

Komparasi Algoritma K-Nearest Neighbor dan Naive Bayes pada Klasifikasi Tingkat Kualitas Udara Kota Tangerang Selatan

Avira Budianita^{1*}, Nurul Iman², Fida Maisa Hana³, Cikita Berlian Hakim⁴

¹Program Studi Bisnis Digital, Fakultas Ekonomi Pendidikan dan Hukum, Universitas Muhammadiyah Kudus

^{2,4}Program Studi Teknik Industri, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kudus

³Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kudus

*Email: avirabudianita@umkudus.ac.id

Abstract

The growth of technology and the impact of industrial activities on the earth have an influence on environmental changes, including changes that are felt are a decrease in air quality or air pollution which has an impact on the health of the human body. Based on this, this research aims to produce a model for solving air quality classification problems based on parameter indicators. A comparative evaluation was also carried out on the classification of the K-Nearest Neighbor and Naive Bayes algorithm methods on the air quality dataset in South Tangerang in 2022. At the same ratio in the classification process, the K-Nearest Neighbor algorithm got an accuracy value of 94.44% and the Naive Bayes algorithm got an accuracy value of 94.44%. Accuracy value 86.11%. From the results of testing the data, it can be concluded that the K-Nearest Neighbor algorithm has high accuracy compared to the Naive Bayes algorithm in air level classification.

Keywords: Data Mining, ISPU, K-Nearest Neighbor, Naive Bayes

Abstrak

Pertumbuhan teknologi dan dampak aktifitas industri di bumi memberikan pengaruh terhadap perubahan lingkungan, diantaranya perubahan yang dirasakan adalah penurunan kualitas udara atau polusi udara yang berdampak pada kesehatan tubuh manusia. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menghasilkan model dalam penyelesaian masalah klasifikasi kualitas udara terhadap indikator parameter. Serta dilakukan evaluasi komparasi terhadap klasifikasi metode algoritma *K-Nearest Neighbor* dan *Naive Bayes* pada dataset kualitas udara di Tangerang selatan pada tahun 2022. Pada rasio yang sama dalam proses klasifikasi, algoritma *K-Nearest Neighbor* mendapatkan nilai *Accuracy* 94.44% dan pada algoritma *Naive Bayes* mendapatkan nilai *Accuracy* 86.11%. Dari hasil pengujian data tersebut dapat disimpulkan bahwa Algoritma *K-Nearest Neighbor* memiliki akurasi tinggi dibandingkan algoritma *Naive Bayes* pada klasifikasi tingkat udara.

Kata kunci: Data Mining, ISPU, K-Nearest Neighbor, Naive Bayes

PENDAHULUAN

Pertumbuhan teknologi dan dampak aktifitas industri di bumi memberikan pengaruh terhadap perubahan lingkungan, diantaranya perubahan lingkungan yang dirasakan adalah penurunan kualitas udara atau polusi udara berdampak pada kesehatan tubuh manusia. Menurut ("PP No 41 Tahun 1999," n.d.) pencemaran udara merupakan dimasukkannya atau masuknya zat, energi, dan komponen lain oleh kegiatan manusia ke dalam udara ambien,

sehingga mutu udara ambien turun sampai ketinggian tertentu yang menyebabkan udara ambien tidak dapat memenuhi fungsinya. Kualitas udara merujuk pada kondisi udara di suatu wilayah yang mencakup berbagai parameter fisik, kimia, dan biologis yang mempengaruhi kesehatan manusia, lingkungan, dan ekosistem secara keseluruhan, parameter ini disebut sebagai ISPU.

Indeks Standar Pencemar Udara merupakan sebuah indikator tanpa satuan yang

digunakan sebagai gambaran tingkat kualitas udara di suatu lokasi wilayah tertentu. Indikator ini berdasarkan pada dampak terhadap kesehatan manusia, nilai estetika, dan makhluk hidup berada di bumi. Tujuan dari penggunaan ISPU adalah untuk menyediakan informasi yang seragam mengenai kualitas udara kepada masyarakat di waktu dan lokasi tertentu, serta sebagai acuan bagi pemerintah baik pemerintah pusat maupun pemerintah daerah sebagai upaya pengendalian pencemaran udara. Nilai ISPU diperoleh melalui penggunaan sistem pemantauan kualitas udara otomatis (*Air Quality Monitoring System/AQMS*) (Dian, Emilia, Rizka, et al., 2023). Adapun berbagai zat polutan udara yaitu Karbon Monoksida (CO), Sulfur Dioksida (SO₂), Ozon permukaan (O₃), Oksida Nitrogen (NO_x), dan Partikulat Matter (PM₁₀).

Penelitian ini dilakukan klasifikasi kualitas udara di Tangerang Selatan pada tahun 2022 menggunakan metode algoritma *K-Nearest Neighbor* dan *Naive Bayes* untuk memberikan informasi berupa klasifikasi tingkatan pencemaran udara dengan indikator parameter berbasis desktop. Data ISPU yang digunakan bersumber dari data publik tersedia di website Kaggle. Dalam laporan ini, akan dilakukan analisis data berupa perbandingan rasio untuk memperoleh tingkat akurasi yang tertinggi

TINJAUAN PUSTAKA

Penelitian ini menggunakan dua algoritma yaitu algoritma *K-Nearest Neighbor* dan algoritma *Naive Bayes*. Hasil akhir yang ingin dicapai dalam penelitian ini adalah membandingkan rasio dan akurasi dua algoritma tersebut dalam proses klasifikasi tingkat kualitas udara, sejauh mana tingkat pencemaran udara menggunakan data kualitas udara di Tangerang Selatan pada tahun 2022. Perbedaan penelitian ini dengan penelitian terdahulu adalah penggunaan data serta jenis dan jumlah algoritma klasifikasi yang berbeda. Pada penelitian (Dian, Emilia, et al., 2023) data yang digunakan adalah data tingkat pencemaran udara kota Jakarta tahun 2021 menggunakan

algoritma klasifikasi *Decision Tree*, hasil dari penelitian tersebut menyatakan bahwa klasifikasi menggunakan algoritma *Decision Tree* menunjukkan tingkat akurasi 100% dan tidak ada pembandingan dengan algoritma klasifikasi lainnya.

2.1. Data Mining

Data Mining merupakan proses yang digunakan untuk mencari informasi yang belum ditemukan dengan cara manual dari suatu kumpulan data (Hana et al., 2020). Proses pada Data mining menggunakan teknik matematika, statistik, kecerdasan buatan, dan machine learning untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakumulasi dari berbagai database besar. Menurut (Novianti et al., 2016) hal penting yang terkait dengan data mining diantaranya meliputi:

- a. Data mining adalah suatu proses otomatisasi terhadap dataset.
- b. Dataset yang akan diproses dalam proses data mining merupakan data yang sangat besar.
- c. Data mining ini memiliki tujuan yaitu untuk mendapatkan pola yang memberikan hasil klasifikasi/rekomendasi/klusterisasi yang bermanfaat.

2.2. Algoritma K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan sebuah algoritma atau Langkah-langkah untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran dengan jarak terdekat (Swantika et al., 2020). Nilai jarak pada metode K-NN didapatkan dengan menggunakan rumus *Euclidian Distance* (Kusuma et al., 2017). Akurasi yang baik terhadap hasil klasifikasi ini didapatkan dari langkah sederhana tersebut. Rumus *Euclidean Distance* seperti pada rumus (1) (Kusuma et al., 2017):

$$d(xy) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

d : jarak kedekatan

n : jumlah atribut antara 1 s.d n
 i : atribut individu antara 1 s.d n
 x : data training
 y : data testing

2.4. Algoritma Naive Bayes

Naive Bayes Classifier (NBC) sebuah metode statistik bayesian sederhana (Jadhav & Channe, 2013). Algoritma NBC merupakan salah satu algoritma pada data mining yang menerapkan teori *Bayes* dalam proses klasifikasi (Budianita, 2023). Proses klasifikasi menggunakan algoritma NBC ini dapat digunakan untuk data yang berukuran kecil sampai dengan data dengan ukuran besar. Berikut merupakan tahapan dalam proses klasifikasi menggunakan metode algoritma *Naive Bayes* :

1. Menghitung kelas probabilitas untuk mencari nilai prior probabilitas dari data training menggunakan rumus (2),

$$P(H) = \frac{x_i}{n} \quad (2)$$

dengan x_i merupakan nilai variabel ke- i dan n merupakan total dari ukuran sampel.

2. Selanjutnya menghitung nilai standar deviasi dan rata-rata dari masing-masing variabel nilai.
Untuk Menentukan nilai standar deviasi, gunakan rumus (3),

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n-1}} \quad (3)$$

simbol σ adalah proses pencarian standar deviasi.

Untuk menghitung nilai rata-rata yaitu pada rumus (4),

$$\mu = \frac{\sum_{i=1}^x x_i}{n} \quad (4)$$

dengan μ adalah nilai rata-rata dan x_i adalah nilai variabel ke- i

3. Menghitung nilai Densitas Gauss yaitu pada rumus (5),

$$P(X_i=x_i|Y=y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad (5)$$

P tersebut merupakan peluang, X_i merupakan variabel I dan Y merupakan kelas yang dicari.

4. Menghitung nilai posterior probabilitas yaitu pada rumus (6),

$$P\left(\frac{H}{X}\right) = \frac{P(X|H).P(H)}{P(X)} \quad (6)$$

Dengan X adalah data dengan kelas yang tidak terklasifikasi, H adalah data untuk hipotesis X yang merupakan kelas tertentu., $P(H|X)$ merupakan peluang H dengan syarat X terjadi, $P(H)$ merupakan kemungkinan yang dihipotesiskan H (probabilitas sebelumnya), $P(X|H)$ merupakan X kelayakan ditentukan oleh keadaan pada hipotesis H dan $P(X)$ adalah kemungkinan X .

2.5. Evaluasi Performance

Kinerja pada suatu sistem klasifikasi menggambarkan tingkat ukuran suatu model dalam klasifikasi data. Evaluasi ini bertujuan menguji model data mining untuk mengetahui kinerja model (Sang et al., 2021). Salah satu metode untuk mengukur evaluasi performa yang adalah menggunakan *confusion matrix*. *Confusion matrix* merupakan alat ukur untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang dipakai berbentuk sebuah matriks. Klasifikasi performa berdasarkan nilai akurasi disajikan pada tabel 1 (Arifin & Sasongko, 2018).

Tabel 1. Klasifikasi Performa Berdasarkan Nilai Akurasi

| Range Nilai | Klasifikasi |
|-------------|--------------|
| $\leq 60\%$ | Sangat Buruk |
| 60% - 70% | Buruk |
| 70% - 80% | Cukup |
| 80% - 90% | Baik |
| 90% - 100% | Sangat Baik |

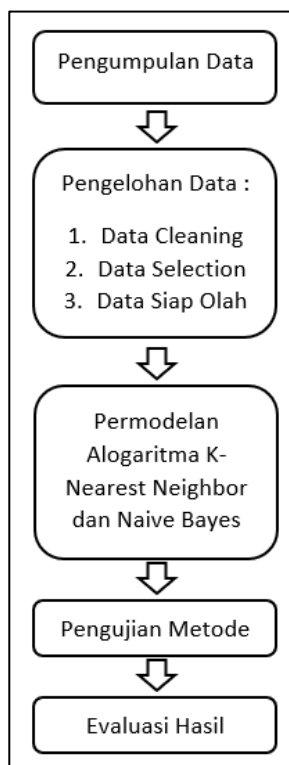
Mengukur valuasi performansi dapat dilakukan dengan precision untuk menunjukkan tingkat akurasi yang dapat didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual (Arifin & Sasongko, 2018).

Berikut persamaan dari perhitungan mencari akurasi:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

METODE

Proses penelitian ini dilakukan berdasarkan metode penelitian secara sistematis yang dirumuskan sebagai berikut:



Gambar 1. Metode Penelitian

1. Pengumpulan Data

Proses ini merupakan pengumpulan data yang dihasilkan berbentuk file spreadsheet berformat excel. Data yang didapatkan merupakan raw data/data mentah yang tidak beraturan pada beberapa sheet dan kolom. Sumber data berasal dari Kaggle.

2. Pembersihan Data

Dari data mentah tersebut selanjutnya dilakukan proses penghapusan atau pembersihan data. Proses ini dilakukan dengan cara mengeliminasi data yang tidak diperlukan atau data yang tidak berpengaruh terhadap penelitian yang dilakukan secara manual dengan bantuan *software Microsoft Excel*.

3. Seleksi Data

Proses seleksi data bertujuan untuk menentukan jenis data, sumber serta variabel yang tepat dan berdampak pada proses penelitian.

4. Pemodelan

Pada tahap ini dilakukan proses implementasi menggunakan metode Algoritma *K-Nearest Neighbor* dan *Naive Bayes* dalam pengolahan dataset.

5. Evaluasi dan validasi hasil

Langkah terakhir proses evaluasi dilakukan bertujuan untuk mengetahui hasil akhir perhitungan serta melihat apakah implementasi algoritma berfungsi dengan baik atau tidak, sehingga menghasilkan perbandingan yang baik dari hasil perhitungan masing-masing algoritma.

HASIL DAN PEMBAHASAN

Dari data ISPU Tangerang Selatan pada tahun 2022 merupakan pencatatan per-hari dari amatan pencemaran udara di Tangerang Selatan berdasarkan Indeks Standar Pencemar Udara diperoleh dari [Kaggle](https://www.kaggle.com/datasets/indonesia-air-quality) dengan jumlah data 365 record yang disajikan dalam 10 entitas

Tabel 2. Kolom Data dan Keterangan

| Entitas | Keterangan |
|---------|--|
| Tanggal | Tanggal dicatatnya tingkat pencemaran udara (harian) |
| PM2.5 | Partikel udara dengan ukuran lebih kecil dari 2.5 mikron (mikrometer). |
| PM10 | partikel udara dengan diameter 10 mikrometer |
| SO2 | Sulfida |
| CO | Karbon Monoksida |

| Entitas | Keterangan |
|--------------------|---|
| O3 | Ozon |
| NO2 | Nitrogen Oksida |
| Critical Component | Kategori hasil perhitungan indeks standar pencemaran udara (Baik, Sedang, Tidak Baik) |
| Category | Lokasi pengukuran tingkat pencemaran di Tangerang Selatan |

4.1 Data Indeks Standar Pencemaran Udara

Dari dataset yang digunakan sebagai bahan penelitian, setelah dilakukan data *preparation* yaitu mengeliminasi data dengan menghapus entri yang duplikat, nilai yang kosong, serta data yang tidak relevan atau tidak lengkap. Maka didapatkan data bersih dengan variabel respon (Y) berupa kolom kategori dan variabel prediktor berupa kolom PM2.5, PM10, SO2, CO, O3, NO2.

Tabel 3. Contoh dataset yang digunakan

| Date | Variabel Prediktor | | | | | | Max | Critical | Category |
|------------|--------------------|-------|------|----|----|-----|-----|----------|-----------|
| | PM 2.5 | PM 10 | SO 2 | CO | O3 | NO2 | | | |
| 01/01/2022 | 51 | 21 | 19 | 13 | 19 | 3 | 21 | PM2.5 | Good |
| 01/02/2022 | 51 | 21 | 19 | 12 | 19 | 3 | 21 | PM2.5 | Good |
| 01/03/2022 | 50 | 20 | 19 | 13 | 20 | 2 | 20 | PM2.5 | Good |
| - | - | - | - | - | - | - | - | - | - |
| - | - | - | - | - | - | - | - | - | - |
| 12/29/2022 | 49 | 19 | 15 | 11 | 31 | 4 | 49 | PM2.5 | Mode rate |
| 12/30/2022 | 50 | 19 | 16 | 12 | 31 | 4 | 50 | PM2.5 | Mode rate |
| 12/31/2022 | 51 | 20 | 16 | 12 | 31 | 4 | 51 | PM2.5 | Mode rate |
| 12/29/2022 | 52 | 21 | 16 | 13 | 31 | 4 | 52 | PM2.5 | Mode rate |

Tabel 4. Dataset hasil preprocessing

| NO | Variabel Prediktor | | | | | | Category |
|-----|--------------------|-------|-----|----|----|-----|----------|
| | PM 2.5 | PM 10 | SO2 | CO | O3 | NO2 | |
| 1 | 51 | 21 | 19 | 13 | 19 | 3 | Good |
| 2 | 51 | 21 | 19 | 12 | 19 | 3 | Good |
| 3 | 50 | 20 | 19 | 13 | 20 | 2 | Good |
| - | - | - | - | - | - | - | - |
| - | - | - | - | - | - | - | - |
| 360 | 49 | 19 | 15 | 11 | 31 | 4 | Moderate |
| 361 | 50 | 19 | 16 | 12 | 31 | 4 | Moderate |
| 362 | 51 | 20 | 16 | 12 | 31 | 4 | Moderate |
| 363 | 52 | 21 | 16 | 13 | 31 | 4 | Moderate |

4.2 Evaluasi Performance Algoritma K-Nearest Neighbor

Tabel 5. Hasil perbandingan nilai Akurasi algoritma K-Nearest Neighbor berdasarkan rasio

| Rasio | Akurasi |
|-------|---------|
| 60:40 | 86.30% |
| 70:30 | 91.74% |
| 80:20 | 87.67% |
| 90:10 | 94.44%. |

Dari tabel 5 diperoleh hasil akurasi algoritma K-Nearest Neighbor dengan akurasi terbaik dari perbandingan berdasarkan rasio pembagian data training dan testing adalah pada pembagian rasio 90 : 10 yaitu dengan tingkat akurasi 94.44%.

4.3 Evaluasi Performance Algoritma Naive Bayes

Tabel 6. Hasil perbandingan nilai Akurasi algoritma Naive Bayes berdasarkan rasio

| Rasio | Akurasi |
|-------|---------|
| 60:40 | 80.14% |
| 70:30 | 79.82% |
| 80:20 | 78.08% |
| 90:10 | 86.11% |

Dari tabel 6 diperoleh algoritma *Naive Bayes* dengan akurasi terbaik dari perbandingan berdasarkan rasio pembagian data training dan data adalah 90:10 yaitu dengan tingkat akurasi 86.11%.

Selanjutnya merupakan tabel evaluasi perbandingan antara algoritma *K-Nearest Neighbor* dan *Naive Bayes* pada setiap rasio yang digunakan.

Tabel 7. Perbandingan algoritma K-NN dan Naïve Bayes berdasarkan rasio

| Rasio | Akurasi | |
|-------|--------------------|-------------|
| | K-Nearest Neighbor | Naive Bayes |
| 60:40 | 86.30% | 80.14% |
| 70:30 | 91.74% | 79.82% |
| 80:20 | 87.67% | 78.08% |
| 90:10 | 94.44% | 86.11% |

Tabel 7 merupakan nilai akurasi yang dihasilkan berdasarkan nilai rata-rata pembobotan pada masing-masing *confusion matrix*. Pada rasio 90:10 memperoleh akurasi lebih tinggi yaitu algoritma *K-Nearest Neighbor* memiliki nilai tingkat akurasi lebih tinggi sebesar 94.44% dibandingkan tingkat akurasi pada algoritma *Naive Bayes* yaitu sebesar 86.11%, Maka peneliti menyimpulkan bahwa algoritma *K-Nearest Neighbor* memiliki nilai akurasi yang lebih unggul dari pada algoritma *Naive Bayes* dalam klasifikasi kualitas udara di Tangerang Selatan. Serta

dibuktikan dengan nilai akurasi disetiap rasio yang digunakan oleh algoritma *K-Nearest Neighbor* lebih tinggi daripada algoritma *Naive Bayes*.

4.4 Evaluasi Dari Kedua Algoritma

Tabel 8 Hasil Evaluasi Performance

| Metode | Accuracy | | True. Good | True. Moderat | Precision |
|-------------|----------|---------------|------------|---------------|-----------|
| K-NN | 94.44 % | Pred. Good | 5 | 1 | 83.33 % |
| | | Pred. Moderat | 1 | 29 | 96.67 % |
| | | Recall | 83.33 % | 96.67 % | |
| Naive Bayes | 86.11 % | Pred. Good | 3 | 2 | 60.00 % |
| | | Pred. Moderat | 3 | 28 | 90.32 % |
| | | Recall | 50.00 % | 93.33 % | |

Pada Tabel 8 merupakan nilai *Precision*, *Recall*, dan *Accuracy* kemudian menghitung akurasi tes *F1-Measure* dicapai berdasarkan nilai rata-rata pembobotan pada masing-masing *confusion matrix* berdasarkan rasio terbaik dari masing-masing algoritma, pada algoritma *K-Nearest Neighbor* menggunakan rasio 90:10

dan pada algoritma Naive Bayes menggunakan rasio 90:10 dengan perhitungan akurasi sebagai berikut:

$$\begin{aligned} \text{Accuracy K-NN} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{5 + 29}{5 + 1 + 1 + 29} \\ &= \frac{34}{36} \\ &= 0,9444 \end{aligned}$$

$$\begin{aligned} \text{Accuracy NBC} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{3 + 28}{3 + 2 + 3 + 28} \\ &= \frac{31}{36} \\ &= 0,8611 \end{aligned}$$

Pada algoritma K-Nearest Neighbor mendapatkan nilai *precision* sebesar 96.67%, *recall* 96.67%, *Accuracy* 94.44% dan *F1-Measure* 96,67 % dan pada algoritma Naive Bayes mendapatkan nilai *precision* sebesar 90.32%, *recall* 93.33%, *Accuracy* 86.11%, dan *F1-Measure* 91,80%. Berdasarkan hasil perhitungan klasifikasi diatas Algoritma K-Nearest Neighbor memiliki performa yang lebih baik dibandingkan dengan algoritma Naive Bayes, baik dari nilai *Precision*, *Recall*, *Accuracy* dan *F1-Measure*.

SIMPULAN

Berdasarkan hasil perhitungan menggunakan data ISPU Tangerang Selatan 2022 dengan metode klasifikasi algoritma K-Nearest Neighbor memiliki tingkat akurasi yang lebih baik dibandingkan dengan algoritma Naive Bayes untuk melakukan klasifikasi kualitas udara di Tangerang Selatan. Hasil tersebut didapatkan dari perhitungan klasifikasi algoritma K-Nearest Neighbor menggunakan rasio pembagian data training dan data testing yang sama yaitu pada rasio 90:10 kemudian

algoritma K-Nearest Neighbor mendapatkan nilai akurasi 94.44% dan pada algoritma Naive Bayes mendapatkan nilai akurasi 86.11%. Sehingga dapat disimpulkan Algoritma K-Nearest Neighbor memiliki akurasi tinggi dibandingkan algoritma Naive Bayes dalam klasifikasi data ISPU Tangerang Selatan 2022.

Berdasarkan hasil yang diperoleh pada penelitian ini diharapkan dapat menjadi suatu acuan bagi penelitian-penelitian selanjutnya dalam mengkaji klasifikasi kualitas udara baik di daerah yang sama maupun di daerah lainnya serta dapat dikembangkan menggunakan algoritma-algoritma klasifikasi yang lainnya.

DAFTAR PUSTAKA

- Arifin, O., & Sasongko, T. B. (2018). Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier. *Seminar Nasional Teknologi Informasi Dan Multimedia 2018*, 6(1), 67–72.
- Budianita, A. (2023). Information Gain Berbasis Algoritma Naive Bayes Classifier Pada Pemodelan Prediksi Kelulusan. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 5(1), 1–10. <https://doi.org/10.46772/intech.v5i1.1116>
- Dian, F., Emilia, R., & Gery, G. P. (2023). KLASIFIKASI TINGKAT PENCEMARAN UDARA KOTA JAKARTA TAHUN 2021 MENGGUNAKAN ALGORITMA DECISION TREE. *SEMINAR NASIONAL STATISTIKA AKTUARIA II* (2023). <https://www.data.jakarta.go.id/>
- Dian, F., Emilia, Rizka, G. G., & Indrayatna, F. (2023). Klasifikasi Tingkat Pencemaran Udara Kota Jakarta Tahun 2021 Menggunakan Algoritma Decision Tree. *Prosiding Nasional SNSA 2*, 127–131.
- Hana, F. M., Negara, D. S., & Khaqiqi, K. U. (2020). PERBANDINGAN ALGORITMA NEURAL NETWORK DENGAN LINIER DISCRIMINANT ANALYSIS (LDA) PADA

- KLASIFIKASI PENYAKIT DIABETES. *Jurnal Bisnis Digital Dan Sistem Informasi Volume, 1*, 1541–1541.
- Jadhav, S. D., & Channe, H. P. (2013). Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. In *International Journal of Science and Research (IJSR) ISSN* (Vol. 5). www.ijsr.net
- Kusuma, S. F., Pawening, R. E., & Dijaya, R. (2017). Otomatisasi klasifikasi kematangan buah mengkudu berdasarkan warna dan tekstur. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 3(1), 17–23. <https://doi.org/10.26594/register.v3i1.576>
- Novianti, B., Rismawan, T., & Bahri, S. (2016). Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak). *Jurnal Coding, Sistem Komputer Untan*, 04(3), 75–84.
- PP No 41 Tahun 1999. (n.d.). *PERATURAN PEMERINTAH REPUBLIK INDONESIA NOMOR 41 TAHUN 1999 TENTANG PENGENDALIAN PENCEMARAN UDARA*.
- Sang, A. I., Sutoyo, E., & Darmawan, I. (2021). Analisis Data Mining untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree dan Support Vector Machine. *E-Proceeding of Engineering*, 8(5), 8954–8963.
- Swantika, I. M. A., Kanata, B., & Suksmadana, I. M. B. (2020). Perancangan Sistem Untuk Mengetahui Kualitas Biji Kopi Berdasarkan Warna Dengan K-Nearest Neighbor. *Jurnal Bakti Nusa*, 1(2), 25–36.