

Penerapan Data Mining untuk Klasifikasi Udara di Daerah Istimewa Yogyakarta Menggunakan Algoritma C4.5

Az Zahra Dwi Nur Adiya¹, Amanda Fitria Desvita², Anindya Fidela³, Dwi Amelia⁴, Tri Astuti⁵

^{1,2,3,4,5}Sistem Informasi, Universitas Amikom Purwokerto, Indonesia

Email: ¹azzahradwinuradiya@email.com, ²amandafitria719@email.com, ³fidelaanindya88@email.com, ⁴dwiameiaaa06@gmail.com, ⁵tri_astuti@amikompurwokerto.ac.id

ABSTRAK

Pencemaran udara merupakan masalah lingkungan global yang menjadi perhatian serius di berbagai wilayah di seluruh dunia, termasuk di Indonesia. Kualitas udara yang buruk memiliki dampak negatif terhadap kesehatan manusia, ekosistem, dan pertumbuhan ekonomi. Tujuan penelitian ini adalah untuk mengklasifikasikan kualitas udara yang ada di Daerah Istimewa Yogyakarta. Klasifikasi dilakukan dengan menggunakan teknik *data mining* dengan algoritma C4.5 atau *decision tree*. Hasil dari analisa yang telah diteliti menggunakan algoritma C4.5 menggunakan data berjumlah 5822 dan *sample* pengganti sebanyak 20 *instances* dengan perulangan sebanyak 100 kali jadi hasil dari penelitian ini didapatkan 7 keputusan atau *leaves* dengan tingkat akurasi keberhasilan sebesar 99.9485% dan tingkat kegagalan 0.0515%.

Kata Kunci: Kualitas Udara, Klasifikasi, Daerah Istimewa Yogyakarta, Algoritma C4.5

ABSTRACT

Air pollution is a global environmental problem that is of serious concern in various regions around the world, including in Indonesia. Poor air quality has negative impacts on human health, ecosystems and economic growth. The purpose of this research is to classify the air quality in the Special Region of Yogyakarta. Classification is done using data mining techniques with the C4.5 algorithm or decision tree. The results of the analysis that has been studied using the C4.5 algorithm used 5822 data and 20 replacement samples with 100 repetitions, so the results of this study obtained 7 decisions or leaves with a success rate of 99.9485% and a failure rate of 0.0515%.

Keywords: Air quality, Classification, Daerah Istimewa Yogyakarta, C4.5 Algorithm

Penulis Korespondensi:

Az Zahra Dwi Nur Adiya

Email: azzahradwinuradiya@gmail.com

Article Info

Diterima: 24 Juni 2024

Direvisi: 25 Juni 2024

Disetujui: 26 Juni 2024

This is an open access article under the [CC BY](#) license.



1. PENDAHULUAN

Pencemaran udara merupakan masalah lingkungan global yang menjadi perhatian serius di berbagai wilayah di seluruh dunia, termasuk di Indonesia. Kualitas udara yang buruk memiliki dampak negatif terhadap kesehatan manusia, ekosistem, dan pertumbuhan ekonomi. Daerah Istimewa Yogyakarta (DIY) sebagai salah satu provinsi di Indonesia juga menghadapi tantangan serupa terkait kualitas udara yang perlu diatasi. *Data mining* adalah salah satu teknik yang digunakan untuk menggali informasi yang berharga dari data yang besar dan kompleks. Salah satu metode *data mining* yang populer untuk klasifikasi adalah algoritma C4.5. Algoritma ini digunakan untuk membangun model klasifikasi berbasis pohon keputusan yang mampu menghasilkan aturan-aturan yang dapat digunakan untuk memprediksi atau mengklasifikasikan data yang belum diketahui.

Dalam konteks ini, penerapan *data mining* menggunakan algoritma C4.5 dapat menjadi solusi yang efektif dalam menganalisis dan mengklasifikasikan kualitas udara di Daerah Istimewa Yogyakarta. Dengan menganalisis data udara yang telah

dikumpulkan dari berbagai sumber, seperti stasiun pemantauan udara, sensor udara, dan data cuaca, algoritma C4.5 dapat digunakan untuk mengidentifikasi pola dan hubungan yang berkaitan dengan kualitas udara.

Melalui penelitian ini, dapat dihasilkan model klasifikasi yang dapat memprediksi kualitas udara di DIY berdasarkan atribut-atribut yang relevan. Hasil dari penelitian ini dapat memberikan informasi yang berharga bagi pemerintah dan masyarakat untuk mengambil langkah-langkah yang tepat dalam menjaga dan meningkatkan kualitas udara di DIY. Selain itu, penelitian ini juga dapat menjadi dasar untuk pengembangan sistem pemantauan dan peringatan dini terkait kualitas udara di DIY.

Dengan demikian, penelitian ini bertujuan untuk mengaplikasikan metode *data mining* menggunakan algoritma C4.5 untuk klasifikasi kualitas udara di Daerah Istimewa Yogyakarta. Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam memahami faktor-faktor yang memengaruhi kualitas udara, serta membantu dalam pengambilan keputusan yang lebih baik dalam upaya menjaga kualitas udara yang sehat dan berkelanjutan di DIY.

2. METODE PENELITIAN

Pada gambar di bawah ini, dapat dilihat bagaimana penelitian ini dilakukan berdasarkan alur yang dapat dijelaskan pada masing-masing poin.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Penelitian ini menggunakan *dataset* yang berasal dari Kaggle berdasarkan data yang dilaporkan pada tahun 2021. Data mencakup informasi seperti tanggal dan waktu pengukuran, pengukuran materi partikular, pengukuran belerang dioksida, pengukuran karbon monoksida, pengukuran ozon, pengukuran natrium dioksida, nilai pengukuran tertinggi, komponen yang memiliki nilai pengukuran tertinggi, dan kategori kualitas udara berdasarkan peraturan yang telah ditetapkan oleh pemerintah.

2.2 Seleksi Fitur

Seleksi fitur bertujuan guna mengurangi atribut, sebab banyaknya fitur pada *dataset* yang digunakan dapat mengakibatkan *overfitting* dan perlu menentukan fitur yang dibutuhkan dalam pengolahan data. Selain itu, dilakukan normalisasi dan pemilihan atribut yang paling berpengaruh dalam klasifikasi kualitas udara.

2.3 Klasifikasi dengan Algoritma C4.5

Klasifikasi merupakan salah satu teknik dalam pengolahan data yang bekerja dengan cara objek yang dipergunakan dibagi menjadi kelas-kelas dengan jumlah kelas sesuai dengan yang diinginkan. Klasifikasi dapat menciptakan suatu pola yang dapat memisahkan tiap-tiap kelas data yang bertujuan guna menentukan objek yang tergolong ke dalam kategori tertentu dilihat dari perilaku serta atribut dari kelompok yang telah didefinisikan. Proses termasuk dalam membangun pohon keputusan untuk mendapatkan model klasifikasi yang optimal.

2.4 Evaluasi

Model klasifikasi yang dihasilkan dievaluasi menggunakan metrik evaluasi klasifikasi seperti akurasi, presisi, recall, dan F1-score. Selain itu, validitas model juga diuji dengan menggunakan data uji yang tidak terlibat dalam proses pembangunan model.

3 HASIL DAN PEMBAHASAN

3.1. Pre-Processing

Dataset yang digunakan didapatkan dari Kaggle dalam bentuk format CSV (*Comma Separated Values*). *Dataset* tersebut berbentuk *file* terpisah berdasarkan bulan, dari bulan Januari hingga bulan Desember 2021. *Dataset* ini berisi mengenai pengukuran kualitas udara di Daerah Istimewa Yogyakarta yang dilaporkan pada tahun 2021. Penjelasan variabel atau parameter yang terdapat pada *dataset* adalah sebagai berikut:

- *Date*: tanggal pengukuran kualitas udara
- *Time*: waktu pengukuran kualitas udara
- PM10: pengukuran partikel udara yang berukuran lebih kecil dari 10 mikron (mikrometer)
- PM2.5: pengukuran partikel udara yang berukuran lebih kecil dari 2.5 mikron (mikrometer)
- SO₂: pengukuran sulfur/belerang dioksida
- CO: pengukuran karbon monoksida
- O₃: pengukuran ozon
- NO₂: pengukuran natrium dioksida
- Max: nilai pengukuran tertinggi dari seluruh komponen yang diukur dalam waktu yang sama
- *Critical Component*: komponen yang memiliki nilai pengukuran tertinggi
- *Category*: kategori kualitas udara yang diperoleh dari perhitungan rata-rata komponen PM10, PM2.5, SO₂, CO, O₃, dan NO₂ kemudian diklasifikasi berdasarkan ISPU (Indeks Standar Pencemaran Udara)

Tabel 1. Indeks ISPU

No	Nilai ISPU	Kategori
1	0 - 15	Baik
2	15 - 65	Sedang
3	66 - 150	Tidak Sehat
4	151 - 250	Sangat Tidak Sehat
5	>250	Berbahaya

Secara keseluruhan, *dataset* ini memuat 8760 *record* data dengan 10 atribut dan 1 kelas. Atribut-atribut yang terdapat dalam *dataset* ini antara lain *Date*, *Time*, PM10, PM2.5, SO₂, CO, O₃, NO₂, Max, *Critical Component*, dan *Category*.

Date	Time	PM10	PM2.5	SO2	CO	O3	NO2	Max	Critical Component	Category
01/01/2021	0:00:00	13	40	0	25	0	0	40	PM2.5	Good
02/02/2021	1:00:00	19	60	19	26	41	7	60	PM2.5	Moderate
03/03/2021	2:00:00	39	89	22	15	12	4	89	PM2.5	Moderate
04/04/2021	3:00:00	55	26	22	22	20	5	26	PM2.5	Good
05/05/2021	4:00:00	25	56	16	7		6	56	PM2.5	Moderate
06/06/2021	5:00:00	40	66		11		7	66	PM2.5	Moderate
07/07/2021	6:00:00	16	34		12		4	34	PM2.5	Good
08/08/2021	7:00:00	36	59		3		4	59	PM2.5	Moderate
09/09/2021	8:00:00	11	23		8		3	23	PM2.5	Good
10/10/2021	9:00:00	12	22	7	5	33	4	33	O3	Good
11/11/2021	10:00:00	3	8	3	13	54	3	54	O3	Moderate
12/12/2021	11:00:00	17	0	1	14	11	1	17	PM10	Good
13/01/2021	12:00:00	5	17	20	19	19	2	20	SO2	Good
14/01/2021	13:00:00	5	16	20	19	27	3	27	O3	Good
15/01/2021	14:00:00	15	50	24	26	29	3	50	PM2.5	Good
16/01/2021	15:00:00	18	55	21	22	37	3	55	PM2.5	Moderate
17/01/2021	16:00:00	19	57	25	28	28	4	57	PM2.5	Moderate
18/01/2021	17:00:00	16	51	23	22	15	4	51	PM2.5	Moderate
19/01/2021	18:00:00	20	59	28	24	35	5	59	PM2.5	Moderate
20/01/2021	19:00:00	25	62	24	26	31	5	62	PM2.5	Moderate
21/01/2021	20:00:00	16	51	24	26	27	4	51	PM2.5	Moderate
22/01/2021	21:00:00	18	55	22	27	44	5	55	PM2.5	Moderate
23/01/2021	22:00:00	27	67	23	26	48	4	67	PM2.5	Moderate
24/01/2021	23:00:00	34	77	25	29	41	7	77	PM2.5	Moderate
25/01/2021	0:00:00	32	79	21	24	25	5	79	PM2.5	Moderate
26/01/2021	1:00:00	26	68	21	24	37	5	68	PM2.5	Moderate
27/01/2021	2:00:00	8	28	21	22	39	3	39	O3	Good
28/01/2021	3:00:00	4	13	17	17	35	2	35	O3	Good
29/01/2021	4:00:00	16	53	21	28	41	4	53	PM2.5	Moderate
30/01/2021	5:00:00	22	68	22	32	51	4	68	PM2.5	Moderate
31/01/2021	6:00:00	25	67	20	26	42	6	67	PM2.5	Moderate

Gambar 2. Tabel Penggabungan Data

Proses selanjutnya adalah data *selection* atau seleksi data. Pada penelitian ini mengambil sebanyak 7 kolom dengan 6 atribut seperti PM10, PM2.5, SO2, CO, O3, NO2, dan *Category*. Kemudian masuk ke tahap *cleaning* data untuk menghapus serta memperbaiki dan menemukan data yang rusak atau tidak akurat dengan melakukan pengaturan kembali pada data-data yang ada pada catatan, tabel, atau *database*. Hasil dari *cleaning* data dari 8760 baris data menghasilkan 5822 data. Sampel data yang perlu dilakukan *cleaning* data adalah sebagai berikut:

PM10	PM2.5	SO2	CO	O3	NO2	Category
13	40	0	25	0	0	Good
19	60	19	26	41	7	Moderate
39	89	22	15	12	4	Moderate
55	26	22	22	20	5	Good
25	56	16	7		6	Moderate
40	66		11		7	Moderate
16	34		12		4	Good
36	59		3		4	Moderate
11	23		8		3	Good
12	22	7	5	33	4	Good
3	8	3	13	54	3	Moderate
17	0	1	14	11	1	Good
5	17	20	19	19	2	Good
5	16	20	19	27	3	Good
15	50	24	26	29	3	Good
18	55	21	22	37	3	Moderate
19	57	25	28	28	4	Moderate
16	51	23	22	15	4	Moderate
20	59	28	24	35	5	Moderate
25	62	24	26	31	5	Moderate

Gambar 3. Tabel Sampel *Cleaning* Data

Setelah dilakukan proses *cleaning* data maka didapatkan data siap olah. Berikut *output* yang dihasilkan setelah melakukan *preprocessing* data. Data inilah yang selanjutnya akan dilakukan klasifikasi menggunakan *Tools* Weka. Berikut tampilan data siap olah:

PM10	PM2.5	SO2	CO	O3	NO2	Category
13	40	0	25	0	0	Good
19	60	19	26	41	7	Moderate
39	89	22	15	12	4	Moderate
55	26	22	22	20	5	Good
12	22	7	5	33	4	Good
3	8	3	13	54	3	Moderate
17	0	1	14	11	1	Good
5	17	20	19	19	2	Good
5	16	20	19	27	3	Good
15	50	24	26	29	3	Good
18	55	21	22	37	3	Moderate
19	57	25	28	28	4	Moderate
16	51	23	22	15	4	Moderate
20	59	28	24	35	5	Moderate
25	62	24	26	31	5	Moderate
16	51	24	26	27	4	Moderate
18	55	22	27	44	5	Moderate
27	67	23	26	48	4	Moderate
34	77	25	29	41	7	Moderate
32	79	21	24	25	5	Moderate

Gambar 4. Tabel Data Siap Olah

3.2. Proses Klasifikasi dan Pohon Keputusan

Algoritma C4.5 menggunakan pohon keputusan dengan langkah-langkah sebagai berikut. Pertama, pilih atribut sebagai akar pohon keputusan. Kemudian, cabang dibuat untuk setiap nilai di dalam akar. Selanjutnya, bagi kasus dalam cabang. Proses ini diulang untuk setiap cabang hingga semua kasus pada cabang memiliki kelas yang sama. Dalam pemilihan atribut akar, algoritma C4.5 menggunakan metode gain. Gain dihitung untuk setiap atribut yang tersedia, dan atribut dengan gain tertinggi dipilih sebagai atribut akar. Rumus yang digunakan untuk menghitung gain adalah sebagai berikut:

$$Gain(S,A) = Entropy(S) \sum_{i=1}^n \frac{|S_i|}{|S|}$$

Keterangan :

- S = Himpunan kasus
- A = Atribut
- N = Jumlah partisi atribut A
- |Si| = Jumlah kasus pada partisi ke-i
- |S| = Jumlah kasus dalam S

Sementara itu, penghitungan nilai entropi dapat dilihat pada persamaan:

$$Entropy(S) = \sum_{i=1}^n - pi * \log_2 pi$$

Keterangan :

- S = Himpunan kasus
- A = Atribut
- N = Jumlah partisi S
- Pi = Proporsi dari Si terhadap S

Berikut hasil klasifikasi dan perhitungan pohon keputusan di Tools Weka.

```
Classifier output

J48 pruned tree
-----

PM2.5 <= 50
|  O3 <= 46
|  |  SO2 <= 49: Good (3893.57/3.0)
|  |  SO2 > 49
|  |  |  SO2 <= 50: Good (12.17)
|  |  |  SO2 > 50: Moderate (17.23/0.23)
|  O3 > 46
|  |  O3 <= 50
|  |  |  SO2 <= 38: Good (29.22)
|  |  |  SO2 > 38: Moderate (18.1/0.1)
|  |  O3 > 50: Moderate (178.23/1.23)
PM2.5 > 50: Moderate (1673.49/9.49)

Number of Leaves :      7

Size of the tree :      13

Time taken to build model: 0.22 seconds
```

Gambar 5. Hasil Klasifikasi

Sementara itu, gambar 6 di bawah ini adalah *output klasifikasi*.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      5819      99.9485 %
Incorrectly Classified Instances     3         0.0515 %
Kappa statistic                    0.9988
Mean absolute error                 0.0048
Root mean squared error             0.0335
Relative absolute error             1.1055 %
Root relative squared error         7.1552 %
Total Number of Instances          5822

=== Detailed Accuracy By Class ===

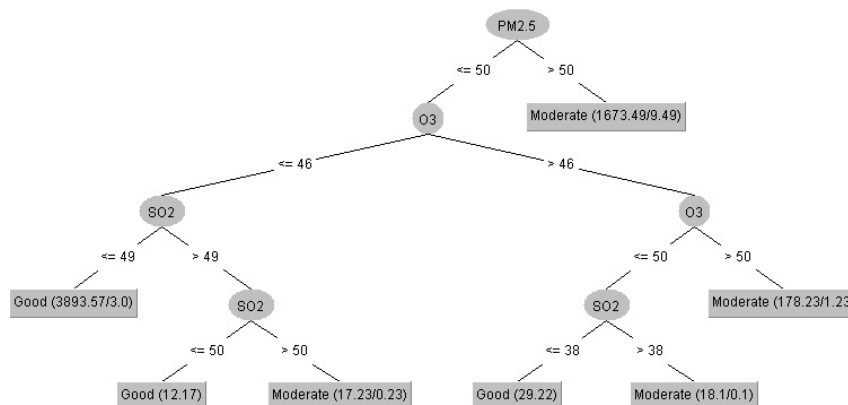
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
1,000    0,002    0,999    1,000    1,000    0,999    0,999    0,998    Good
0,998    0,000    1,000    0,998    0,999    0,999    0,999    0,999    Moderate
Weighted Avg.  0,999    0,001    0,999    0,999    0,999    0,999    0,999    0,999

=== Confusion Matrix ===

  a  b  <-- classified as
3943  0  |  a = Good
  3 1876 |  b = Moderate
    
```

Gambar 6. Output Klasifikasi

Akhirnya didapatkan pohon keputusan yang dapat dilihat pada gambar 7 di bawah ini.



Gambar 7. Pohon Keputusan

Berdasarkan pohon keputusan dari algoritma C4.5 untuk melakukan klasifikasi kualitas udara Daerah Istimewa Yogyakarta pada Gambar 4, diperoleh nilai sebagai berikut :

1. Jika nilai PM2.5 lebih besar dari 50 maka masuk kedalam kategori *moderate* (sedang).
2. Jika nilai PM2.5 kurang dari sama dengan 50 tetapi O₃ lebih dari 46 dan O₃ lebih dari 50 maka masuk kedalam kategori *moderate* (sedang).
3. Jika nilai PM2.5 kurang dari sama dengan 50 tetapi O₃ lebih dari 46 dan O₃ kurang dari sama dengan 50 dan SO₂ lebih dari 38 maka masuk kedalam kategori *moderate* (sedang).
4. Jika nilai PM2.5 kurang dari sama dengan 50 tetapi O₃ lebih dari 46 dan O₃ kurang dari sama dengan 50 dan SO₂ kurang dari sama dengan 38 maka masuk kedalam kategori *good* (baik).
5. Jika nilai PM2.5 kurang dari sama dengan 50 tetapi O₃ kurang dari sama dengan 46 dan SO₂ lebih dari 49 dan SO₂ lebih dari 50 maka masuk kedalam kategori *moderate* (sedang).
6. Jika nilai PM2.5 kurang dari sama dengan 50 tetapi O₃ kurang dari sama dengan 46 dan SO₂ lebih dari 49 dan SO₂ kurang dari sama dengan 50 maka masuk kedalam kategori *good* (baik). Jika nilai PM2.5 kurang dari sama dengan 50 tetapi O₃ kurang dari sama dengan 46 dan SO₂ kurang dari sama dengan 49 maka masuk kedalam kategori *good* (baik).

4 KESIMPULAN

Berdasarkan analisis dan pembahasan data yang telah diuji dengan menggunakan metode Algoritma C.45 untuk mengklasifikasi kualitas udara di Daerah Istimewa Yogyakarta berdasarkan data yang diambil dari Kaggle, data tersebut menggunakan 7 parameter pengukuran kualitas udara yaitu PM10, PM2.5, SO₂, CO, O₃, NO₂, dan *Category*. Melalui penerapan algoritma C4.5 dalam *data mining* untuk klasifikasi kualitas udara di Daerah Istimewa Yogyakarta, ditemukan bahwa kualitas

udara diklasifikasikan menjadi dua kategori, yaitu *good* (baik) dan *moderate* (sedang), dengan rasio yang berbeda setiap harinya. Setelah melakukan proses klasifikasi, dapat disimpulkan bahwa kualitas udara yang dianalisis dari Januari hingga Desember 2021 masuk dalam kategori *good* (baik), hal ini dapat dilihat dari tingkat akurasi klasifikasi yang mencapai 99.9485%. Hasil ini menunjukkan bahwa model klasifikasi berhasil mengidentifikasi dan memprediksi dengan akurasi yang tinggi bahwa sebagian besar data pengukuran kualitas udara berada dalam kategori yang memenuhi standar kualitas udara yang baik.

REFERENSI

- [1] Firmansyah, A., Dewi, A. L., Hirna, E. S., Briliyanto, M. B. A., Nurjannah, M., & Nooraeni, R., "Pengelompokan Titik Wilayah di Provinsi Daerah Istimewa Yogyakarta Berdasarkan Kualitas Udara Menggunakan Algoritma Fuzzy C-Means", *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 8, no. 2, Juli-Desember 2020.
- [2] Novianto, H., Azis, M. M., & Arini, H. M., "Analisis Perubahan Sistem Kualitas Udara Kota Yogyakarta Pada Masa Pandemi Covid-19", *Jurnal Rekayasa Proses*, vol. 16, no. 2, 2020.
- [3] Ergianzah R S. 2021. Pemetaan Sebaran Kualitas Udara Ambien Kawasan Perkotaan Yogyakarta Dengan Parameter SO₂, CO, dan NO₂ Metode Inverse Distance Weighting (IDW). Tugas Akhir. Fakultas Teknik Sipil dan Perencanaan. Universitas Islam Indonesia.
- [4] Astriyani, M., Laela, I. N., Lestari, D. P., Anggraeni, L., & Astuti, T., "Analisis Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma C4.5", *Jurnal Sistem & Teknologi Informasi Komunikasi*, vol. 6, no.1.
- [5] Kirono, A. A. H., Asror, I., & Wibowo, Y. F. A., "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Menggunakan Algoritma Naive Bayes", *e-Proceeding of Engineering*, vol. 9, no. 3, Juni 2022.
- [6] Julianto H. 2021. Klasifikasi Kualitas Udara Menggunakan Algoritma Learning Vector Quantization di Kota Pekanbaru. Tugas Akhir. Fakultas Sains dan Teknologi. Universitas Islam Negeri Sultan Syarif Kasim.
- [7] Sodiq, M. J., & Sela, E. I. 2019. Perbandingan Metode Naive Bayes dan K-Nearest Neighbor pada Klasifikasi Kualitas Udara di DKI Jakarta. Tugas Akhir. Fakultas Teknologi Informasi dan Elektro. Universitas Teknologi Yogyakarta.
- [8] Alkabbani, H., Ramadan, A., Zhu, Q., Elkamel, A. An Improved Air Quality Index Machine Learning-Based Forecasting with Multivariate Data Imputation Approach. *Atmosphere* 2022, 13, 1144. doi: 10.3390/atmos3071144.
- [9] A. Amalia, A. Zaidiah and I. N. Isnainiyah, *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, no. 2, pp. 496-507, 2021.
- [10] K. y. Ananta, E. Anugrayningtyas, F. A. Abdul Aziz and D. Hartanti, "Penerapan Metode K-Nearest Neighbor (K-NN) Dalam Klasifikasi Kualitas Udara di Provinsi DKI Jakarta," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, pp. 709-712, 2023.
- [11] A. S. Handayani, S. Soim, T. E. Agusdi, Rumiasih and A. Nurdin, "KLASIFIKASI KUALITAS UDARA DENGAN METODE SUPPORT VECTOR MACHINE," *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, vol. 3, no. 2, pp. 187-199, 2020.
- [12] A. Inez Sang, E. Sutoyo and I. Darmawan, "ANALISIS DATA MINING UNTUK KLASIFIKASI DATA KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE," *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 8964-8963, 2021.
- [13] S. Nurjanah and M. A. K. D. S. Siregar, "PENERAPAN ALGORITMA K – NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI PENCEMARAN UDARA DI KOTA JAKARTA," *Scientific Student Journal for Information, Technology and Science*, vol. 1, no. 2, pp. 71-76, 2020.
- [14] I. I. Ridho and G. Mahalisa, "ANALISIS KLASIFIKASI DATASET INDEKS STANDAR PENCEMARAN UDARA (ISPU) DI MASA PANDEMI MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Technologia*, vol. 14, no. 1, pp. 38-41, 2023.
- [15] A. Umri, S. S, M. S. Firdaus and A. Primajaya, "ANALISIS DAN KOMPARASI ALGORITMA KLASIFIKASI DALAM INDEKS PENCEMARAN UDARA DI DKI JAKARTA," *JIKO (Jurnal Informatika dan Komputer)*, vol. 4, no. 2, pp. 98-104, 2021.