

ANALISIS KUALITAS UDARA MENGGUNAKAN ALGORITMA K-MEANS

Karina Auliasari¹, Mariza Kertaningtyas²

¹Program Studi Teknik Informatika, ²Program Studi Teknik Industri, Institut Teknologi Nasional Malang

Jln. Raya Karanglo Km. 2 Malang 65145

¹karina.auliasari@lecturer.itn.ac.id, ²marizakertaningtyas@lecturer.itn.ac.id

Abstract

This research uses the K-Means method in analyzing data from gas sensors, the main purpose was to identify indoor and outdoor air quality. The sensor system has applied the IoT concept, which is design an Arduino microcontroller device that equipped with an MQ-135 sensor and an MQ-7 sensor. The sensor system has duty to collecting the pollutant data and then it can be sent to the gateway. The gateway used a web service program to provide polutan data transmission and it save in MySQL. The dataset is obtained from from monitoring results that taken in eight months. The data it self contain six variables that is NH3, CO, CO2, moth, day and time. K-Means method that applied in dataset produce 2 clusters. From the experimental plot, it can be seen that K-Means algorithm has successfully clustered six variables. By observing the pattern, it can be seen that CO and day is the only good combination that produce well-grouped data patterns without being mixed with each other. Based on comparison between clustering results by K-Means algorithm and actual data from monitoring shown that the accuracy, sensitivity, specificity, precision and recall values are 77.78%.

Keywords : *Clustering, Data, Sensor, Gas, K-Means, IoT*

Abstrak

Penelitian ini menggunakan metode K-Means dalam menganalisis data dari sensor gas untuk mengidentifikasi udara di dalam dan luar ruangan. Perangkat keras sistem sensor pada penelitian ini mengaplikasikan konsep IoT, yang dimulai dengan merancang perangkat mikrokontroler arduino yang dilengkapi dengan sensor MQ-135 dan sensor MQ-7. Sistem sensor bertanggung jawab mengumpulkan data polutan untuk selanjutnya dikirimkan kepada gateway. Gateway menggunakan layanan web service untuk menyediakan transmisi dan penyimpanan data di MySQL. Dataset yang digunakan merupakan data hasil monitoring yang diambil dalam rentang waktu delapan bulan dengan enam variabel data yaitu NH3, CO2, CO, bulan, hari dan waktu. Dari hasil klastering yang dilakukan terhadap dataset terbentuk 2 klaster. Dari hasil pengelompokan data yang dilakukan oleh algoritma K-Means terlihat bahwa hanya variabel hari dan co yang memiliki visualisasi pola data terkelompok dengan baik. Berdasarkan pengujian hasil klasterisasi yang dilakukan oleh algoritma K-Means dengan data aktual indoor dan outdoor dari hasil monitoring sensor gas menghasilkan nilai accuracy, sensitivity, specificity, precision dan recall sebesar 77,78 %.

Kata kunci : *Klastering, Data, Sensor, Gas, K-Means, Polutan.*

1. PENDAHULUAN

Lingkungan seperti rumah, tempat perbelanjaan, kantor, pabrik dan laboratorium

menjadi tempat manusia menghabiskan sebagian waktunya dalam beraktifitas[1]. Oleh karena itu, kualitas udara pada lingkungan dalam ruangan memiliki peran penting dalam kehidupan

manusia. Banyak penelitian lebih berfokus untuk membahas kualitas udara di luar ruangan untuk membahas seberapa berbahayanya polutan dari asap kendaraan, asap pabrik dan asap pembakaran lahan terbuka [2]. Namun yang sering dilupakan bahwa polutan di dalam ruangan juga memiliki potensi membahayakan kesehatan, seperti kemungkinan kebocoran tabung gas LPG di dalam rumah maupun di pabrik, senyawa organik yang menguap dari kebocoran AC di kantor, tempat perbelanjaan atau dari bahan kimia pada laboratorium. Beberapa pabrik dan laboratorium mengantisipasi adanya polutan berbahaya dengan memasang alat pemantau gas berbahaya yang memiliki harga yang mahal, konsumsi listrik yang besar dan bising dalam pengoperasiannya. Alat ini pun kinerjanya terbatas tidak bisa digunakan secara bersamaan di beberapa ruangan.

Lebih lanjut alat monitoring gas berkembang dengan adanya konsep wireless sensor network (WSN) dimana beberapa alat yang dilengkapi sensor ditempatkan pada ruangan yang berbeda dapat terhubung melalui jaringan wireless untuk mengirim data sensor secara real time. Kemudian juga konsep WSN diterapkan secara luas dalam beberapa wilayah sehingga muncul konsep *Internet of Things* (IoT). Terlepas dari kemajuan teknologi alat pemantau gas menggunakan konsep WSN ataupun IoT saat ini, perlu dilakukan ekstraksi informasi dari data sensor yang telah terkumpul. Cara kerja *Internet of Things* sendiri adalah dengan cara interaksi antara sesama mesin yang terhubung secara otomatis tanpa campur tangan user dan dalam jarak berapapun. Internet menjadi penghubung di antara kedua interaksi mesin tersebut, sementara user hanya bertugas sebagai pengatur dan pengawas bekerjanya alat tersebut secara langsung [3]. Data yang dihimpun oleh alat pemantau gas perlu dianalisis lebih lanjut untuk bisa mengidentifikasi sumber polusi dan tren polutan yang ada di dalam ruangan. Selain itu hasil dari analisis dapat dikomunikasikan kepada penghuni ruangan tentang hubungan aktifitas sehari-hari dengan polusi yang terjadi dan untuk memperingatkan penghuni ruangan apabila ada potensi gas berbahaya di dalam ruangan.

2. TINJAUAN PUSTAKA DAN TEORI

2.1. Penelitian terkait

Beberapa penelitian yang melakukan analisis data polutan udara diantaranya seperti penelitian yang dilakukan oleh Sugiarto et al. menggunakan

algoritma decision tree C.4.5 untuk menghasilkan pemodelan keputusan status udara baik (good) atau buruk (warning) [4]. Selanjutnya Arya et al. melakukan analisis data polutan menggunakan metode K-Nearest Neighbour (K-NN) untuk mengelompokkan data ke dalam tiga status yaitu baik, sedang dan tidak sehat [5]. Pada penelitian yang dilakukan oleh Aini et al menggunakan algoritma naïve bayes untuk mengklasifikasikan data polutan udara di Kota Makasar dimana pada penelitian ini jumlah data training yang digunakan mempengaruhi tingkat akurasi hasil klasifikasi ke dalam tiga label yaitu baik, sedang dan tidak sehat [6]. Selanjutnya penelitian yang dilakukan Handayani et al. mengklasifikasikan kualitas udara menggunakan metode *support vector machine* (SVM) menggunakan tiga buah node sensor, dimana metode SVM membedakan data ke dalam dua kelas yaitu kelas *normal* dan *unnormal* dari 3 lokasi area parkir yang diteliti [7]. Sejauh ini penelitian-penelitian yang ada dilakukan untuk menganalisis kualitas udara di luar ruangan, ada penelitian yang menganalisis kualitas udara di dalam ruangan seperti yang dilakukan Shidki et al. di tahun 2020 namun belum menerapkan suatu metode atau algoritma untuk mengklasifikasikan data polutan [8].

Oleh karena itu pada penelitian ini digunakan metode K-Means dalam melakukan pengelompokan data dari sensor gas untuk mengidentifikasi udara di dalam dan luar ruangan. Metode K-Means dipilih karena metode K-Means memiliki kelebihan kemudahan implementasi metode dan mampu meminimalkan variasi antar data yang ada di dalam suatu *cluster* serta memaksimalkan variasi data yang ada di *cluster* lainnya [9].

2.2. Kualitas udara di dalam ruangan

Environmental protection agency (EPA) dari USA mendefinisikan kualitas udara di dalam ruangan (indoor air quality) sebagai hasil dari interaksi antara tempat, suhu, sistem gedung (desain konstruksi, interior dan mekanisasi ruangan), sumber polutan (material, peralatan gedung, proses kelembapan, aktifitas di dalam maupun luar gedung) dan manusia di dalamnya [10]. Berdasarkan standar ASHRAE 62-2001 kualitas udara yang baik adalah udara di dalam ruangan yang tidak terkontaminasi oleh polutan, dimana konsentrasi polutan yang membahayakan di dalam ruangan sebesar 80% atau lebih sehingga manusia di dalam ruangan merasa tidak nyaman [11].

Hasil riset dari *the national institute of occupational safety and health* (NIOSH),

menyebutkan ada 5 sumber pencemaran di dalam ruangan yaitu [12]:

a) Pencemaran dari alat-alat di dalam gedung seperti asap rokok, pestisida, bahan-bahan pembersih ruangan.

b) Pencemaran di luar gedung meliputi masuknya gas buangan kendaraan bermotor, gas dari cerobong asap atau dapur yang terletak di dekat gedung, dimana kesemuanya dapat terjadi akibat penempatan lokasi lubang udara yang tidak tepat.

c) Pencemaran akibat bahan bangunan meliputi pencemaran formaldehid, lem, asbes, fibreglass dan bahan-bahan lain yang merupakan komponen pembentuk gedung tersebut.

d) Pencemaran akibat mikroba dapat berupa bakteri, jamur, protozoa dan produk mikroba lainnya yang dapat ditemukan di saluran udara dan alat pendingin beserta seluruh sistemnya.

e) Gangguan ventilasi udara berupa kurangnya udara segar yang masuk, serta buruknya distribusi udara dan kurangnya perawatan sistem ventilasi udara.

2.3. Algoritma K-Means

Algoritma *K-Means* diperkenalkan oleh J.B. MacQueen pada tahun 1976. Metode ini mempartisi data ke dalam *cluster* (kelompok) sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam *cluster* yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam kelompok yang lain [13]. Berikut adalah langkah-langkah dari algoritma *K-Means* [13]:

Langkah 1 : Menentukan banyak *K-cluster* yang ingin dibentuk.

Langkah 2 : Membangkitkan nilai random untuk pusat *cluster* awal (*centroid*) sebanyak *k*.

Langkah 3 : Menghitung jarak setiap data input terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidian* (*Euclidian Distance*) hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Berikut adalah persamaan *Euclidian Distance* (persamaan 1):

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \quad (1)$$

Langkah 4 : Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).

Langkah 5 : Mengupdate nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus yang ditunjukkan pada persamaan 2:

$$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum_{j=S_j} x_j \quad (2)$$

dimana:

$\mu_j(t+1)$ = centroid baru pada iterasi ke $(t+1)$,

N_{S_j} = banyak data pada cluster S_j

Langkah 6 : Melakukan perulangan dari langkah 2 hingga 5 hingga anggota tiap *cluster* tidak ada yang berubah.

Langkah 7 : Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat *cluster* (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk *Radial Basis Function* yang ada di *hidden layer*.

3. METODOLOGI PENELITIAN

3.1. Tahapan penelitian

Beberapa proses dilakukan di dalam penelitian ini diantaranya perancangan dan pembuatan sistem sensor, pengujian sistem sensor pada laboratorium, pemasangan sistem sensor, proses monitoring data dan pengolahan data. Pada detail proses pengolahan data dilakukan beberapa tahapan yaitu membaca dataset, menghitung jarak dari masing-masing data ke data lainnya, menentukan jumlah kelompok klaster yang optimal, membuat klasterisasi dari data sesuai jumlah kelompok yang telah ditentukan dan membandingkan hasil klaster. Gambaran blok diagram tahapan penelitian dari tiap proses ditunjukkan pada Gambar 1.

3.2. Perancangan arsitektur sistem dan perangkat

Perancangan perangkat keras sistem sensor pada penelitian ini mengaplikasikan konsep IoT, yang dimulai dengan merancang perangkat mikrokontroler arduino yang dilengkapi dengan sensor MQ-135 dan sensor MQ-7, perangkat ini diharapkan menjadi solusi alat otomatis pemantau kualitas udara di dalam ruangan berbiaya rendah. Sensor MQ-135 digunakan untuk melakukan monitoring nilai kadar polutan CO2 dan NH3 (gas ammonia) sedangkan sensor MQ-7 digunakan untuk memantau nilai kadar polutan CO. Selanjutnya masing-masing sistem sensor ditempatkan pada lokasi di dalam ruangan (*indoor*) dan di luar ruangan (*outdoor*).

Skenario penempatan sistem sensor untuk menggambarkan bahwa pihak pengawas gedung dapat memonitoring kondisi udara baik di dalam maupun luar gedung. Sistem sensor bertanggung

jawab mengumpulkan data polutan untuk selanjutnya dikirimkan kepada *gateway*. *Gateway* menggunakan layanan *web service* untuk menyediakan transmisi dan penyimpanan data di MySQL. *Gateway* sendiri merupakan layanan *web service* yang dibuat menggunakan Bahasa pemrograman PHP.

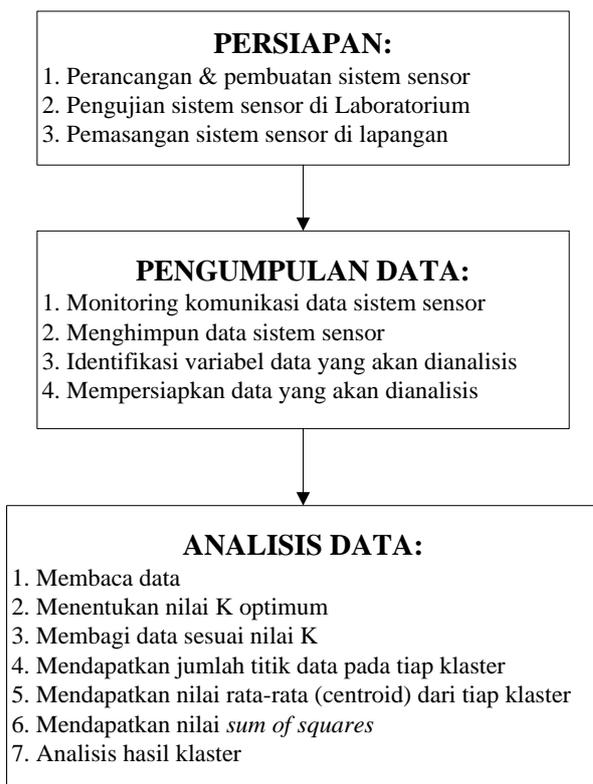
Perangkat yang digunakan pada penelitian ini menggabungkan mikrokontroler Arduino UNO sebagai unit pengolah dan chip ESP8266 sebagai modul komunikasi data. Bagian dari perangkat ditunjukkan pada Gambar 2. Data yang dipantau oleh sensor diunggah ke API platform IoT ThingSpeak yang bersifat open source menggunakan protocol HTTP. Chip ESP8266 terhubung ke internet menggunakan Wi-fi dan bertanggung jawab mengunggah data yang diterima platform *ThingSpeak*. Pengguna dapat memonitoring data polutan dari halaman website yang sudah disediakan oleh *ThingSpeak*. Data nilai kadar polutan hasil monitoring disajikan dalam bentuk numerik yang juga disimpan dalam database MySQL. Data nilai kadar polutan dalam bentuk numerik inilah yang selanjutnya diolah untuk dianalisis menggunakan metode K-Means.



Gambar 2. Bagian perangkat sistem sensor gas

3.3. Prapemrosesan data

Dataset yang digunakan dalam penelitian ini merupakan data hasil monitoring yang diambil dalam rentang waktu dari bulan Agustus 2020 hingga Maret 2021 dengan enam variabel data yaitu NH3, CO, CO2, waktu, bulan dan hari, yang dikategorikan berdasarkan lokasi. Data uji yang digunakan dalam penelitian ini berjumlah 200 baris data dan data uji berjumlah 153 baris data. Pada variabel waktu data sensor gas diambil pada 4 waktu yaitu pagi (pukul 08:00-10:00 WIB), siang (11:00-14:00 WIB), sore (15:00-17:00 WIB) dan malam (18:00-21:00 WIB). Proses awal pengolahan data hasil monitoring dilakukan dengan membaca dataset seperti yang ditunjukkan pada Gambar 3.



Gambar 1. Tahapan penelitian

	nh3	co2	co	waktu	bulan	hari
1	0.00	0.35	157	Malam	Agustus	1
2	0.00	0.35	157	Malam	Agustus	2
3	0.00	0.35	156	Malam	Agustus	3
4	0.00	0.35	156	Malam	Agustus	4
5	0.00	0.35	157	Malam	Agustus	5
6	0.00	0.47	162	Malam	Agustus	6
7	0.00	0.35	158	Pagi	Agustus	7
8	0.00	0.34	160	Pagi	Agustus	8
9	0.00	0.35	160	Pagi	Agustus	9
10	NA	0.35	160	Pagi	Agustus	10
11	0.00	0.35	158	Pagi	Agustus	11
12	0.00	0.35	157	Pagi	Agustus	12
13	0.00	NA	158	Pagi	Agustus	13
14	0.00	0.47	160	Pagi	Agustus	14
15	0.00	0.35	158	Pagi	Agustus	15

Gambar 3. Dataset monitoring sensor gas

Jika diamati isi data dari dua variabel yaitu waktu dan bulan memiliki format teks sehingga perlu diubah terlebih dahulu ke dalam format numerik, yang pada penelitian ini diubah menggunakan fungsi `data.matrix` pada bahasa pemrograman R. Pada variabel waktu, data yang berisi pagi diubah menjadi angka 1, siang menjadi angka 2, sore menjadi angka 3 dan malam menjadi

angka 4, hal serupa juga dilakukan pada variabel bulan dimana, misalnya jika isi data bulan Agustus diubah menjadi angka 8, September menjadi angka 9 dan seterusnya hingga bulan Desember menjadi angka 12.

Setelah semua isi baris data bernilai numerik langkah selanjutnya yang dilakukan mencari variabel data yang tidak berisi nilai atau *not available* (NA). Pada hasil pencarian didapatkan bahwa hanya variabel nh3 dan co2 yang memiliki nilai NA hal ini ditunjukkan dengan nilai TRUE seperti pada tampilan Gambar 5. Variabel nh3 dan co2 yang tidak berisi nilai selanjutnya dilakukan normalisasi data sehingga tidak mempengaruhi pada saat diterapkan algoritma K-Means. Normalisasi dilakukan dengan cara menghitung terlebih dahulu masing-masing nilai rata-rata variabel nh3 dan co2. Nilai rata-rata variabel nh3 akan menggantikan isi data pada variabel nh3 yang bernilai NA. Hal yang sama juga dilakukan pada variabel co2, data yang bernilai NA akan digantikan oleh nilai rata-rata dari variabel co2, sehingga dari proses normalisasi ini semua baris data memiliki nilai.

	nh3	co2	co	waktu	bulan	hari
1	0.00	0.35	157	4	8	1
2	0.00	0.35	157	4	8	2
3	0.00	0.35	156	4	8	3
4	0.00	0.35	156	4	8	4
5	0.00	0.35	157	4	8	5
6	0.00	0.47	162	4	8	6
7	0.00	0.35	159	1	8	7
8	0.00	0.34	160	1	8	8
9	0.00	0.35	160	1	8	9
10	0.00	0.35	160	1	8	10
11	0.00	0.35	158	1	8	11
12	0.00	0.35	157	1	8	12
13	0.00	NA	158	1	8	13
14	0.00	0.47	160	1	8	14
15	0.00	0.35	158	1	8	15

Gambar 4. Mengubah isi data ke bentuk numerik

```
nh3  co2  co waktu bulan hari kelas
TRUE TRUE FALSE FALSE FALSE FALSE
```

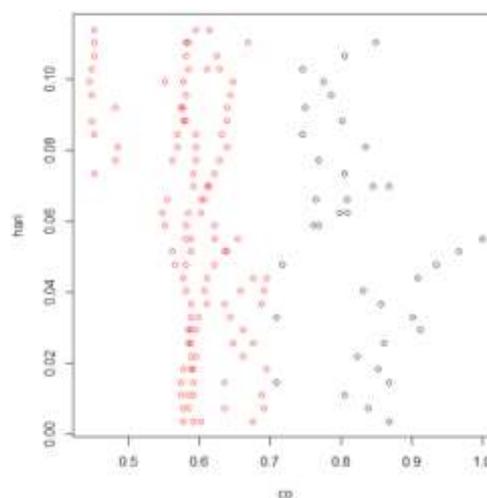
Gambar 5. Variabel yang memiliki nilai NA

4. HASIL DAN PEMBAHASAN

Untuk mengetahui pola pengelompokan data dari keenam variabel pada data sensor gas perlu dilakukan proses visualisasi pengelompokan data atau visualisasi klater dari masing-masing variabel, sehingga bisa dilihat kombinasi variabel mana yang memiliki pola data yang

terkelompokkan dengan baik tanpa saling tercampur.

Dari hasil visualisasi pengelompokan data menggunakan algoritma K-Means pada keenam variabel (nh3, co2, co, waktu, bulan dan hari) terlihat hanya ada satu kombinasi yang memiliki pola data yang terkelompokkan dengan baik yaitu variabel co dan hari seperti yang diperlihatkan grafik yang dilingkari merah pada Gambar 8. Detail gambaran pengelompokan data kombinasi dari variabel co dan hari diperlihatkan pada grafik Gambar 6.



Gambar 6. Grafik pengelompokan kombinasi variabel hari dan co

Hasil pengelompokan data yang dihasilkan oleh algoritma K-Means ditunjukkan pada Gambar 8 yang memperlihatkan tiga bagian. Nilai K ditentukan bernilai 2 karena disesuaikan dengan jumlah kelas lokasi yang dihasilkan data sensor yaitu indoor dan outdoor. Pada bagian **pertama** menunjukkan ukuran atau jumlah titik data pada dua klaster/kelompok. Untuk bagian **kedua** memperlihatkan nilai rata-rata (*centroid*) dari tiap klaster. Di bagian **ketiga** yaitu berupa *clustering vector* dimana menunjukkan *vector* berisi angka 1 sampai dengan 2 sesuai dengan jumlah klaster yang ditentukan di awal. Hasil *vector* menunjukkan dimulai dari angka 1 ini artinya data pertama dari dataset dialokasikan pada klaster nomor 1. Dari gambar juga terlihat isi *vector* pada baris kedua atau data ke52 bernilai 1 yang artinya data ke52 dari dataset kita dialokasikan pada klaster nomor 1, dan seterusnya hingga data ke154 teralokasi pada klaster nomor 1.

Untuk mengetahui seberapa akurat algoritma K-Means dalam melakukan pengelompokan data dengan data hasil monitoring sensor gas maka dilakukan pengukuran *accuracy*, *sensitivity*, *specificity*, *precision* dan *recall*. Pengukuran *accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Nilai akurasi diperoleh melalui persamaan 3, 4, 5, 6 dan 7 [13]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$= \frac{119 + 119}{119 + 119 + 34 + 34}$$

$$Accuracy = 0,7778$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$= \frac{119}{119 + 34}$$

$$Sensitivity = 0,7778$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$= \frac{119}{119 + 34}$$

$$Specificity = 0,7778$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$= \frac{119}{119 + 34}$$

$$Precision = 0,7778$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$= \frac{119}{119 + 34}$$

$$Recall = 0,7778$$

dimana:

TP: True positif, TN: True negatif,
 FP: False positif, FN: False negatif.

Nilai TP, TN, FP dan FN seperti yang ditunjukkan pada tabel 1 didapatkan melalui *confusion matrix* perbandingan hasil klasterisasi yang dilakukan oleh algoritma K-Means dengan data aktual indoor dan outdoor hasil monitoring yang ada pada lampiran. Dari hasil perhitungan

accuracy sesuai dengan persamaan 3, maka didapatkan nilai *accuracy*, *sensitivity*, *specificity*, *precision* dan *recall* sebesar 0,7778. Nilai *sensitivity*, *specificity*, *precision* dan *recall* dalam prosentase disajikan pada Tabel 2.

Tabel 1. nilai TP, TN, FP dan FN

N=153 data		Klaster hasil prediksi		Σ
		Klaster 1	Klaster 2	
Kelas data aktual sensor gas	Kelas indoor	TP=119	FN=34	153
	Kelas outdoor	FP=34	TN=119	153
Σ		153	153	

Tabel 2. Nilai *sensitivity*, *specificity*, *precision* dan *recall* dalam prosentase

	% (dalam persen)
<i>Accuracy</i>	77,78
<i>Sensitivity</i>	77,78
<i>Specificity</i>	77,78
<i>Precision</i>	77,78
<i>Recall</i>	77,78

5. Kesimpulan dan Saran

Dari hasil pengelompokan data yang dilakukan oleh algoritma K-Means terlihat bahwa hanya variabel hari dan co yang memiliki visualisasi pola data terkelompok dengan baik. Berdasarkan pengujian hasil klasterisasi yang dilakukan oleh algoritma K-Means dengan data aktual indoor dan outdoor dari hasil monitoring sensor gas menghasilkan nilai *accuracy*, *sensitivity*, *specificity*, *precision* dan *recall* sebesar 77,78 %.

Pada penelitian selanjutnya untuk dapat melihat variasi klasterisasi data hasil monitoring perlu diujicoba dengan menambahkan variabel data gas NO₂, O₃, kelembapan dan suhu baik di dalam maupun luar ruangan. Selain itu juga perlu dilakukan perbandingan klasterisasi data dengan metode KNN untuk bisa membandingkan akurasi hasil pengelompokan data sensor gas.

6. UCAPAN TERIMA KASIH

Ucapan terima kasih kami sampaikan kepada Institut Teknologi Nasional Malang yang telah memberikan dana pada penelitian ini yang dilaksanakan pada tahun 2020. Terima kasih juga kami sampaikan kepada program studi Teknik Informatika S1 ITN Malang dan Kepala Laboratorium Multimedia Teknik Informatika S1

ITN Malang yang telah mengakomodir selama pelaksanaan penelitian ini.

Daftar Pustaka:

- [1] I. Muhammad, "Manajemen Kualitas Udara dalam Gedung Bertingkat", Hiperkes. Jakarta, 2003.
- [2] K. Haryoto, "Kesehatan Lingkungan Pemukiman dan Perkantoran", Fakultas Kesehatan Masyarakat Universitas Indonesia, Depok : 2002.
- [3] D. Citra, I. Hadi dan Sarjana, "Platformweb Sebagai Penampil Data Monitoring Kotak Sampah Berbasis IoT," JIRE (Jurnal Informatika & Rekayasa Elektronika) Volume 3, No 2, Hal. 162-175, November 2020.
- [4] Sugiarto, Bambang dan R. Sustika, "Data Classification for Air Quality on Wireless Sensor Network Monitoring System Using Decision Tree Algorithm," 2nd International Conference on Science and Technology-Computer (ICST), hal. 172-176, 2016.
- [5] T. F. Arya, M. Faiqurahman, dan Y. Azhar, "Aplikasi Wireless Sensor Network Untuk Sistem Monitoring Dan Klasifikasi Kualitas Udara," Jurnal Sistem Informasi Vol. 14 Issue Oktober 2018 ISSN 2088-7043, hal. 74-82, 2018.
- [6] N. Aini, R. Ruktiari, M. R. Pratama dan A. F. Buana, "Sistem Prediksi Tingkat Pencemaran Polusi Udara dengan Algoritma Naïve Bayes di Kota Makassar," Prosiding Seminar Nasional Komunikasi dan Informatika ke-3, Hal. 83-90, 2019.
- [7] A. S. Handayani, S. Soim, T. E. Agusdi, Rumiasih dan A. Nurdin, "Klasifikasi Kualitas Udara Dengan Metode Support Vector Machine," JIRE (Jurnal Informatika & Rekayasa Elektronika) Volume 3, No 2, November 2020.
- [8] H. Shidki, I. Chandra dan E. Djunaedy, "Analisis Kualitas Udara Dalam Ruangan Pada Kantor Terbuka Di Universitas Telkom," e-Proceeding of Engineering Vol. 7 No. 1 ISSN: 2355-9365, Hal. 1121-1128, 2020.
- [9] Ediyanto, M. N. Mara dan N. Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis", BIMASTER(Buletin Ilmiah Matematika Statistika dan Terapannya), Vol. 2, No. 2, 2013.
- [10] EPA, "An Office Building Occupational's Guide to Indoor Air Quality,"[Daring], Tersedia pada: <https://www.epa.gov/indoor-air-quality-iaq/office-building-occupants-guide-indoor-air-quality>, [Diakses: 1-Aug-2020].
- [11] American Society for Heating, Refrigerating and Air Conditioning Engineers (ASHRAE), ASHRAE Standard 62.1-2007, "Standard for Acceptable Indoor Air Quality", Atlanta[Daring], Tersedia pada: <https://aws.state.ak.us/OnlinePublicNotices/Notices/Attachment.aspx?id=113385>, [Diakses: 10-Aug-2020].
- [12] NIOSH, "Indoor Air Quality and Work Environment Symptoms," Survey NIOSH, 2001.
- [13] P. Eko, "Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB," Penerbit ANDI OFFSET, Yogyakarta, 2014.

Lampiran

Tabel nilai *confussion matrix* perbandingan hasil klusterisasi yang dilakukan oleh algoritma K-Means dengan data aktual indoor dan outdoor

Data ke-	Data aktual	Prediksi (Klaster)	TP	TN	FP	FN	Σ
1.	Indoor	1	1	1	0	0	2
2.	Indoor	1	1	1	0	0	2
3.	Indoor	1	1	1	0	0	2
4.	Indoor	1	1	1	0	0	2
5.	Indoor	1	1	1	0	0	2
6.	Indoor	1	1	1	0	0	2
7.	Indoor	1	1	1	0	0	2
8.	Indoor	1	1	1	0	0	2
9.	Indoor	1	1	1	0	0	2
10.	Indoor	1	1	1	0	0	2
11.	Indoor	1	1	1	0	0	2
12.	Indoor	1	1	1	0	0	2
13.	Indoor	1	1	1	0	0	2
14.	Indoor	1	1	1	0	0	2
15.	Indoor	1	1	1	0	0	2
16.	Indoor	1	1	1	0	0	2
17.	Indoor	1	1	1	0	0	2
18.	Indoor	1	1	1	0	0	2
19.	Indoor	1	1	1	0	0	2
20.	Indoor	1	1	1	0	0	2
21.	Indoor	1	1	1	0	0	2
22.	Indoor	1	1	1	0	0	2
23.	Indoor	1	1	1	0	0	2
24.	Indoor	1	1	1	0	0	2
25.	Indoor	1	1	1	0	0	2
26.	Indoor	1	1	1	0	0	2
27.	Indoor	1	1	1	0	0	2
28.	Indoor	1	1	1	0	0	2
29.	Indoor	1	1	1	0	0	2
30.	Indoor	1	1	1	0	0	2
31.	Indoor	1	1	1	0	0	2
32.	Indoor	1	1	1	0	0	2
33.	Indoor	1	1	1	0	0	2
34.	Indoor	1	1	1	0	0	2
35.	Indoor	1	1	1	0	0	2
36.	Indoor	1	1	1	0	0	2
37.	Indoor	1	1	1	0	0	2
38.	Indoor	1	1	1	0	0	2
39.	Indoor	1	1	1	0	0	2
40.	Indoor	1	1	1	0	0	2
41.	Indoor	1	1	1	0	0	2
42.	Indoor	1	1	1	0	0	2
43.	Indoor	1	1	1	0	0	2
44.	Indoor	1	1	1	0	0	2
45.	Indoor	1	1	1	0	0	2
46.	Indoor	1	1	1	0	0	2
47.	Indoor	1	1	1	0	0	2
48.	Indoor	1	1	1	0	0	2
49.	Indoor	1	1	1	0	0	2
50.	Indoor	1	1	1	0	0	2
51.	Indoor	1	1	1	0	0	2
52.	Indoor	1	1	1	0	0	2
53.	Indoor	1	1	1	0	0	2
54.	Indoor	1	1	1	0	0	2
55.	Indoor	1	1	1	0	0	2
56.	Indoor	1	1	1	0	0	2
57.	Indoor	1	1	1	0	0	2
58.	Indoor	1	1	1	0	0	2

59.	Indoor	1	1	1	0	0	2
60.	Indoor	1	1	1	0	0	2
61.	Indoor	1	1	1	0	0	2
62.	Indoor	1	1	1	0	0	2
63.	Indoor	1	1	1	0	0	2
64.	Indoor	1	1	1	0	0	2
65.	Indoor	2	0	0	1	1	2
66.	Indoor	1	1	1	0	0	2
67.	Indoor	1	1	1	0	0	2
68.	Indoor	1	1	1	0	0	2
69.	Indoor	1	1	1	0	0	2
70.	Indoor	2	0	0	1	1	2
71.	Indoor	1	1	1	0	0	2
72.	Indoor	1	1	1	0	0	2
73.	Indoor	1	1	1	0	0	2
74.	Indoor	1	1	1	0	0	2
75.	Indoor	1	1	1	0	0	2
76.	Indoor	1	1	1	0	0	2
77.	Indoor	1	1	1	0	0	2
78.	Indoor	1	1	1	0	0	2
79.	Indoor	1	1	1	0	0	2
80.	Indoor	1	1	1	0	0	2
81.	Indoor	1	1	1	0	0	2
82.	Indoor	1	1	1	0	0	2
83.	Indoor	1	1	1	0	0	2
84.	Indoor	1	1	1	0	0	2
85.	Indoor	1	1	1	0	0	2
86.	Indoor	1	1	1	0	0	2
87.	Outdoor	1	0	0	1	1	2
88.	Outdoor	1	0	0	1	1	2
89.	Outdoor	1	0	0	1	1	2
90.	Outdoor	1	0	0	1	1	2
91.	Outdoor	1	0	0	1	1	2
92.	Outdoor	1	0	0	1	1	2
93.	Outdoor	1	0	0	1	1	2
94.	Outdoor	1	0	0	1	1	2
95.	Outdoor	1	0	0	1	1	2
96.	Outdoor	1	0	0	1	1	2
97.	Outdoor	1	0	0	1	1	2
98.	Outdoor	1	0	0	1	1	2
99.	Outdoor	1	0	0	1	1	2
100.	Outdoor	1	0	0	1	1	2
101.	Outdoor	1	0	0	1	1	2
102.	Outdoor	1	0	0	1	1	2
103.	Outdoor	1	0	0	1	1	2
104.	Outdoor	1	0	0	1	1	2
105.	Outdoor	2	1	1	0	0	2
106.	Outdoor	1	0	0	1	1	2
107.	Outdoor	1	0	0	1	1	2
108.	Outdoor	2	1	1	0	0	2
109.	Outdoor	2	1	1	0	0	2
110.	Outdoor	2	1	1	0	0	2
111.	Outdoor	2	1	1	0	0	2
112.	Outdoor	2	1	1	0	0	2
113.	Outdoor	2	1	1	0	0	2
114.	Outdoor	2	1	1	0	0	2
115.	Outdoor	2	1	1	0	0	2
116.	Outdoor	2	1	1	0	0	2
117.	Outdoor	2	1	1	0	0	2
118.	Outdoor	2	1	1	0	0	2

119.	Outdoor	2	1	1	0	0	2
120.	Outdoor	2	1	1	0	0	2
121.	Outdoor	2	1	1	0	0	2
122.	Outdoor	2	1	1	0	0	2
123.	Outdoor	2	1	1	0	0	2
124.	Outdoor	2	1	1	0	0	2
125.	Outdoor	2	1	1	0	0	2
126.	Outdoor	2	1	1	0	0	2
127.	Outdoor	2	1	1	0	0	2
128.	Outdoor	2	1	1	0	0	2
129.	Outdoor	2	1	1	0	0	2
130.	Outdoor	2	1	1	0	0	2
131.	Outdoor	2	1	1	0	0	2
132.	Outdoor	2	1	1	0	0	2
133.	Outdoor	2	1	1	0	0	2
134.	Outdoor	2	1	1	0	0	2
135.	Outdoor	2	1	1	0	0	2
136.	Outdoor	2	1	1	0	0	2
137.	Outdoor	2	1	1	0	0	2
138.	Outdoor	2	1	1	0	0	2
139.	Outdoor	2	1	1	0	0	2
140.	Outdoor	2	1	1	0	0	2
141.	Outdoor	2	1	1	0	0	2
142.	Outdoor	1	0	0	1	1	2
143.	Outdoor	1	0	0	1	1	2
144.	Outdoor	1	0	0	1	1	2
145.	Outdoor	1	0	0	1	1	2
146.	Outdoor	1	0	0	1	1	2
147.	Outdoor	1	0	0	1	1	2
148.	Outdoor	1	0	0	1	1	2
149.	Outdoor	1	0	0	1	1	2
150.	Outdoor	1	0	0	1	1	2
151.	Outdoor	1	0	0	1	1	2
152.	Outdoor	1	0	0	1	1	2
153.	Outdoor	1	0	0	1	1	2
Σ			119	119	34	34	