

KLASIFIKASI DATA MINING UNTUK MENENTUKAN KUALITAS UDARA DI PROVINSI DKI JAKARTA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (K-NN)

Ade Davy Wiranata¹, Soleman², Irwansyah³, I Ketut Sudaryana⁴, Rizal⁵

^{1,3,4,5}Teknik Informatika

²Sistem Informasi

^{1,3}Universitas Muhammadiyah Prof. Dr. HAMKA Jakarta, Indonesia.

²Universitas Borobudur Jakarta, Indonesia.

^{4,5}Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Widuri Jakarta, Indonesia.

Correspondence email: solemediagrafik@gmail.com

Article history:

Submission date: Juni 25, 2023

Revised date: Juni 29, 2023

Published date: Juni 30, 2023

ABSTRACT

Air plays an important role in maintaining the life of living things on earth. Metabolic processes that occur in the bodies of living things cannot take place without oxygen from the air. The air pollution problem in DKI Jakarta is very serious and can cause health problems such as irritation of the respiratory tract, respiratory diseases, and long-term health problems such as cardiovascular disease and lung cancer. Air pollution can also affect environmental quality, reduce visibility, and damage ecosystems. The purpose of this study is to determine the accuracy of classifying air quality in DKI Jakarta province. The data mining method that the author uses is the K-Nearest Neighbors (K-NN) algorithm. From the results of the evaluation process of the K-Nearest Neighbors (K-NN) algorithm using the K-5 fold that has been carried out using the RapidMiner tool, the results of K-2 fold accuracy of 73.97%, K-3 fold accuracy of 72.60%, K-4 fold accuracy of 72.60%, and K-5 fold accuracy of 75.35%.

Keywords: Air Pollution, DKI Jakarta, Data Mining, K-NN.

ABSTRAK

Udara mengambil peran penting dalam menjaga kehidupan makhluk hidup di bumi. Proses metabolisme yang terjadi dalam tubuh makhluk hidup tidak dapat berlangsung tanpa adanya oksigen yang berasal dari udara. Permasalahan polusi udara di DKI Jakarta sangat serius dan dapat menyebabkan masalah kesehatan seperti iritasi saluran pernapasan, penyakit pernapasan, dan masalah kesehatan jangka panjang seperti penyakit kardiovaskular dan kanker paru-paru. Polusi udara juga dapat mempengaruhi kualitas lingkungan, mengurangi visibilitas, dan merusak ekosistem. Tujuan dari penelitian ini adalah untuk mengetahui akurasi dari klasifikasikan kualitas udara yang ada di provinsi DKI Jakarta. Metode data mining yang penulis gunakan adalah algoritma *K-Nearest Neighbors* (K-NN). Dari hasil proses evaluasi algoritma *K-Nearest Neighbors* (K-NN) menggunakan K-5 fold yang telah dilakukan menggunakan *tools* RapidMiner diperoleh hasil akurasi K-2 fold sebesar 73,97%, akurasi K-3 fold sebesar 72,60%, akurasi K-4 fold sebesar 72,60%, dan akurasi K-5 fold sebesar 75,35%.

Kata Kunci: Polusi Udara, DKI Jakarta, Data Mining, K-NN.

PENDAHULUAN

Udara mengambil peran penting dalam menjaga kehidupan makhluk hidup di bumi. Proses metabolisme yang terjadi dalam tubuh makhluk hidup tidak dapat berlangsung tanpa adanya oksigen yang berasal dari udara.

Kualitas udara melibatkan kandungan polutan seperti partikel padat, gas-gas beracun, ozon, sulfur dioksida, nitrogen dioksida, karbon monoksida, dan timbal di udara yang dapat mempengaruhi kesehatan manusia dan lingkungan. Pencemaran udara sebagai ancaman terbesar terhadap kesehatan lingkungan (Andrianto, 2017).



Menurut WHO, polusi udara menyebabkan sekitar 7 juta kematian setiap tahunnya dan berkontribusi pada berbagai masalah kesehatan seperti penyakit pernapasan, penyakit kardiovaskular, dan kanker. Pencemaran udara membuat kualitas udara disuatu daerah menjadi turun, dan kurang bagus untuk makhluk hidup, bahkan bisa membahayakan. Banyak faktor yang memicu penurunan kualitas udara disuatu daerah diantaranya aktivitas transportasi, industri, jasa, kegiatan lainnya yang meningkat, dan buangan sisa-sisa kegiatan ke udara (Sani, 2016). Di kota besar seperti DKI Jakarta, pencemaran udara merupakan salah satu masalah yang cukup serius yang perlu dihadapi oleh masyarakat DKI Jakarta. Permasalahan polusi udara di DKI Jakarta sangat serius dan dapat menyebabkan masalah kesehatan seperti iritasi saluran pernapasan, penyakit pernapasan, dan masalah kesehatan jangka panjang seperti penyakit kardiovaskular dan kanker paru-paru. Polusi udara juga dapat mempengaruhi kualitas lingkungan, mengurangi visibilitas, dan merusak ekosistem (Rahmawati, 2016).

Data mining adalah proses ekstraksi pengetahuan atau informasi yang bermanfaat dari kumpulan data yang besar dan kompleks. Tujuan dari data mining adalah menemukan pola, hubungan, dan tren tersembunyi yang tidak dapat dengan mudah diidentifikasi melalui analisis manusia biasa (Budiyantara et al., 2020).

Dalam kaitannya data mining memiliki beberapa teknik salah satunya adalah teknik klasifikasi. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Klasifikasi memiliki beberapa Algoritma klasifikasi data mining salah satunya yaitu *K-Nearest Neighbors* (K-NN) (Budiyantara et al., 2020).

Algoritma *K-Nearest Neighbors* (K-NN) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk masalah klasifikasi dan regresi. Algoritma ini mengambil pendekatan berbasis instansi atau non-parametrik, yang berarti tidak ada model yang dihasilkan dari proses pelatihan. Sebaliknya, algoritma ini menyimpan seluruh set data pelatihan dalam memori dan mengklasifikasikan atau memprediksi label untuk data baru berdasarkan kesamaan dengan tetangga terdekat (Tupan Tri M dan Irwansyah, 2020).

Penelitian terdahulu dari Abdul Aziiz Hendrie Kirono, Ibnu Asror, Yanuar Firdaus Arie Wibowo dengan judul klasifikasi tingkat kualitas udara DKI Jakarta menggunakan algoritma *naive bayes*. Algoritma *naive bayes* adalah salah satu algoritma klasifikasi berdasarkan probabilitas yang membandingkan data *training* dan data *testing*. Keduanya dibandingkan melalui beberapa tahap persamaan, yang akhirnya diperoleh probabilitas tertinggi yang ditetapkan sebagai informasi. Membuat sebuah model uji pada klasifikasi menggunakan algoritma *naive bayes* yang bertujuan

mencari hasil yang baik. Hasil dari pengklasifikasian pada data Indeks Standar Pencemaran Udara pada kota DKI Jakarta menghasilkan yaitu, dengan rata-rata akurasi 88%, precision 85%, recall 96%, f1-score 90% (Aziiz et al., 2022).

Penelitian terdahulu dari Maya Astriyani, Ida Nur Laela, Dwi Puji Lestari, Laudiana Anggraeni, Tri Astuti dengan judul analisis klasifikasi data kualitas udara DKI Jakarta menggunakan algoritma C.45. Hasil dari penerapan algoritma C.45 data mining untuk klasifikasi kualitas udara di DKI Jakarta didapatkan bahwa kualitas udara dikategorikan menjadi tiga macam yaitu sedang, baik, dan tidak sehat dengan rasio yang berbeda-beda setiap harinya. Setelah dilakukan klasifikasi maka dapat disimpulkan kualitas udara yang dianalisis pada bulan Januari-Desember 2021 masuk kedalam kategori sedang dimana tingkat kualitas udara masih dapat diterima untuk kesehatan manusia, hewan, dan tumbuhan (Astriyani et al., 2023).

METODE PENELITIAN

Metode Penelitian

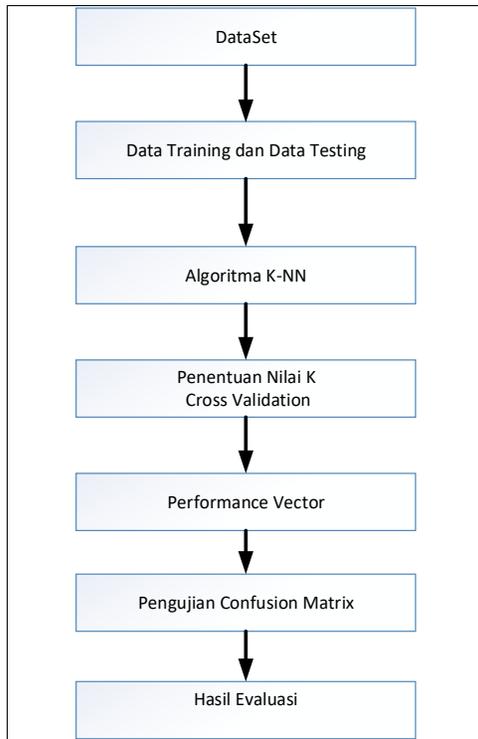
K-Nearest Neighbor (K-NN) adalah sebuah algoritma klasifikasi dan regresi yang populer dalam pembelajaran mesin (*machine learning*) (Hakim Tanjung, 2021). Algoritma ini bekerja berdasarkan prinsip "nearest neighbor", di mana prediksi untuk sebuah instansi baru didasarkan pada label atau nilai-nilai yang dimiliki oleh tetangga terdekatnya dalam ruang fitur. Dalam algoritma K-NN, nilai K adalah jumlah tetangga terdekat yang akan dipertimbangkan untuk melakukan prediksi. Langkah-langkah utama dalam algoritma K-NN adalah sebagai berikut:

1. Menentukan nilai K: K adalah parameter yang harus ditentukan sebelumnya. Ini mengatur jumlah tetangga terdekat yang akan digunakan dalam proses prediksi.
2. Menghitung jarak: Algoritma menghitung jarak antara instansi baru yang akan diprediksi dengan setiap instansi dalam dataset pelatihan. Jarak ini dapat dihitung menggunakan berbagai metrik seperti *Euclidean distance*, *Manhattan distance*, atau metrik jarak lainnya tergantung pada konteks masalah. Yang paling sering digunakan adalah *euclidean distance* (Bramer, 2007), yaitu sebagai berikut:

$$X = \sqrt{(a1 - b1)^2 + (a2 - b2)^2 + \dots + (an - bn)^2} \dots (1)$$

3. Memilih K tetangga terdekat: Algoritma memilih K tetangga terdekat berdasarkan jarak yang dihitung sebelumnya. Ini bisa dilakukan dengan mencari K instansi dengan jarak terpendek dari instansi baru.
4. Menentukan label atau nilai prediksi: Untuk masalah klasifikasi, label yang paling umum atau mayoritas dari

tetangga terdekat akan diambil sebagai prediksi. Untuk masalah regresi, nilai prediksi dapat diambil sebagai rata-rata atau median nilai tetangga terdekat.



Sumber: (Wiranata et al., 2023)

Gambar 1. Langkah-Langkah Penelitian Dengan Algoritma *K-Nearest Neighbors*

Berikut ini merupakan penjelasan dari langkah-langkah penelitian dengan algoritma *K-Nearest Neighbors*.

1. Data set : kumpulan data yang terorganisir dan terstruktur yang digunakan untuk melakukan analisis dan pemodelan. Data yang digunakan dalam penelitian ini sebanyak 244 data (Tupan et al., 2021).
2. Data *training* dan data *testing* : Dalam penelitian ini 171 data digunakan untuk data latih (*training*) dan data uji (*testing*) sebanyak 73 data.

3. Algoritma K-NN : Pada penelitian ini penulis menggunakan algoritma K-NN sebagai model klasifikasi.
4. Penentuan nilai K *cross validation* : Nilai K dalam *cross-validation* mengacu pada jumlah lipatan (*fold*) yang digunakan dalam proses validasi silang (*cross-validation*). Pada penelitian ini penulis menggunakan *K-5 fold cross validation*.
5. *Performance vector* : Pada penelitian ini penulis mengacu pada evaluasi dan pengukuran kinerja model yang dibangun dengan menggunakan *platform RapidMiner* yaitu akurasi.
6. Pengujian *confusion matrix* : alat evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan prediksi model atau hasil dari *performance vector* dengan nilai sebenarnya dari dataset.
7. Hasil Evaluasi : Membandingkan antara *performance vector* dengan hasil pengujian *confusion matrix* apakah sama atau berbeda hasil akurasi.

HASIL DAN PEMBAHASAN

Pengolahan Data (*Data Preparation*)

Pada tahapan ini dilakukan persiapan data awal, pemilihan atribut atau variabel yang akan dianalisis, validasi variabel yang ada serta persiapan transformasi data. Adapun hal-hal yang dihasilkan pada fase persiapan data ini adalah :

Dilakukan persiapan data *training* sebanyak 171 data dan data *testing* sebanyak 73 data. Atribut yang digunakan terdiri dari 9 atribut yaitu : partikel udara (Pm10), sulfur dioksida (so2), karbon monoksida (co), ozon (o3), nitrogen dioksida (no2), nilai max, kritikal, lokasi_spku, kategori sebagai atribut hasil.

Berikut ini data atribut yang akan digunakan pada penelitian ini dapat kita lihat pada tabel 1 sebagai berikut :

Pengelompokan Data

Pada tabel 1 dibawah ini merupakan 10 contoh tabel data uji (*testing*) yang digunakan dalam penelitian ini sebanyak 73 data.

Tabel 1. Contoh Sampel Data Testing

Partikel udara (Pm10)	Sulfur Dioksida (so2)	Karbon monoksida (co)	Ozon (o3)	Nitrogen dioksida(no2)	Nilai Max	Kritikal	lokasi_spku	Kategori
38	36	25	46	9	46	O3	DKI5	Baik
45	36	39	102	8	102	O3	DKI5	Tidak Sehat
51	37	27	63	10	63	O3	DKI5	Sedang
51	38	19	85	10	85	O3	DKI5	Sedang
52	39	25	62	9	62	O3	DKI5	Sedang
62	37	39	64	9	64	O3	DKI5	Sedang

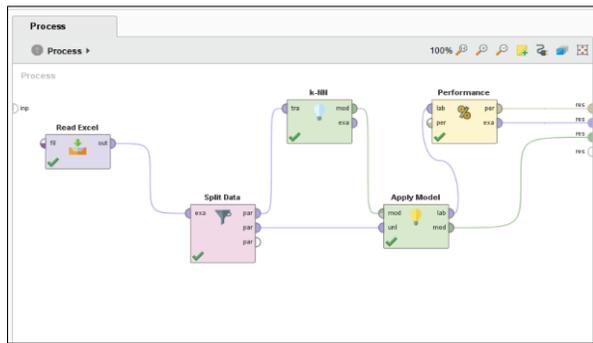


Partikel udara (Pm10)	Sulfur Dioksida (so2)	Karbon monoksida (co)	Ozon (o3)	Nitrogen dioksida(no2)	Nilai Max	Kritikal	lokasi_spku	Kategori
50	38	39	66	11	66	O3	DKI5	Sedang
52	40	22	70	11	70	O3	DKI5	Sedang
82	38	58	71	13	82	PM10	DKI4	Sedang
44	37	27	47	9	47	O3	DKI5	Baik

Sumber : (data.jakarta.go.id, 2020)

Desain Proses Menggunakan RapidMiner

Desain proses menggunakan *Tools* RapidMiner dapat kita lihat pada gambar 2 sebagai berikut:



Sumber: (Wiranata et al., 2023)

Gambar 2. Proses Desain dan *Import Data* Dengan RapidMiner

1. *Read Excel* : Operator yang dapat digunakan untuk membaca *file Excel*. Operator yang paling umum digunakan adalah "*Read Excel*" atau "*Read Excel Spreadsheet*". Operator ini memungkinkan Anda untuk menentukan lembar kerja (*sheet*) yang ingin dibaca dan rentang sel (*range*) yang ingin diproses pada tools RapidMiner.
2. *Split Data*: operator yang digunakan untuk membagi data menjadi set pelatihan (*training set*) dan set pengujian (*testing set*). Operator ini memungkinkan Anda untuk membagi data ke dalam dua set berdasarkan proporsi yang ditentukan. Pada penelitian ini penulis membagi data 0,7 untuk data *training* dan 0,3 untuk data *testing*.
3. *K-Nearest Neighbors* : Algoritma klasifikasi yang digunakan penulis pada penelitian ini
4. *Apply Model* : Digunakan untuk menerapkan model yang telah dilatih pada data baru. Operator ini memungkinkan Anda untuk menggunakan model yang sudah ada untuk melakukan prediksi atau klasifikasi pada data yang belum dilihat sebelumnya.
5. *Performance* : Digunakan dalam menganalisis kinerja model yaitu untuk mengetahui hasil akurasi dari algoritma K-NN.

Performance Vector K-5 dan Confusion Matrix

1. Hasil Akurasi Performance Vector K-2 Fold

Penulis menggunakan parameter k-2 *fold* dapat kita lihat pada gambar 3 sebagai berikut :

	true BAIK	true TIDAK SEHAT	true SEDANG	class precision
pred BAIK	0	0	1	0.00%
pred TIDAK SEHAT	1	7	7	46.67%
pred SEDANG	2	8	47	82.46%
class recall	0.00%	46.67%	85.45%	

Sumber: (Wiranata et al., 2023)

Gambar 3. Hasil Akurasi K-2 *Fold*

Dari proses parameter k-2 *fold* algoritma *K-Nearest Neighbors* terbentuk hasil matrix *Accuracy* sebesar 73.97%.

Hasil *performance vector* dan *confusion matrix* dapat dilihat pada gambar 4 berikut ini.

PerformanceVector

PerformanceVector:
accuracy: 73.97%

ConfusionMatrix:

True:	BAIK	TIDAK SEHAT	SEDANG
BAIK:	0	0	1
TIDAK SEHAT:	1	7	7
SEDANG:	2	8	47

kappa: 0.295

ConfusionMatrix:

True:	BAIK	TIDAK SEHAT	SEDANG
BAIK:	0	0	1
TIDAK SEHAT:	1	7	7
SEDANG:	2	8	47

Sumber: (Wiranata et al., 2023)

Gambar 4. Performance Vector K-2 *Fold*

2. Hasil Akurasi Performance Vector K-3 Fold

Penulis menggunakan parameter k-3 *fold* dapat kita lihat pada gambar 5 sebagai berikut :

	true BAIK	true TIDAK SEHAT	true SEDANG	class precision
pred BAIK	0	0	1	0.00%
pred TIDAK SEHAT	0	3	4	42.86%
pred SEDANG	3	12	50	76.92%
class recall	0.00%	20.00%	90.91%	

Sumber: (Wiranata et al., 2023)

Gambar 5. Hasil Akurasi K-3 fold

Dari proses parameter k-3 fold algoritma *K-Nearest Neighbors* terbentuk hasil matrix *Accuracy* sebesar 72.60%.

Hasil *performance vector* dan *confusion matrix* dapat dilihat pada gambar 6. berikut ini.

```

PerformanceVector:
accuracy: 72.60%
ConfusionMatrix:
True:  BAIK  TIDAK SEHAT  SEDANG
BAIK:  0      0            1
TIDAK SEHAT:  0      3            4
SEDANG:  3      12           50
kappa: 0.113
ConfusionMatrix:
True:  BAIK  TIDAK SEHAT  SEDANG
BAIK:  0      0            1
TIDAK SEHAT:  0      3            4
SEDANG:  3      12           50
    
```

Sumber: (Wiranata et al., 2023)

Gambar 6. *Performance Vector* K-3 Fold

3. Hasil Akurasi *Performance Vector* K-4 Fold

Penulis menggunakan parameter k-4 fold dapat kita lihat pada gambar 7 sebagai berikut :

	true BAIK	true TIDAK SEHAT	true SEDANG	class precision
pred BAIK	0	0	2	0.00%
pred TIDAK SEHAT	0	3	3	50.00%
pred SEDANG	3	12	50	76.92%
class recall	0.00%	20.00%	90.91%	

Sumber: (Wiranata et al., 2023)

Gambar 7. Hasil Akurasi K-4 Fold

Dari proses parameter k-4 fold algoritma *K-Nearest Neighbors* terbentuk hasil matrix *Accuracy* sebesar 72.60%.

Hasil *performance vector* dan *confusion matrix* dapat dilihat pada gambar 8 berikut ini

```

PerformanceVector:
accuracy: 72.60%
ConfusionMatrix:
True:  BAIK  TIDAK SEHAT  SEDANG
BAIK:  0      0            2
TIDAK SEHAT:  0      3            3
SEDANG:  3      12           50
kappa: 0.119
ConfusionMatrix:
True:  BAIK  TIDAK SEHAT  SEDANG
BAIK:  0      0            2
TIDAK SEHAT:  0      3            3
SEDANG:  3      12           50
    
```

Sumber: (Wiranata et al., 2023)

Gambar 8. *Performance Vector* K-4 Fold

4. Hasil Akurasi *Performance Vector* K-5 Fold

Penulis menggunakan parameter k-5 fold dapat kita lihat pada gambar 9 sebagai berikut :

	true BAIK	true TIDAK SEHAT	true SEDANG	class precision
pred BAIK	0	0	0	0.00%
pred TIDAK SEHAT	0	3	3	50.00%
pred SEDANG	3	12	52	77.81%
class recall	0.00%	20.00%	94.55%	

Sumber: (Wiranata et al., 2023)

Gambar 9. Hasil Akurasi K-5 Fold

Dari proses parameter k-4 fold algoritma *K-Nearest Neighbors* terbentuk hasil matrix *Accuracy* sebesar 75.34%.

Hasil *performance vector* dan *confusion matrix* dapat dilihat pada gambar 10 berikut ini.

```

PerformanceVector:
accuracy: 75.34%
ConfusionMatrix:
True:  BAIK  TIDAK SEHAT  SEDANG
BAIK:  0      0            0
TIDAK SEHAT:  0      3            3
SEDANG:  3      12           52
kappa: 0.154
ConfusionMatrix:
True:  BAIK  TIDAK SEHAT  SEDANG
BAIK:  0      0            0
TIDAK SEHAT:  0      3            3
SEDANG:  3      12           52
    
```

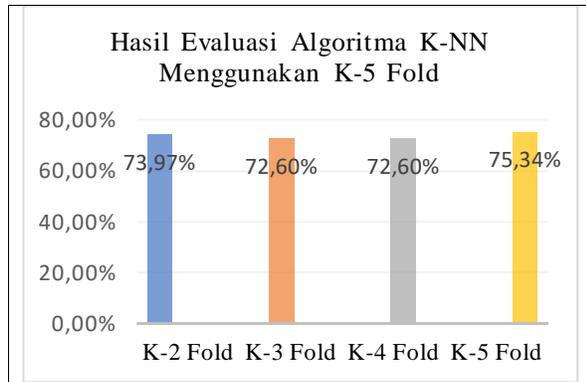
Sumber: (Wiranata et al., 2023)

Gambar 10. *Performance Vector* K-5 Fold

Hasil Evaluasi Menggunakan K-5 Fold

Berikut ini merupakan grafik hasil evaluasi algoritma K-NN menggunakan parameter K-5 fold dapat dilihat pada gambar 11. Berikut ini:





Sumber: (Wiranata et al., 2023)

Gambar 11. Grafik Hasil Akurasi Algoritma K-NN

Pada *diagram chart* dapat kita lihat hasil akurasi pada masing-masing parameter. Akurasi K-2 fold sebesar 73,97%, akurasi K-3 *fold* sebesar 72,60%, akurasi K-4 *fold* sebesar 72,60%, dan akurasi K-5 *fold* sebesar 75,35%.

KESIMPULAN

Data yang digunakan pada penelitian ini sebanyak 244 data yang terbagi menjadi 171 data digunakan untuk data latih (*training*) dan data uji (*testing*) sebanyak 73 data. Atribut yang digunakan terdiri dari 9 atribut yaitu : Partikel Udara (Pm10), Sulfur Dioksida (so2), Karbon Monoksida (co), Ozon (o3), Nitrogen Dioksida (no2), nilai max, kritikal, lokasi_spku, kategori sebagai atribut hasil. Dari hasil proses evaluasi algoritma *K-Nearest Neighbors* (K-NN) menggunakan *K-5 fold* yang telah dilakukan menggunakan *tools* RapidMiner diperoleh hasil akurasi K-2 *fold* sebesar 73,97%, akurasi K-3 *fold* sebesar 72,60%, akurasi K-4 *fold* sebesar 72,60%, dan akurasi K-5 *fold* sebesar 75,35%.

DAFTAR PUSTAKA

- Ade Davy Wiranata, Soleman, Irwansyah, I Ketut Sudaryana, R. (2023). *Klasifikasi Data Mining Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta Menggunakan Algoritma K-Nearest Neighbors (K-NN)*.
- Agus Budiyantra, Irwansyah, Egi Prengki, P. A. P. (2020). *Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu*.
- Andrianto, P. (2017). Sistem Informasi Pelayanan Kesehatan Berbasis Web di Puskesmas. *Prosiding Seminar Nasional Komputer Dan Informatika (SENASKI)*, 2017, 978–602.
- Astriyani, M., Laela, I. N., Lestari, D. P., Anggraeni, L., & Astuti, T. (2023). Analisis Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma C.45. *JuSiTik: Jurnal Sistem Dan Teknologi Informasi Komunikasi*, 6(1), 36–41. <https://doi.org/10.32524/jusitik.v6i1.790>
- Aziiz, A., Kirono, H., Asror, I., Firdaus, Y., & Wibowo, A. (2022). Klasifikasi Tingkat Kualitas Udara Dki Jakarta Menggunakan Algoritma Naive Bayes. *EProceedings ...*, 9(3), 1962–1969. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/18002%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/18002/17631>
- Bramer, M. (2007). *Principles of Data Mining: Undergraduate Topics in Computer Science*. Springer-Verlag.
- Budiyantra, A., Irwansyah, I., Prengki, E., Pratama, P. A., & Wiliani, N. (2020). Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 265–270. <https://doi.org/10.33480/jitk.v5i2.1214>
- Hakim Tanjung, D. Y. (2021). Optimalisasi Algoritma C4.5 untuk Prediksi Kerusakan Mesin ATM. *Infosys (Information System) Journal*, 6(1), 12. <https://doi.org/10.22303/infosys.6.1.2021.12-21>
- Rahmawati, E. (2016). Sistem Pakar Diagnosis Penyakit Paru-Paru Menggunakan Metode Forward Chaining. *Jurnal Teknik Elektro*, 8(2), 64–69. <https://doi.org/10.15294/jte.v8i2.7436>
- Sani, A. (2016). *Analisa Penjualan Retail Dengan Metode Association Rule Untuk AssociationRule Untuk Pengambilan Keputusan Strategis Perusahaan* : 2(June 2016), 34–50.
- Tupan Tri M dan Irwansyah. (2020). *Implementasi Data Mining Untuk Menentukan Kelayakan Pemberian Kredit Dengan Menggunakan Algoritma K-Nearest Neighbors (K-Nn)*.
- Tupan Tri Muryono, Ahmad Taufik, I. (2021). *Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit*.