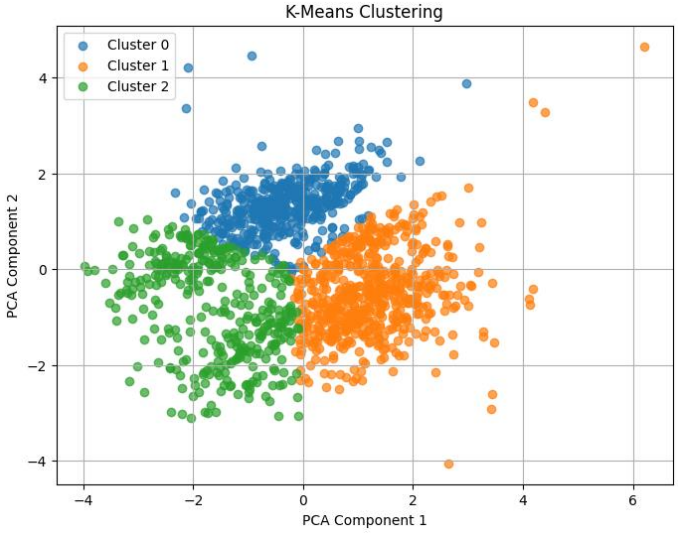
Hasil eksperimen digambarkan dalam tabel. Tujuan penelitian ini adalah untuk mengetahui bagaimana model klasterisasi *K-Means, Gaussian Mixture Model*, dan *DBSCAN* bekerja saat klasterisasi Indeks Standar Pencemar Udara (ISPU) Kota Jakarta, seperti yang ditunjukkan pada Tabel 3. *Silhouette Score, Davis Bouldin*, dan *Calinski Harabasz Index* digunakan untuk menilai.

Tabel 2 Hasil Percobaan

Algoritma	Silhouette Score	Calinski Harabaz	Davies Bouldin
K-Means Clustering	0.2313	403.111	1.3956
K-Means Clustering	0.2173	368.995	1.404
K-Means Clustering	0.2498	379.1544	1.2619
K-Means Clustering	0.2557	364.4438	2.2627
K-Means Clustering	0.2784	361.7609	1.2481
K-Means Clustering	0.2552	341.0109	1.2860
K-Means Clustering	0.2506	314.9247	1.2881
K-Means Clustering	0.2470	302.1520	1.2971

Algoritma	Silhouette Score	Calinski Harabaz	Davies Bouldin
K-Means Clustering	0.2338	289.5527	1.3254
K-Means Clustering	0.2386	284.8145	1.3229
GMM	0.1613	221.754	2.7367
GMM	0.1723	199.781	2.5644
GMM	0.1834	189.4945	2.7075
GMM	0.1254	196.4117	2.8979
GMM	0.2001	238.6479	1.5771
GMM	0.2146	240.3337	1.5125
GMM	0.2229	228.2827	1.6302
GMM	0.1736	176.7432	2.3878
GMM	0.1938	185.9600	2.0315
GMM	0.1825	166.7535	2.0517
DBSCAN	-0.2854	23.9932	1.5328
DBSCAN	-0.0123	71.4608	1.8187
DBSCAN	0.1421	92.3708	4.5192
DBSCAN	0	0	0
DBSCAN	0	0	0
DBSCAN	0	0	0
DBSCAN	0	0	0
DBSCAN	0	0	0
DBSCAN	0	0	0
DBSCAN	0	0	0

Algoritma *K-Means Clustering* menunjukkan kinerja terbaik dibandingkan dengan algoritma *Gaussian Mixture Model (GMM)* dan *DBSCAN*, menurut hasil evaluasi yang dilakukan dengan menggunakan metrik *Silhouette Score*, *Calinski-Harabasz Score*, dan *Davies-Bouldin Index*. Selain itu, algoritma *K-Means Clustering* juga memperoleh nilai *Calinski-Harabasz Score* sebesar 403.111, *Silhouette Score* tertinggi sebesar 0.2784, dan nilai *Davis Bouldin Index* 1.2481 yang menunjukkan bahwa *cluster* yang dihasilkannya memiliki kekompakan yang baik dapat dilihat pada Gambar 2



Gambar 2 K-Means Clustering

Sebaliknya, dalam semua metrik evaluasi, GMM dan DBSCAN menunjukkan kinerja yang lebih buruk. Meskipun GMM memiliki nilai *Silhouette Score* tertinggi sebesar 0.2229 dan nilai *Calinski-Harabasz* tertinggi sebesar 240.3337, nilai *Davies-Bouldin Index*nya sebesar 1.5125 masih lebih tinggi dari *K-Means*. *DBSCAN* memiliki nilai *Davies-Bouldin Index* yang lebih rendah sebesar 1.5328, tetapi nilai *Silhouett*e Skor Nya sebesar 0.1421 dan nilai C*alinski-Harabasz* tertinggi 92.3708, yang merupakan nilai terendah dari semua algoritma.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

4. KESIMPULAN

Studi ini menganalisis data Indeks Standar Pencemar Udara (ISPU) Provinsi DKI Jakarta menggunakan tiga algoritma clustering: *K-Means Clustering, Gaussian Mixture Model (GMM)*, dan *DBSCAN*. *K-Means Clustering* menunjukkan performa terbaik berdasarkan evaluasi dengan menggunakan *Silhouette Score*, *Calinski-Harabasz Score*, dan *Davies-Bouldin Index*. *Cluster* dengan kekompakan yang tinggi, pemisahan yang ideal, dan jarak yang relatif besar dibandingkan ukurannya dapat dibuat dengan algoritma ini. *K-Means Clustering* mengungguli *GMM* dan *DBSCAN* dengan nilai tertinggi dari metrik evaluasi, seperti *Silhouette Score* sebesar 0.2784, *Calinski-Harabasz Score* sebesar 403.111, dan *Davies-Bouldin Index* serendah 1.2481.

Sementara *GMM* menunjukkan kinerja yang cukup baik, mereka masih kalah dari *K-Means* dalam hal kekompakan dan pemisahan *cluster*. Sebaliknya, *DBSCAN* tidak bekerja dengan baik dengan *dataset* ini, menghasilkan nilai metrik evaluasi yang jauh lebih rendah. Ini mungkin karena parameter epsilon dan jumlah sampel yang rendah yang tidak sesuai dengan distribusi data. Penelitian ini memiliki keuntungan dari pendekatan perbandingan algoritma, yang memungkinkan analisis menyeluruh terhadap hasil *clustering*. Namun, penelitian ini memiliki beberapa keterbatasan, seperti ketergantungan pada parameter algoritma dan kemungkinan bias hasil karena metode *preprocessing* atau pengisian nilai yang tidak ada.

Pengembangan lebih lanjut dapat dilakukan dengan mempelajari algoritma clustering lain, seperti *clustering* hierarchical atau spectral, dan melakukan pengujian dengan berbagai kombinasi parameter untuk meningkatkan kualitas hasil *clustering*. Selain itu, penelitian ini dapat diperluas dengan menggabungkan analisis clustering dengan data tambahan, seperti faktor meteorologi, untuk memperoleh pemahaman yang lebih baik tentang polusi udara di Jakarta.

DAFTAR PUSTAKA

1. A. Riyanto, A. Maheswara, R. Zulianty, V. M. Alegra, A. N. Muhammad, and P. Hukum, “Tanggung Jawab Pemerintah dalam Penyelesaian Masalah Polusi Udara di DKI Jakarta”.
2. A. Amalia *et al.*, “Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor”, [Online]. Available: https://data.jakarta.go.id/.
3. A. Budianita, N. Iman, F. Maisa Hana, and C. Berlian Hakim, “Algoritma K-Nearest Neighbor dan Naive Bayes pada Klasifikasi Tingkat Kualitas Udara Kota Tangerang Selatan,” *Jurnal Informatika dan Rekayasa Perangkat Lunak Komparasi*.
4. S. Sosnegher Ndelawa, R. Dwi Bekti, M. T. Jatipaningrum, F. Astuti, and P. S. Statistika, “Penerapan Metode K-Means Pada Data Ordinal Untuk Pengelompokan Daerah Berdasarkan Kualitas Udara di Daerah Istimewa Yogyakarta,” *Jurnal Statistika Industri dan Komputasi*, vol. 09, no. 02, pp. 60–71, 2024.
5. A. Davyn Daniel, E. Dewayani, and T. Sutrisno, “Analisis Dan Prediksi Data Pemantauan Coronavirus Disease 2019 Di Provinsi Daerah Khusus Ibukota Jakarta Dengan Metode Double Exponential Smoothing,” *Computatio: Journal of Computer Science and Information Systems*, vol. 6, no. 2, pp. 98–106, 2022.
6. A. Prawira and C. Ariya, “Loan Prediction App Using Polynomial Regression,” *Computatio: Journal of Computer Science and Information Systems*, vol. 8, no. 1, pp. 73–85, 2024, [Online]. Available: https://www.kaggle.com/altruistdelhite04/loan-prediction-problem-
7. F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, “Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 273–281, Jan. 2024, doi: 10.57152/malcom. v4i1.1085.
8. N. Hadi and J. Benedict, “Implementasi Machine Learning Untuk Prediksi Harga Rumah Menggunakan Algoritma Random Forest,” *Computatio: Journal of Computer Science and Information Systems*, vol. 8, no. 1, pp. 50–61, 2024, [Online]. Available . https://www.kaggle.com/harlfoxem/housesalesprediction
9. M. Anjelita, A. P. Windarto, A. Wanto, and I. Sudahri, *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Pengembangan Datamining Klastering Pada Kasus Pencemaran Lingkungan Hidup*.

10. I. Virgo, S. Defit, and Y. Yuhandri, "Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering," Jurnal Sistim Informasi dan Teknologi, pp. 23–28, Mar. 2020, doi: 10.37034/jsisfotek. v2i1.17.
11. P. Alkhairi and A. P. Windarto, Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. [Online]. Available: https://seminar-id.com/semnas-sainteks2019.html
12. D. Faidah Yusti, A. Maula Hudzaifa, N. Theresia, and C. Egytia Widiantoro, "Optimalisasi Strategi Pengelompokkan Potensi Padi Sebagai Solusi Efektif Kelangkaan Beras di Jawa Barat".
13. M. Farid, "Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Density Based Spatial Clustering of Application with Noise (DBSCAN) Tugas Akhir," 2024.
14. I. Widaningrum, D. Mustikasari, R. Arifin, S. L. Tsaqila, and D. Fatmawati, "Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen."
15. I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," 2023. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC
16. I. T. Umagapi, B. Umaternate, S. Komputer, P. Pasca Sarjana Universitas Handayani, B. Kepegawaian Daerah Kabupaten Pulau Morotai, and B. Riset dan Inovasi, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa."
17. M. P. A. Budiman and D. Winarso, "Penerapan Algoritma K-Medoids Clusteringuntuk Pengelompokan Bulan Rawan Bencana Kabut Asap di Kota Pekanbaru," *Jurnal Fasilkom*, Apr. 2024.