

Penerapan Data Mining Untuk Clustering Kualitas Udara

Ahmad Rifqi¹, Rima Tamara Aldisa^{2*}

¹Fakultas Teknologi Komunikasi dan Informatika, Informatika, Universitas Nasional, Jakarta, Indonesia

²Fakultas Teknologi Komunikasi dan Informatika, Sistem Informasi, Universitas Nasional, Jakarta, Indonesia

Email: ¹ahmadrifqi1508@gmail.com, ^{2,*}rimatamaraa@gmail.com

Email Koresponding: rimatamaraa@gmail.com

Submitted: 13/12/2023; Accepted: 25/12/2023; Published: 26/12/2023

Abstrak—Kesehatan manusia pada saat ini merupakan kunci terhadap keberlangsungan kehidupan. Kesehatan manusia sangat diperlukan dalam proses perkembangan hidup manusia. Kesehatan lingkungan merupakan terkait dengan keadaan ataupun kondisi yang tersedia pada daerah sekitar dikawasan tempat tinggal baik lingkungan yang kecil ataupun lingkungan yang besar. Kualitas udara merupakan kondisi terhadap udara disekitar. Kualitas udara merupakan hal yang sangat penting bagi kehidupan manusia dikarenakan udaralah yang membantu manusia untuk hidup dengan bernafas. Dengan tersedianya kualitas udara yang baik, maka tentu akan menjadi faktor penting terhadap suatu daerah tersebut, bukan saja terhadap kesehatan tetapi juga terhadap dengan sektor lainnya yang berinteraksi langsung pada area terbuka. Peran pentingnya kualitas udara bagi manusia, maka perlu dilakukan perhatian yang lebih dan diberikan penanganan khusus bagi daerah yang terpapar udara buruk. Hal diatas merupakan sebuah masalah yang sangat penting untuk segera diselesaikan, jika masalah tersebut tidak segera diselesaikan akan berdampak bagi kesehatan. Proses penyelesaian permasalahan diperlukan sebuah cara untuk penyelesaian. Dimana proses pengukuran kualitas udara dapat dilihat berdasarkan dengan kondisi ataupun kriteria tertentu yang terjadi pada suatu daerah. Data mining merupakan sebuah cara yang dilakukan untuk melakukan proses pemecahan masalah dengan melakukan pengolahan data. Dalam proses yang dilakukan pada data mining, terdapat berbagai macam cara penyelesaian. Salah satu yang dapat dipergunakan adalah clustering. Pada clustering sendiri terdapat bebergai macam algoritma seperti DBSCAN, K-Means dan K-Medoids. Pada penelitian ini proses penyelesaian akan menggunakan ketiga algoritma tersebut K-Means, K-Medoids dan DBSCAN. Tujuan penggunaan dari ketiga algoritma tersebut sebagai proses perbandingan hasil yang didapatkan. Dalam proses yang dilakukan pada penyelesaian data mining digunakan teknik clustering dengan menggunakan 3 (tiga) algoritma yaitu K-Means, K-Medoids dan DBSCAN. Adapun hasil yang didapatkan bahwa algoritma K-Means nilai akurasi tertinggi didapatkan pada K=4 dengan nilai 0,843, untuk algoritma K-Medoids nilai tertinggi didapatkan pada K=5 dengan nilai 0,896 dan untuk algoritma DBSCAN nilai tertinggi didapatkan pada K=2 dengan nilai 0,885.

Kata Kunci: Data Mining; Klasterisasi; K-Means; K-Medoids; DBSCAN

Abstract—Human health at this time is the key to the continuity of life. Human health is very necessary in the process of development of human life. Environmental health is related to the circumstances or conditions that exist in the surrounding area where you live, whether in a small environment or a large environment. Air quality is the condition of the surrounding air. Air quality is very important for human life because air is what helps humans to live by breathing. With the availability of good air quality, it will certainly be an important factor for an area, not only for health but also for other sectors that interact directly in open areas. The important role of air quality for humans means that more attention needs to be paid and special treatment is given to areas exposed to bad air. The above is a very important problem that must be resolved immediately, if the problem is not resolved immediately it will have an impact on health. The process of solving problems requires a way to resolve them. Where the process of measuring air quality can be seen based on certain conditions or criteria that occur in an area. Data mining is a method used to carry out the problem solving process by processing data. In the process carried out in data mining, there are various ways of solving it. One thing that can be used is clustering. In clustering itself there are various kinds of algorithms such as DBSCAN, K-Means and K-Medoids. In this research, the solution process will use the three algorithms K-Means, K-Medoids and DBSCAN. The purpose of using these three algorithms is to compare the results obtained. In the process carried out in completing data mining, clustering techniques are used using 3 (three) algorithms, namely K-Means, K-Medoids and DBSCAN. The results obtained were that the K-Means algorithm had the highest accuracy value obtained at K=4 with a value of 0.843, for the K-Medoids algorithm the highest value was obtained at K=5 with a value of 0.896 and for the DBSCAN algorithm the highest value was obtained at K=2 with a value of 0.885.

Keywords: Data Mining; Clustering; K-Means; K-Medoids; DBSCAN

1. PENDAHULUAN

Kesehatan manusia pada saat ini merupakan kunci terhadap keberlangsungan kehidupan. Kesehatan manusia sangat diperlukan dalam proses perkembangan hidup manusia. Dengan kesehatan yang dimiliki maka manusia akan terjamin terhadap keberlangsungan hidup dari manusia tersebut baik untuk orang yang sudah lanjut usia ataupun juga anak bayi. Salah satu yang menjadi dari kesehatan manusia adalah kesehatan dari lingkungan[1], [2].

Kesehatan lingkungan merupakan terkait dengan keadaan ataupun kondisi yang tersedia pada daerah sekitar dikawasan tempat tinggal baik lingkungan yang kecil ataupun lingkungan yang besar. Banyak faktor yang mempengaruhi terhadap kesehatan lingkungan seperti manusia itu sendiri ataupun terkait dengan industri yang berlangsung didaerah tersebut. Salah satu kesehatan lingkungan yang sangat penting adalah kualitas udara[3], [4].

Kualitas udara merupakan kondisi terhadap udara disekitar. Kualitas udara merupakan hal yang sangat penting bagi kehidupan manusia dikarenakan udaralah yang membantu manusia untuk hidup dengan bernafas. Udara yang bersih tentu saja akan berguna dan bermanfaat bagi tubuh manusia. Berlaku sebaliknya, jika udara yang kotor dihirup oleh manusia dikernakan terpapar polusi maka akan menimbulkan penyakit bagi manusia baik terhadap penyakit Infeksi Saluran Pernafasan Akut (ISPA) ataupun penyakit yang fatal lainnya[5], [6].

Dengan tersedianya kualitas udara yang baik, maka tentu akan menjadi faktor penting terhadap suatu daerah tersebut, bukan saja terhadap kesehatan tetapi juga terhadap dengan sektor lainnya yang berinteraksi langsung pada area terbuka. Kualitas udara yang baik bukan saja bermanfaat bagi manusia saja, tetapi juga bagi makhluk hidup lainnya seperti hewan dan tumbuhan yang menjadi pendukung terhadap ekosistem di kehidupan manusia. Maka dari itu sudah sepatutnya untuk diberikan perhatian lebih ataupun penanganan lebih terhadap kualitas udara disuatu daerah agar kiranya dapat membantu dalam menjamin kesehatan manusia[7], [8].

Peran pentingnya kualitas udara bagi manusia, maka perlu dilakukan perhatian yang lebih dan diberikan penanganan khusus bagi daerah yang terpapar udara buruk. Namun permasalahan yang terjadi masih minimnya informasi terkait dengan kondisi dimana suatu daerah memiliki kualitas udara yang buruk. Dengan tidak terdapatnya informasi tersebut maka akan sangat sulit terhadap penanganan yang akan dilakukan. Sehingga proses penyelesaian yang dilakukan tidak berdampak berarti bagi manusia.

Hal diatas merupakan sebuah masalah yang sangat penting untuk segera diselesaikan, jika masalah tersebut tidak segera diselesaikan akan berdampak bagi kesehatan terhadap manusia dan juga akan berdampak pada tumbuhan juga hewan. Dampak yang dihasilkan bisa sampai dengan penyakit akut yang diderita hingga memiliki dampak cukup fatal dalam keberlangsungan hidup. Proses penyelesaian permasalahan diperlukan sebuah cara untuk penyelesaian. Dimana proses pengukuran kualitas udara dapat dilihat berdasarkan dengan kondisi ataupun kriteria tertentu yang terjadi pada suatu daerah. Dimana dalam proses pengambilan keputusan terhadap penentuan kualitas udara tersebut nantinya harus melihat terhadap kondisi yang terjadi sebenarnya dan juga memperhatikan terhadap data – data kualitas udara sebelumnya. Maka dari itu cara yang tepat untuk menyelesaikan permasalahan tersebut dengan menggunakan data mining. Data mining merupakan sebuah cara yang dilakukan untuk melakukan proses pemecahan masalah dengan melakukan pengolahan data. Pengolahan data yang dilakukan pada data mining berdasarkan dengan data – data yang terjadi pada masa lampau dan tersimpan pada gudang data kemudian dilakukan pencocokan kembali dengan data dimasa sekarang untuk nantinya dilakukan proses pengolahan data dan mendapatkan informasi yang berguna untuk pengambilan keputusan. Informasi yang dihasilkan berdasarkan dengan penggalian informasi tersebut berbagai macam bentuk dan hasilnya sesuai dengan kebutuhan data yang diinginkan[9]–[11]. Dalam proses yang dilakukan pada data mining, terdapat berbagai macam cara penyelesaian. Salah satu yang dapat dipergunakan adalah clustering. Clustering merupakan proses pengelompokan dari objek baru terhadap beberapa kelompok bagian yang ditentukan. Proses pembentukan kelompok clustering berdasarkan dengan objek – objek yang dimiliki terhadap data[12]–[14]. Pada clustering sendiri terdapat berbagai macam algoritma seperti DBSCAN, K-Means dan K-Medoids.

Pada penelitian ini proses penyelesaian akan menggunakan ketiga algoritma tersebut K-Means, K-Medoids dan DBSCAN. Tujuan penggunaan dari ketiga algoritma tersebut sebagai proses perbandingan hasil yang didapatkan. Perbandingan yang dilakukan bertujuan untuk mendapatkan hasil yang sangat layak dan dapat dipercaya serta dipergunakan untuk proses pengambilan keputusan.

Beberapa penelitian terdahulu sebagai referensi pendukung dari penelitian seperti yang dilakukan oleh Asri Samsiar Ilmananda dan Habel David Ranglalin pada tahun 2023 dengan judul penelitian “Klasterisasi Negara Pengekspor Beras ke Indonesia Menggunakan Algoritma K-Means Clustering” dimana hasil penelitian yang didapatkan bahwasannya algoritma K-Means Clustering dapat diterapkan untuk mengelompokkan negara-negara pengekspor beras ke Indonesia. Klasterisasi dilakukan dalam tiga cluster, yaitu cluster impor tingkat tinggi (C1), cluster impor tingkat sedang (C2) dan cluster impor tingkat rendah (C3)[15].

Penelitian lainnya yang dilakukan oleh Ahzril Pria Adistya, dkk pada tahun 2023 dengan judul penelitian “Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa” didapatkan hasil bahwasannya Dalam penelitian menggunakan tiga kategori kelulusan yaitu nilai bagus, cukup, dan tidak lulus. Hasil perhitungan dataset dengan metode K-Means didapatkan hasil Accuracy 81%, hal ini menunjukkan bahwa sistem dapat mengklasifikasikan data secara benar[16]. Selain terhadap K-Means juga terdapat penelitian terdahulu dari metode K-Medoids seperti yang dilakukan oleh Yusvi Diana dan Febri Hadi pada tahun 2023 dengan judul penelitian “Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang” maka hasil yang didapatkan dari penelitian Hasil penelitian ini menggunakan algoritma K-Medoids dengan 21 data latih. Hasil akurasi pengelompokan algoritma K-Medoids adalah 97%. Dari hasil yang didapatkan diperoleh data penjualan yang minat dibeli konsumen yaitu : Pantofel Casual Pria dan Pantofel Mary Jane Pria[17]. Penelitian lainnya terhadap algoritma K-Medoids yaitu seperti dilakukan oleh Eka Prasetyaningrum dan Puji Susanti pada tahun 2023 dengan judul penelitian yang diangkat adalah “Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan” dimana pada penelitian yang dilakukan didapatkan hasil Hasil yang didapatkan pada algoritma K-Means mendapatkan hasil nilai DBI (Davies Bouldin Index) yang lebih kecil dengan nilai 0,296 sedangkan hasil algoritma K-Medoids dengan nilai DBI (Davies Bouldin Index) sebesar 0,507. Algoritma terbaik untuk klasterisasi hasil produksi buah-buahan yang ada di kabupaten Kotawaringin Timur adalah algoritma K-Means berdasarkan hasil nilai DBI yang diperoleh[18]. Penelitian terdahulu lainnya sebagai acuan terhadap algoritma DBSCAN seperti yang dilakukan oleh Ipa Rezky Cahyani Pata, dkk pada tahun 2023 dengan judul penelitian yang dilakukan “Segmentasi Pemetaan Pelanggan Potensial Menggunakan Algoritma Dbscan Dengan Rfm Model Berbasis Web” dimana hasil dari penelitian yaitu Berdasarkan pengujian nilai silhouette coefficient yang ditunjukkan menunjukkan bahwa hasil cluster yang terbentuk dapat dikatakan sebagai hasil cluster dengan struktur baik dikarenakan nilai silhouette coefficient

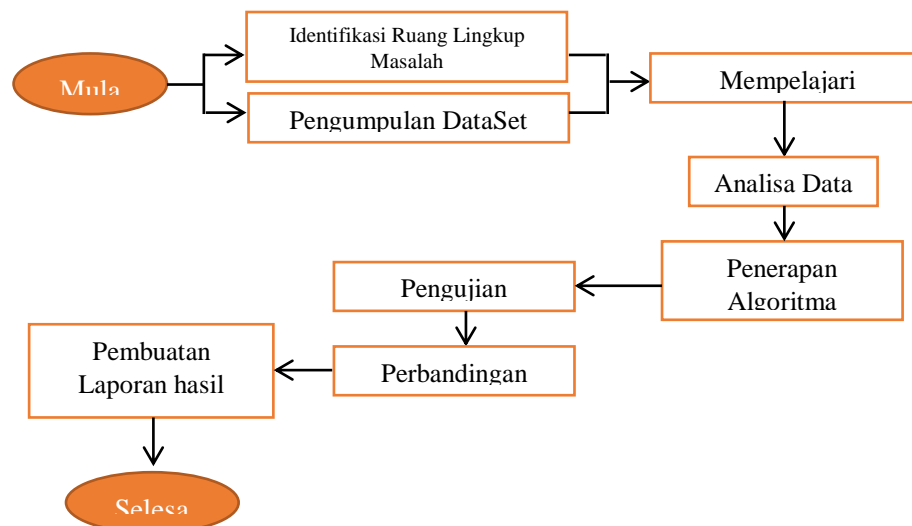
keseluruhan memenuhi berdasarkan interval nilai struktur baik yakni 0,51 – 0,70[19]. Penelitian terakhir yang digunakan pada penelitian yaitu dilakukan oleh Andri Syafrianto dan Eko Riswanto ditahun 2023 dengan judul penelitian “Pengelompokkan Jumlah Kunjungan Mahasiswa ke Perpustakaan Kampus Menggunakan Algoritma DBSCAN” dengan hasil penelitian yang didapatkan bahwa Berdasarkan hasil analisa terhadap 75 mahasiswa dengan nilai epsilon 0,4 dan MinPts 2 maka metode DBSCAN mampu membentuk cluster sebanyak 7 dengan 5 data noise[20].

Berdasarkan dengan penelitian terdahulu diatas maka dari itu, pada penelitian ini akan dilakukan proses perbandingan hasil terhadap clustering kualitas udara. Tujuan dilakukan perbandingan untuk mendapaktan hasil yang lebih baik dipergunakan dalam proses pengambilan keputusannya. Dimana hasil berdasarkan dengan nilai terbaik dari kinerja yang dilakukan.

2. METODOLOGI PENELITIAN

2.1 Metodologi Penelitian

Metodologi penelitian merupakan proses atau alur yang dilakukan pada penelitian dimulai dari proses identifikasi masalah hingga proses penarikan kesimpulan. Adapun metodologi penelitian yang digunakan pada penelitian ini dapat dilihat pada gambar 1 berikut:



Gambar 1. Metodologi Penelitian

2.2 Data Mining

Data mining adalah proses yang menggunakan statistic, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi yang bermanfaat. Data mining didefinisikan sebagai proses penemuan pola dalam data. Berdasarkan tugasnya, data mining dikelompokkan menjadi deskripsi, estimasi prediksi, klasifikasi, klastering, dan asosiasi. Suatu proses penambangan informasi penting dari suatu data. Informasi penting ini didapat dari suatu proses yang amat rumit seperti menggunakan artificial intelligence, teknik statistik, ilmu matematika, machine learning dan lain sebagainya[21], [22]. Jadi, berdasarkan penjelasan diatas dapat disimpulkan bahwa data mining merupakan suatu proses penambangan data dalam jumlah data yang sangat besar dengan menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence terkini. Menurut para ahli tujuan dari penambangan data ini untuk mengekstraksi serta mengidentifikasi suatu data demi informasi tertentu yang berhubungan dengan suatu database besar atau big data[23], [24].

2.3 Algoritma K-Means

K-Means adalah salah satu algoritma clustering yang sangat populer karenakesederhanaan dan kemampuannya dalam menangani data dengan skala besar. Namun demikian algoritma ini sangat sensitif terhadap centroid awal. Perbedaan centroid awal akan memberikan perbedaan hasil clustering dan apabila centroid awal yang diberikan adalah centroid yang tidak baik maka dapat dipastikan hasil clusteringnya juga tidak baik.

Dengan metode pengelompokan berdasarkan jarak, teknik K-means yang dianggap sebagai metode pengelompokan klasik. Salah satu fitur utama yang membuat K-means berguna dalam penelitian ini adalah kemungkinan menentukan terlebih dahulu berapa banyak cluster yang sedang dicari. Jumlah ini mengacu pada parameter K. Dari jumlah ini, poin K dipilih dalam cluster random sebagai pusat cluster. Semua variabel

ditugaskan ke pusat cluster terdekat mereka sesuai dengan metrik jarak Euclidean. Dengan rumus jarak Euclidean Distance sebagai berikut[25]–[28]:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Dimana:

d = Jarak (Euclidean Distance)

x = data

y = centroid

2.4 Algoritma K-Medoids

Algoritma K-Medoids atau Partitioning Around Medoids (PAM) adalah algoritma clustering yang mirip dengan K-Means. Perbedaan dari kedua algoritma ini yaitu algoritma K-Medoids atau PAM menggunakan objek sebagai perwakilan (medoid) sebagai pusat cluster untuk setiap cluster, sedangkan K-Means menggunakan nilai rata-rata (mean) sebagai pusat cluster. Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada algoritma K-Means yang sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada distribusi data. Kelebihan lainnya yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset. Langkah-langkah algoritma K-Medoids[29]–[32]:

1. Inisialisasi pusat cluster sebanyak k (jumlah cluster)
2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan:

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (2)$$

3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru
4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru.
5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan objek baru sebagai medoid.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan cluster beserta anggota cluster masing-masing.

2.5 Algoritma DBSCAN

DBSCAN merupakan algoritma clustering yang dapat menemukan area berdasarkan kerapatan yang tinggi untuk membentuk cluster. Cluster pada algoritma DBSCAN didefinisikan sebagai area yang mempunyai sampel yang padat atau kerapatan yang tinggi dan dibedakan dari area kerapatan yang rendah (noise). Urutan algoritma DBSCAN memiliki 5 langkah secara umum yaitu[33]–[36]:

1. Tentukan parameter Eps dan MinPts
2. Tetapkan titik awal (p) secara random.
3. Hitung jarak p dengan seluruh titik yang lain. Perhitungan jarak dengan rumus Euclidean Distance.

$$\text{Jarak} = \sqrt{(x - xp)^2 + (y - yp)^2} \quad (3)$$

4. Ambil semua titik yang density reachable dari p yang memenuhi Eps dan MinPts. Jika p adalah core point maka cluster terpenuhi.
5. Ulangi langkah 3 sampai 6 sampai semua poin telah diproses. Jika p merupakan border point dan tidak terdapat titik yang memiliki hubungan density-reachable dengan titik awal maka selanjutnya akan memasuki poin lainnya.

2.6 Pengumpulan Data

Sebelum dilakukan proses penelitian terlebih dahulu dilakukan pengumpulan data, dimana pada penelitian ini menggunakan dataset terhadap kualitas udara. Dataset penelitian dapat diakses melalui UCI Repository Machine Learning <https://archive.ics.uci.edu/dataset/360/air+quality> dimana dataset tersebut sudah diakui untuk digunakan oleh banyak penelitian.

3. HASIL DAN PEMBAHASAN

Hal yang utama dalam hidup manusia adalah kesehatan, dimana jika manusia sehat maka dapat melakukan berbagai macam aktifitas. Salah satu pengaruh terhadap kesehatan manusia adalah kesehatan lingkungan. Kesehatan lingkungan merupakan suatu keadaan sekitar didaerah ataupun tempat tinggal. Dimana jika suatu tempat memiliki lingkungan yang sehat maka akan terjadi juga kehidupan yang sehat. Salah satu contoh lingkungan yang sehat adalah kualitas udara. Kualitas udara merupakan hal yang sangat penting bagi manusia, dimana udara sangat penting digunakan untuk bernafas. Kualitas udara juga sangat mempengaruhi terhadap

kesehatan manusia, jika kualitas udara baik maka akan jauh dari penyakit bagi manusia terkhususnya penyakit pernafasan, namun sebaliknya jika kualitas udara buruk maka akan mudah terkena penyakit terhadap saluran pernafasan. Permasalahan tersebut sangatlah penting dan harus segera dilakukan, dimana harus dilakukan pengukuran dan mengetahui keadaan sekitar terhadap kualitas udara. Kualitas udara dapat dilihat berdasarkan dengan kondisi dan kriteria tertentu. Dimana dengan melihat kondisi dari suatu daerah maka nantinya dapat menentukan apakah kualitas udara sudah baik atau belum agar kiranya dapat dilakukan penanganan. Permasalahan tersebut harus segera diselesaikan, jika terlalu lama dibiarkan akan memiliki dampak yang cukup fatal. Penyelesaian permasalahan dapat diselesaikan dengan melakukan pengolahan data, dengan melakukan pengolahan data maka didapatkan informasi untuk proses pengambilan keputusan. Data mining merupakan salah satu cara yang tepat dipergunakan untuk melakukan proses pengolahan data tersebut. Dimana pada data mining nantinya akan melakukan pengolahan data dan mendapatkan kembali informasi – informasi yang dapat dipergunakan untuk pengambilan keputusan. Pada data mining terdapat salah satu teknik yang dapat dipergunakan untuk menyelesaikan masalah yaitu clustering. Clustering merupakan sebuah proses untuk mengelompokan data pada kelompok tertentu. Pada clustering terdapat berbagai macam algoritma seperti K-Means, K-Medoids dan DBSCAN. Pada penelitian ini akan dilakukan proses perbandingan dari ketiga algoritma tersebut. Tujuan dilakukan perbandingan untuk mendapatkan algoritma dengan kinerja lebih baik dan dipergunakan untuk proses pengambilan keputusan.

3.1 Hasil Pengumpulan Data

Sebelum dilakukan proses penelitian terlebih dahulu harus diketahui terhadap data yang akan digunakan nantinya. Pada penelitian ini data merupakan sebuah dataset yang diperoleh dari UCI Repository Machine Learning, pada data tersebut terdapat 13 atribut data dengan jumlah data sebanyak 9357. Ataupun data tersebut dapat dilihat berikut:

Tabel 1. Sampel Data

No	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)
1	2,6	1360	150	11,9	1046	166
2	2	1292	112	9,4	955	103
3	2,2	1402	88	9,0	939	131
4	2,2	1376	80	9,2	948	172
5	1,6	1272	51	6,5	836	131
6	1,2	1197	38	4,7	750	89
7	1,2	1185	31	3,6	690	62
8	1	1136	31	3,3	672	62
9	0,9	1094	24	2,3	609	45
10	0,6	1010	19	1,7	561	-200
...
...
9357	2,2	1071	-200	11,9	1047	265

Lanjutan Tabel 1. Sampel Data

No	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)	T	RH	AH
1	1056	113	1692	1268	13,6	48,9	0,7578
2	1174	92	1559	972	13,3	47,7	0,7255
3	1140	114	1555	1074	11,9	54,0	0,7502
4	1092	122	1584	1203	11,0	60,0	0,7867
5	1205	116	1490	1110	11,2	59,6	0,7888
6	1337	96	1393	949	11,2	59,2	0,7848
7	1462	77	1333	733	11,3	56,8	0,7603
8	1453	76	1333	730	10,7	60,0	0,7702
9	1579	60	1276	620	10,7	59,7	0,7648
10	1705	-200	1235	501	10,3	60,2	0,7517
...
...
9357	654	168	1129	816	28,5	13,1	0,5028

3.2 Pengujian Algoritma K-Means

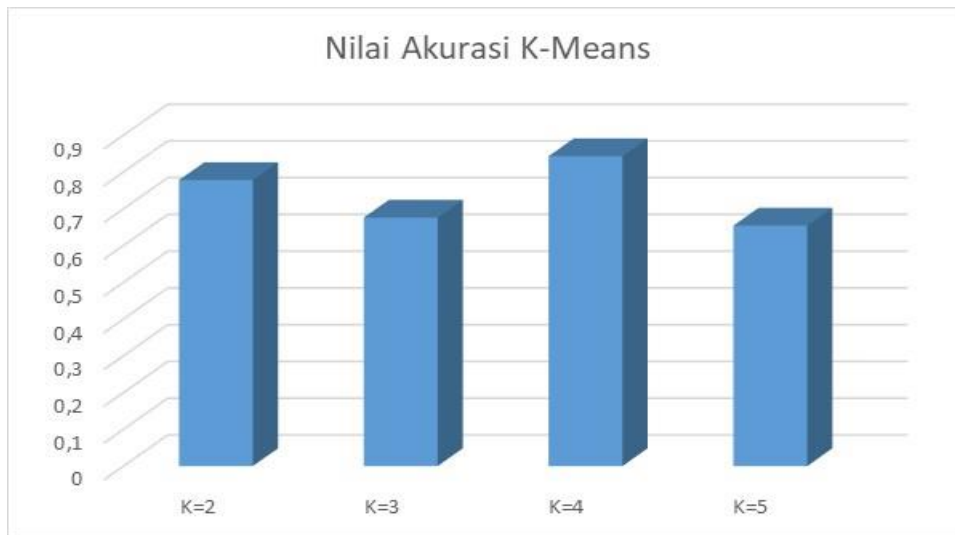
Setelah dilakukan proses pengumpulan data, maka selanjutnya proses yang dilakukan adalah proses pengujian atau pengolahan data. Proses pengujian pertama dilakukan dengan menggunakan algoritma K-Means. Proses yang dilakukan terhadap algoritma K-Means dengan menggunakan tools pengujian rapid miner. Proses pengujian pada

rapid K-Means dengan membentuk terhadap beberapa cluster K=2, K=3, K=4 dan K=5. Dan hasil pengujian dapat dilihat pada tabel 2 berikut:

Tabel 2. Hasil Pengujian Algoritma K-Means

No	Jumlah Klaster	Nilai Akurasi
1	K=2	0,778
2	K=3	0,676
3	K=4	0,843
4	K=5	0,654

Gambar 2 merupakan hasil pengujian K-Means dalam bentuk grafik.



Gambar 2. Hasil Pengujian K-Means

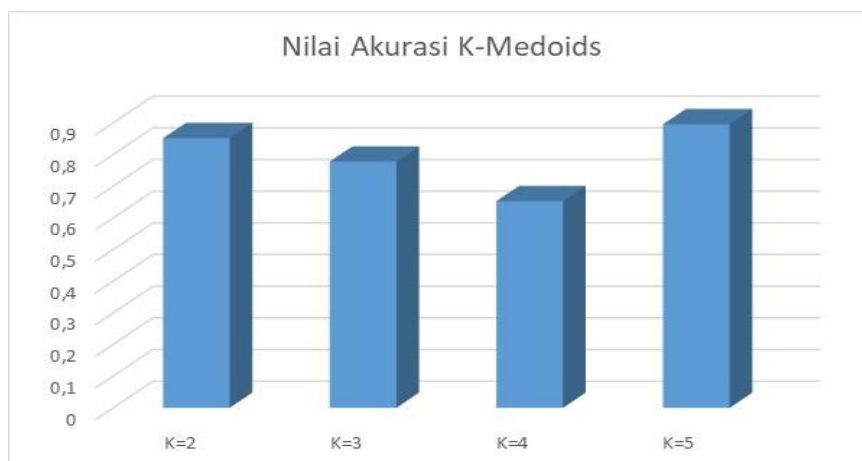
3.3 Pengujian Algoritma K-Medoids

Setelah dilakukan pengujian terhadap algoritma K-Means, maka selanjutnya proses pengujian dilakukan dengan menggunakan K-Medoids dengan terdapat beberapa cluster K=2, K=3, K=4 dan K=5. Dan hasil pengujian dapat dilihat pada tabel 3 berikut:

Tabel 3. Hasil Pengujian Algoritma K- Medoids

No	Jumlah Klaster	Nilai Akurasi
1	K=2	0,852
2	K=3	0,778
3	K=4	0,653
4	K=5	0,896

Gambar 3 merupakan hasil pengujian K-Medoids dalam bentuk grafik.



Gambar 3. Hasil Pengujian K-Medoids

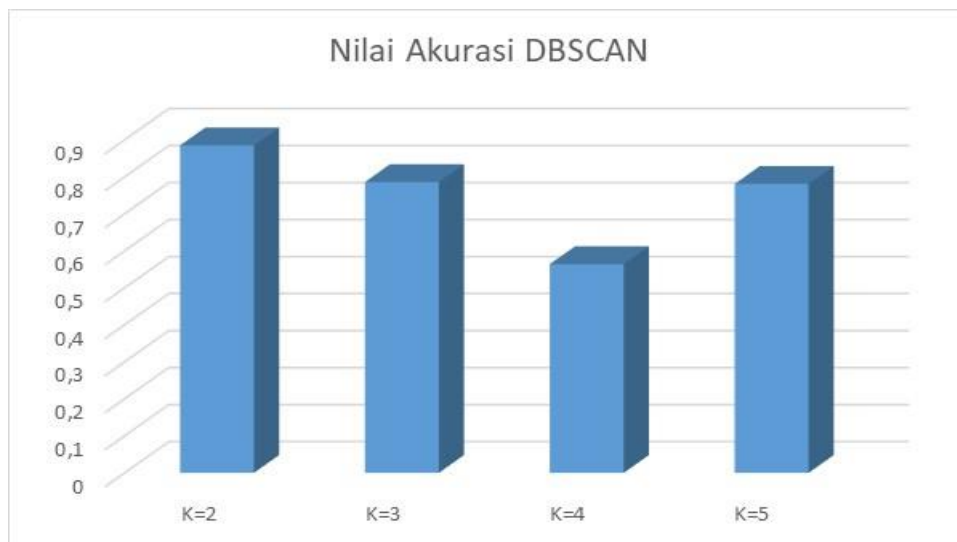
3.4 Pengujian Algoritma DBSCAN

Pengujian terakhir yang dilakukan adalah dengan menggunakan algoritma DBSCAN. Pada proses pengujian dengan algoritma DBSCAN juga terdapat beberapa cluster yang terbentuk K=2, K=3, K=4 dan K=5. Dan hasil pengujian dapat dilihat pada tabel 4 berikut:

Tabel 4. Hasil Pengujian Algoritma DBSCAN

No	Jumlah Klaster	Nilai Akurasi
1	K=2	0,885
2	K=3	0,786
3	K=4	0,564
4	K=5	0,781

Gambar 4 merupakan hasil pengujian DBScan dalam bentuk grafik.



Gambar 4. Hasil Pengujian DBSCAN

3.5 Pembahasan

Pembahasan merupakan proses pengambilan keputusan yang dilakukan dari pengujian, dimana setelah selesai proses pengujian dengan algoritma K-Means, K-Medoids dan DBSCAN. Dalam memudahkan membaca terhadap hasil pengujian dapat dilihat pada tabel berikut

Tabel 5. Hasil Pengujian

No	Jumlah Cluster	Algoritma K-Means	Algoritma K-Medoids	Algoritma DBSCAN
1	K=2	0,778	0,852	0,885
2	K=3	0,676	0,778	0,786
3	K=4	0,843	0,653	0,564
4	K=5	0,654	0,896	0,781

Gambar 5 merupakan hasil perbandingan dari 3 metode yang digunakan.



Gambar 4. Hasil Pengujian Keseluruhan

Dari gambar 4. Dapat dilihat bahwasannya hasil dari pengujian bahwasannya untuk algoritma K-Means nilai akurasi tertinggi didapatkan pada K=4 dengan nilai 0,843, untuk algoritma K-Medoids nilai tertinggi didapatkan pada K=5 dengan nilai 0,896 dan untuk algoritma DBSCAN nilai tertinggi didapatkan pada K=2 dengan nilai 0,885

4. KESIMPULAN

Tahapan terakhir setelah dilakukan proses penelitian yaitu pengambilan keputusan. Pengambilan keputusan dapat disusun dalam bentuk kesimpulan adapun hasil – hasil yang didapatkan pada penelitian bahwasannya peran data mining sudah sangat baik untuk membantu penyelesaian permasalahan terhadap kualitas udara, dimana data mining dapat membantu melakukan proses pengolahan data dan mendapatkan informasi baru untuk pengambilan keputusan. Dalam proses yang dilakukan pada penyelesaian data mining digunakan teknik clustering dengan menggunakan 3 (tiga) algoritma yaitu K-Means, K-Medoids dan DBSCAN. Adapun hasil yang didapatkan bahwa algoritma K-Means nilai akurasi tertinggi didapatkan pada K=4 dengan nilai 0,843, untuk algoritma K-Medoids nilai tertinggi didapatkan pada K=5 dengan nilai 0,896 dan untuk algoritma DBSCAN nilai tertinggi didapatkan pada K=2 dengan nilai 0,885.

REFERENCES

- [1] A. A. H. Kirono, I. Asror, and Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara Dki Jakarta Menggunakan Algoritma Naïve Bayes," *e-Proceeding Eng.*, vol. 9, no. 3, p. 1962, 2022.
- [2] I. I. Ridho and G. Mahalisa, "Analisis Klasifikasi Dataset Indeks Standar Pencemaran Udara (Ispu) Di Masa Pandemi Menggunakan Algoritma Support Vector Machine (Svm)," *Technol. J. Ilm.*, vol. 14, no. 1, p. 38, 2023, doi: 10.31602/tji.v14i1.8005.
- [3] S. S. A. Umri *et al.*, "Analysis and Comparison of Classification Algorithm in Air," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 2, pp. 98–104, 2021, doi: 10.33387/jiko.
- [4] A. Nugroho, I. Asror, and Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *eProceedings Eng.*, vol. 10, No. 2, no. 2, pp. 1824–1834, 2023, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20030%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20030/19395>.
- [5] M. Astriyani, I. N. Laela, D. P. Lestari, L. Anggraeni, and T. Astuti, "Analisis Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma C.45," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 6, no. 1, pp. 36–41, 2023, doi: 10.32524/jusitik.v6i1.790.
- [6] A. Khumaidi, R. Raafi, I. Permana Solihin, and J. Rs Fatmawati, "Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung," *J. Telemat.*, vol. 15, no. 1, pp. 13–18, 2020.
- [7] 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.
- [8] A. D. Wiranata, S. Soleman, I. Irwansyah, I. K. Sudaryana, and R. Rizal, "Klasifikasi Data Mining Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta Menggunakan Algoritma K-Nearest Neighbors (K-Nn)," *Infotech J. Technol. Inf.*, vol. 9, no. 1, pp. 95–100, 2023, doi: 10.37365/jti.v9i1.164.
- [9] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [10] Y. A. Suwitono and F. J. Kaunang, "Implementasi Algoritma Convolutional Neural Network (CNN) Untuk Klasifikasi Daun Dengan Metode Data Mining SEMMA Menggunakan Keras," *J. Komtika (Komputasi dan Inform.*, vol. 6, no. 2, pp. 109–121, 2022, doi: 10.31603/komtika.v6i2.8054.
- [11] N. R. S. Purba and F. Riandari, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Analisis Keranjang Belanja Pada Transaksi Penjualan Pada PT Madu Kembang Joyo," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 4, no. 1, pp. 69–74, 2021, doi: 10.32672/jnkti.v4i1.2745.
- [12] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.
- [13] A. Wibowo and R. Sasongko, "PENERAPAN DATA MINING PADA SUKU BUNGA INVESTASI DEPOSITO DI INDONESIA MENGGUNAKAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PROFITABILITAS (Application of Data Mining on Deposit Investment Rate in Indonesia Using K-Means Clustering Method for Profitab," *Ascarya*, vol. 2, no. 1, 2022, [Online]. Available: <https://doi.org/10.53754/iscs.v2i1.369>.
- [14] Dewi Eka Putri and Eka Praja Wiyata Mandala, "Hybrid Data Mining berdasarkan Klasterisasi Produk untuk Klasifikasi Penjualan," *J. KomtekInfo*, vol. 9, pp. 68–73, 2022, doi: 10.35134/komtekinfo.v9i2.279.
- [15] A. S. Imananda and H. D. Ranglalin, "Klasterisasi Negara Pengekspor Beras ke Indonesia Menggunakan Algoritma K-Means Clustering," *J. SISFOTENIKA*, vol. 13, no. 2, pp. 139–150, 2023, [Online]. Available: <http://sisfotenika.stmikpontianak.ac.id/index.php/ST>.
- [16] A. P. Adistyana, N. Lutfiyani, P. Tara, Rifaldi, R. Adriyan, and P. Rosyani, "Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa," *OKTAL J. Ilmu Komput. dan Sci.*, vol. 2, no. 8, pp. 2301–2306, 2023.
- [17] Y. Diana and F. Hadi, "Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang," *J. Inf. Syst. Informatics Eng. Vol.*, vol. 7, no. 1, pp. 97–103, 2023.
- [18] E. Prasetyaningrum and P. Susanti, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan," *J. Media Inform. Budidarma*, vol. 7, no. 4, pp. 1775–1783, 2023, doi: 10.30865/mib.v7i4.6477.

- [19] I. R. C. Pata, Stasiswaty, and N. Ransi, “SEGMENTASI PEMETAAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA DBSCAN DENGAN RFM MODEL BERBASIS WEB,” *ANIMATOR*, vol. 1, no. 2, pp. 63–71, 2023, [Online]. Available: <http://animator.uho.ac.id/index.php/journal/article/view/25%0Ahttp://animator.uho.ac.id/index.php/journal/article/download/25/16>.
- [20] A. Syafrianto and E. Riswanto, “Pengelompokan Jumlah Kunjungan Mahasiswa ke Perpustakaan Kampus Menggunakan Algoritma DBSCAN,” *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 75–81, 2023, doi: 10.33379/gtech.v7i1.1925.
- [21] S. Widaningsih and S. Yusuf, “Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2598–2611, 2022, doi: 10.35957/jatisi.v9i3.859.
- [22] R. R. Andarista and A. Jananto, “Penerapan Data Mining Algoritma C4 . 5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor,” vol. 16, no. 2, pp. 29–43.
- [23] M. A. Wiratama and W. M. Pradnya, “Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes,” *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 1, p. 1, 2022, doi: 10.23887/janapati.v11i1.45282.
- [24] I. Maryani, O. Revianti, H. M. Nur, and S. Sunanto, “Implementasi Data Mining Pada Penjualan Di Toko GOC Kosmetik Dengan Menggunakan Metode Algoritma Apriori,” *Indones. J. Softw. Eng.*, vol. 8, no. 1, pp. 92–98, 2022, doi: 10.31294/ijse.v8i1.13017.
- [25] N. N. Hasanah and A. S. Purnomo, “Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta),” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 4, no. 2, pp. 300–311, 2022, doi: 10.47233/jteksis.v4i2.499.
- [26] V. Ramadhan and A. Voutama, “Clustering Menggunakan Algoritma K-Means Pada Penyakit ISPA di Puskesmas Kabupaten Karawang,” *J. Pendidik. dan Konseling*, vol. 4, pp. 462–473, 2022.
- [27] Reza Gustrianda and D. I. Mulyana, “Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids,” *J. Media Inform. Budidarma*, vol. 6, no. 1, pp. 27–34, 2022, doi: 10.30865/mib.v6i1.3294.
- [28] N. L. P. P. Dewi, I. N. Purnama, and N. W. Utami, “Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara),” *J. Ilm. Teknol. Inf. Asia*, vol. 16, no. 2, p. 105, 2022, doi: 10.32815/jitika.v16i2.761.
- [29] Luth Fimawahib, Imam Rangga Bakti, and Asep Supriyanto, “Algoritma K-Medoids untuk Pengelompokan Produksi Padi dan Beras sebagai Upaya Optimalisasi Ketahanan Pangan di Provinsi Riau,” *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 13–24, 2022, doi: 10.33372/stn.v8i2.877.
- [30] N. Arminarahmah, A. G. Daengs, J. Tata Hardinata, and I. Kalimantan Muhammad Arsyad Al Banjari, “Klusterisasi Impor Beras Di Indonesia Menurut Negara Asal Utama Menggunakan Algoritma K-Medoids,” *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 8, no. 2, pp. 793–801, 2023, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>.
- [31] R. N. H. Hutasuhut, H. Okprana, and B. E. Damanik, “Penerapan Data Mining Untuk Menentukan Penerima Program Bidikmisi Menggunakan Algoritma K-Medoids,” *TIN Terap. Inform. Nusantara*, vol. 2, no. 11, pp. 667–672, 2022, doi: 10.47065/tin.v2i11.1516.
- [32] T. Ramayanti, E. Haerani, J. Jasril, and L. Oktavia, “Penerapan Algoritma K-Medoids Pada Clustering Penerima Bantuan Pangan Non Tunai (BPNT),” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1287–1296, 2023, doi: 10.30865/mib.v7i3.6475.
- [33] R. Anggara and A. Rahman, “Implementasi Algoritma DBSCAN Dalam Mengelompokan Data Pasien Terdiagnosa Penyakit Ginjal Kronis (PGK),” *J. Algoritm.*, vol. 3, no. 1, pp. 114–123, 2022, doi: 10.35957/algoritme.v3i1.3593.
- [34] D. P. Indini, S. R. Siburian, Nurhasanah, and D. P. Utomo, “Implementasi Algoritma DBSCAN untuk Clustering Seleksi Penentuan Mahasiswa yang Berhak Menerima Beasiswa Yayasan,” in *Prosiding Seminar Nasional Sosial, Humaniora, dan Teknologi*, 2022, pp. 325–331.
- [35] I. W. Septiani, A. C. Fauzan, and M. M. Huda, “Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru,” *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 556, 2022, doi: 10.30865/json.v3i4.4055.
- [36] N. A. Putri, T. W. Utami, and R. Wasono, “Spatial Temporal Density-Based Spatial Clustering Applications With Noise (ST-DBSCAN) Untuk Pengelompokan Titik Kebakaran Hutan dan Lahan di Provinsi Riau Pada Tahun 2021,” pp. 436–445, 2021.